University of Massachusetts Medical School
eScholarship@UMMS

GSBS Dissertations and Theses | Graduate School of Biomedical Sciences

2020-04-17

Molecular Mechanisms of Resistance and Structure-Based Drug Design in Homodimeric Viral Proteases

Gordon J. Lockbaum
University of Massachusetts Medical School

Let us know how access to this document benefits you.
Follow this and additional works at: https://escholarship.umassmed.edu/gsbs_diss

Part of the Biochemistry Commons, and the Structural Biology Commons

Repository Citation
Lockbaum GJ. (2020). Molecular Mechanisms of Resistance and Structure-Based Drug Design in Homodimeric Viral Proteases. GSBS Dissertations and Theses. https://doi.org/10.13028/368b-kn69. Retrieved from https://escholarship.umassmed.edu/gsbs_diss/1072

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

Molecular Mechanisms of Resistance and Structure-Based Drug Design in Homodimeric Viral Proteases

A Dissertation Presented

By

GORDON J. LOCKBAUM

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

April 17th, 2020

BIOCHEMISTRY & MOLECULAR PHARMACOLOGY

Molecular Mechanisms of Resistance and Structure-Based Drug Design in Homodimeric Viral Proteases

A Dissertation Presented By

GORDON J. LOCKBAUM

This work was undertaken in the Graduate School of Biomedical Sciences

BIOCHEMISTRY & MOLECULAR PHARMACOLOGY PROGRAM

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

_____

Daniel N. A. Bolon, Ph.D., Member of Committee

_____

William Royer, Ph.D., Member of Committee

_____

William Kobertz, Ph.D., Member of Committee

_____

Lizbeth Hedstrom, Ph.D., External Member of Committee

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

_____

Sean Ryder, Ph.D., Chair of Committee

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

_____

Mary Ellen Lane, Ph.D.

Dean of the Graduate School of Biomedical Sciences

APRIL 17th, 2020

ACKNOWLEDGEMENTS

First and foremost, thank you to my advisor, Dr. Celia Schiffer. I will always be grateful for your mentorship, for allowing me to join your lab, and for making me a better scientist and person. My sincerest thanks to Dr. Akbar Ali for always being there for me and for being an outstanding collaborator. I also thank Dr. Nese Kurt Yilmaz for your excellent feedback on my numerous writing projects. To Ellen Nalivaika, thank you for showing me the ropes and for being a rock in the lab. Special thanks to Dr. William Royer for sharing your structural biology expertise with me and for taking the time to answer all my questions.

Thank you to my Thesis Research Advisory Committee (TRAC) and Dissertation Exam Committee (DEC) members: Daniel N. A. Bolon, Bill Kobertz, Sean Ryder, Bill Royer, Brian Kelch, and Liz Hedstrom for giving me a generous amount of time and feedback and for pushing me to be a better scientist.

I thank the Schiffer Lab members, past and present. Florian Leidner, Shurong Hou, Jacqueto Zephyr, Anne Jecrois, Heather Loring, Adam Hedger, Mina Henes, Klajdi Kosovrasti, Dr. Ashley Matthew, Dr. Linah Rusere, Dr. Nagesh Desaboini, Dr. Jeong Min Lee, Dr. Jennifer Timm, Dr. Tania Silvas, Dr. Brendan Hilbert, and Dr. Djade Soumana. Thank you to those who directly helped me complete my projects and to all who made the lab a great environment for science.

I would like to take a moment to acknowledge and thank my friends and family. Thank you to my college roommates who stay in touch and always have my back: Eric Bunker, Jon Grenier, Ian Mahoney, Trevor Rappa, Matt Rawson, Read Servis, Sam Sperling, and Alex Stone. Thank you to my wonderful parents, Gordon C. Lockbaum and Linda Denise Lockbaum, for your unwavering support. Thank you to my sister, Olivia, and my brother, Luke, for being cool. Lastly, thank you to Gwen Walsh for making life fun during a brutal five years of graduate school. I love you all.

ABSTRACT

Drug resistance is a global health threat costing society billions of dollars and impacting millions of lives each year. Current drug design strategies are inadequate because they focus on disrupting target activity and not restricting the evolutionary pathways to resistance. Improved strategies would exploit the structural and dynamic changes in the enzyme–inhibitor system integrating data from many inhibitors and variants.

Using HIV-1 protease as a model system, I aimed to elucidate the underlying resistance mechanisms, characterize conserved protease-inhibitor interactions, and generate more robust inhibitors by applying these insights. For primary mechanisms of resistance, comparing interactions at the protease–inhibitor interface showed how specific modifications affected potency. For mutations distal to the active site, molecular dynamics simulations were necessary to elucidate how changes propagated to reduce inhibitor binding. These insights informed inhibitor design to improve potency against highly resistant variants by optimizing hydrogen bonding. A series of hybrid inhibitors was also designed that showed excellent potency by combining key moieties of multiple FDA-approved inhibitors. I characterized the structural basis for alterations in binding affinity in HIV-1 protease both from mutations and inhibitors.

I applied these strategies to HTLV-1 protease, a potential drug target. I identified the HIV-1 inhibitor darunavir as a viable scaffold and evaluated analogues, leading to a low-nanomolar compound with potential for optimization. Hopefully, insights from this thesis will lead to the development of potent HTLV-1 protease inhibitors. More broadly, these inhibitor design strategies are applicable to other rapidly evolving targets, thereby reducing drug resistance rates in the future.

TABLE OF CONTENTS Page #

ACKNOWLEDGEMENTS ..... **iii**
ABSTRACT ..... **iv**
LIST OF TABLES ..... **x**
LIST OF FIGURES ..... **xi**
LIST OF SCHEMES ..... **xiv**
LIST OF THIRD-PARTY COPYRIGHTED MATERIAL ..... **xv**
PREFACE ..... **xvii**
1 CHAPTER 1: Introduction **1**
1.1 Drug Resistance is a Major Biomedical Problem **2**
1.2 HIV is a Good Model of Evolution **3**
1.3 HIV Protease **5**
1.3.1 HIV Protease: The Enzyme 5
1.3.2 HIV Protease is a Good Drug Target 8
1.3.3 HIV Protease is an Ideal Drug Resistance Model System 12
1.3.4 HIV Protease Mutations 12
1.3.5 HIV Protease Inhibitors 15
1.3.6 HIV Protease Structures 16
1.3.7 HIV Protease Dynamics 17
1.4 The Rise of HTLV-1 and HTLV-1 Protease as a Drug Target **19**
1.5 Scope of Thesis **22**
2 CHAPTER 2: Structural Adaptation of Darunavir Analogs Against Primary Mutations in HIV-1 Protease **24**
2.1 Abstract **25**
2.2 Introduction **26**
2.3 Results **32**
2.3.1 Enzymatic Activity of Protease Variants 33
2.3.2 Effect of P1′ Modifications on Inhibition of Variants with Active Site Mutations 35
2.3.3 Structural Rearrangements Underlying DRV Susceptibility to Primary Resistance Mutations 36
2.3.4 Structural Response to Modifying P1′ Moiety Against Primary Mutations 42
2.3.5 I50V Mutation Induces Conformational Changes in Both UMass6 and Protease Flaps 44
2.3.6 Principal Component Analysis (PCA) of Residue-Specific Interactions 53
2.4 Discussion **57**
2.5 Materials and Methods **60**

2.5.1 Protease gene construction 60
2.5.2 Protein expression and purification 61
2.5.3 Enzymatic Assays to Determine K_m and K_i 62
2.5.4 Gag Polyprotein Cleavage Assay 63
2.5.5 Protein Crystallization 64
2.5.6 Data Collection and Structure Solution 64
2.5.7 Structural Analysis and PCA 65
2.6 Acknowledgements 67
3 CHAPTER 3: pM to μM–Elucidating the Role of Distal Mutations in HIV-1 Protease in Conferring Drug Resistance 68
3.1 Abstract 69
3.2 Introduction 70
3.3 Results 73
3.3.1 Viral passaging under drug pressure selected for highly mutated variants 73
3.3.2 Resistance mutations are selected to maintain catalytic activity 75
3.3.3 Accumulation of mutations progressively increases DRV resistance 79
3.3.4 Double mutant cycle reveals the interdependency of A71V and L76V 81
3.3.5 Crystal structures show resistance mutations reduce vdW contacts with DRV 81
3.3.6 Resistance mutations perturb packing in the hydrophobic core 90
3.3.7 Resistance mutations increase protease fluctuations 90
3.3.8 Inhibitor fluctuations increase and interactions are destabilized due to resistance mutations 96
3.3.9 Protease–DRV cross-correlations of fluctuations are lost with increasing resistance 101
3.3.10 Distal mutations impact the dynamic ensemble to confer resistance 104
3.4 Discussion 107
3.5 Methods 111
3.5.1 Protease Gene Construction. 111
3.5.2 Protein Expression and Purification 111
3.5.3 K_M Assay 112
3.5.4 K_i Assay 113
3.5.5 Gag Polyprotein Cleavage Assay 113
3.5.6 Crystallography 114
3.5.7 Data Collection and Structure Solution 115
3.5.8 Internal Distance Analysis of Crystal Structures 116
3.5.9 Intermolecular vdW Contact Analysis of Crystal Structures 116
3.5.10 System Preparation 116
3.5.11 Molecular Dynamics Simulations 117

3.5.12 Intermolecular vdW Contact Analysis of MD Trajectories 117
3.5.13 Cross-Correlations of Protease/DRV Fluctuations.....117
3.5.14 Distance-Difference Matrices 118
3.5.15 Jensen-Shannon Divergence Analysis of Dihedral Angles 119
3.6 Acknowledgements..... 119
4 CHAPTER 4: HIV-1 Protease Inhibitors Incorporating Stereochemically Defined P2′ Ligands to Optimize Hydrogen Bonding in the Substrate Envelope 120
4.1 Abstract..... 121
4.2 Introduction 122
4.3 Chemistry..... 129
4.4 Results and Discussion..... 133
4.4.1 Enzyme Inhibition Assays 133
4.4.2 Antiviral Assays.....139
4.4.3 Analysis of Protease-Inhibitor Complexes 146
4.5 Conclusions..... 164
4.6 Experimental Section..... 165
4.6.1 General 165
4.6.2 Antiviral Assays.....166
4.6.3 Protease Gene Construction 168
4.6.4 Protein Expression and Purification.....168
4.6.5 Enzyme Inhibition Assays 169
4.6.6 Protein Crystallization.....170
4.6.7 X-Ray Data Collection and Structure Solution 170
4.6.8 Molecular Modeling 171
4.7 Acknowledgments..... 172
5 CHAPTER 5: Structural Analysis of Potent Hybrid HIV-1 Protease Inhibitors Containing Bis-Tetrahydrofuran in a Pseudo-symmetric Dipeptide Isostere 173
5.1 Abstract..... 174
5.2 Introduction 175
5.3 Chemistry..... 180
5.4 Results and Discussion..... 182
5.4.1 Enzyme Inhibition and Antiviral Assays 182
5.4.2 Analysis of Protease-Inhibitor Complexes 186
5.4.3 Molecular Dynamics (MD) Simulations.....208
5.5 Conclusions..... 220
5.6 Experimental Section..... 221

5.6.1 General 221
5.6.2 Protease Gene Construction 223
5.6.3 Protein Expression and Purification 223
5.6.4 Enzyme Inhibition Assays 224
5.6.5 Antiviral Assays 224
5.6.6 Protein Crystallization 225
5.6.7 X-Ray Data Collection and Structure Solution 226
5.6.8 Intermolecular vdW Contact Analysis of Crystal Structures 227
5.6.9 Molecular Dynamics System Preparation 227
5.6.10 Molecular Dynamics Simulations 228
5.7 Acknowledgements 228
6 CHAPTER 6: Inhibiting HTLV-1 Protease: A Viable Target for Potent Antivirals 230
6.1 Abstract 231
6.2 Introduction 232
6.3 Results and Discussion 237
6.3.1 DRV and Analogs Inhibit HTLV-1 Protease 237
6.3.2 Inhibition of Gag Processing in HTLV-1 VLP's 240
6.3.3 Cocrystal Structures of DRV and Analogs Bound to HTLV-1 Protease 242
6.3.4 Interactions Lost in the Binding of DRV to HTLV-1 Versus HIV-1 Protease 244
6.3.5 Structural Basis of Improved Potency of DRV Analogs Toward HTLV-1 Protease 248
6.4 Conclusions 255
6.5 Methods 255
6.5.1 Construction of the HTLV-1 Protease Expression Plasmid 255
6.5.2 Protein Expression and Purification 255
6.5.3 Enzyme Binding Assays to Determine K_M 256
6.5.4 Enzyme Inhibition Assays to Determine K_i 257
6.5.5 Protein Crystallization 257
6.5.6 Data Collection and Structure Solution 258
6.5.7 Structural Analysis 259
6.5.8 Molecular Dynamics Simulations 259
6.5.9 Analysis from MD Simulations 260
6.6 Acknowledgements 261
7 CHAPTER 7: Optimal Refinement of Pseudomerohedrally Twinned $P6_1$ HIV-1 Protease–Inhibitor Cocrystal Structures 262
7.1 Abstract 263
7.2 Introduction 264

7.3 Materials and Methods 269
7.3.1 Protein Preparation and Crystallization 269
7.3.2 Data Collection and Processing 269
7.3.3 Molecular Replacement 270
7.3.4 Refinement 270
7.3.5 Visualization of Structures 270
7.4 Results and Discussion 271
7.4.1 Evaluating Twin Law and Inhibitor Orientation 271
7.4.2 Re-Solving Structure with a More Asymmetric Inhibitor 275
7.4.3 Re-Solving Structure with a Highly Asymmetric Inhibitor 280
7.5 Conclusion 284
8 CHAPTER 8: Discussion 285
8.1 Structural and Dynamic Analyses are Essential to Elucidate Complex Mechanisms of Resistance 286
8.2 Elucidating Conserved Static and Dynamic Protein-Inhibitor Interactions is Key to Inhibitor Optimization 287
8.3 Recipe for Robustness Against Resistance (Re-Format) 288
8.4 Future Directions: HTLV-1 Protease 290
8.5 Other Rapidly Evolving Disease Targets that Could Benefit From this Methodology 291
8.6 The Future of SBDD 296
8.7 Concluding Remarks 296
References 298

LIST OF TABLES

Page #

Table 2.1 Km of HIV-1 Protease Variants..... 34
Table 2.2 Enzyme Inhibition Constants (Ki)..... 36
Table 2.3 X-Ray Crystallography Statistics 37
Table 2.4 Change in Protease-Inhibitor vdW Interactions. 40
Table 2.5 Total Protease-Inhibitor vdW Interactions..... 43
Table 2.6 Hydrogen Bonding Distances (Å). 49
Table 3.1 Enzyme Kinetics of HIV-1 Protease Variants..... 76
Table 3.2 X-Ray Crystallography Statistics 83
Table 3.3 Protease–DRV Per Residue vdW Contacts..... 86
Table 4.1 Protease Inhibitory Activity. 135
Table 4.2 Antiviral Potency..... 140
Table 4.3 Antiviral Potency..... 141
Table 4.4 X-ray Data Collection & Crystallographic Refinement Statistics. 147
Table 4.5 Intermolecular Hydrogen Bonds (Å)..... 155
Table 5.1 Inhibitory Activity..... 183
Table 5.2 X-ray Data Collection & Crystallographic Refinement Statistics. 187
Table 5.3 Intermolecular Hydrogen Bonds (Å)..... 195
Table 6.1 HIV-1 Protease & HTLV-1 Protease Substrates/Cleavage Sites. 236
Table 6.2 Enzymatic activity of HTLV-1 protease. 238
Table 6.3 Inhibition Constants. 239
Table 6.4 X-ray Data Collection and Crystallographic Refinement Statistics.... 243
Table 7.1 X-ray Data Collection and Crystallographic Refinement Statistics.... 277

LIST OF FIGURES

Page #

Figure 1.1 HIV-1 Protease Structure and Dynamics. 6
Figure 1.2 Chemical Structures of FDA-Approved HIV-1 Protease Inhibitors 9
Figure 1.3 Diagram of Protease-Substrate Binding and Cleavage 10
Figure 1.4 Clinically Relevant Protease Inhibitors. 11
Figure 1.5 HIV-1 Protease Primary and Secondary Mutations 13
Figure 1.6 HIV-1 Protease Resistance Mutations 14
Figure 1.7 Comparison of HIV and HTLV-1 Protease 21
Figure 2.1 HIV-1 Protease-DRV Complex. 28
Figure 2.2 DRV and Analogues Aligned. 30
Figure 2.3 HIV-1 Gag Polyprotein Cleavage Assay 34
Figure 2.4 Alignment of Active Site Side Chains 41
Figure 2.5 Differences in Flap Orientation with Inhibitors Bound. 46
Figure 2.6 Comparison of Protease-Inhibitor Hydrogen Bonds. 48
Figure 2.7 Comparison of Distance-Difference Matrices 51
Figure 2.8 Mean Changes in C_α Distance Differences 52
Figure 2.9 Principal Component Analysis 54
Figure 2.10 Principal Component Analysis 55
Figure 3.1 Proximal and Distal HIV-1 Protease Mutations 74
Figure 3.2 Full-Length Gag Polyprotein Cleavage Assay 78
Figure 3.3 DRV Inhibition vs Enzyme Catalytic Efficiency. 80
Figure 3.4 Deviation of Internal C_α Distances 84
Figure 3.5 Structural Comparison of Resistant Variants. 87
Figure 3.6 Internal C_α Distance Differences 89
Figure 3.7 RMSD of Protein C_α Atoms During MD Simulations 91
Figure 3.8 RMSF of C_α Atoms from MD Simulations 92
Figure 3.9 RMSF of C_α Atoms from MD Simulations 92
Figure 3.10 Mean internal C_α Distance Differences from MD Simulations 94
Figure 3.11 Distance Between Specific C_α Atoms from MD Simulations 95
Figure 3.12 Alignment of DRV from Crystal Structures 97
Figure 3.13 RMSF of DRV Atoms During MD Simulations. 98
Figure 3.14 Hydrogen Bonding Frequencies from MD Simulations 100
Figure 3.15 Cross-Correlation of Fluctuations 102
Figure 3.16 Cross-Correlation of Fluctuations 103

Figure 3.17 Jensen-Shannon Divergence of Dihedral Angles 106
Figure 4.1 Chemical Structures of Synthesized Compounds..... 124
Figure 4.2 Comparison of Hydrogen Bonding Networks..... 127
Figure 4.3 EC50 vs cLogP 145
Figure 4.4 Comparison of Binding Modes of Synthesized Inhibitors..... 149
Figure 4.5 H-Bonding of Inhibitors Bound to WT HIV-1 Protease..... 150
Figure 4.6 Comparison of Binding Interactions in the S2′ Subsite. 151
Figure 4.7 Comparison of the P2′ Benzene Ring..... 152
Figure 4.8 vdW Packing of Inhibitors in the S2′ Subsite. 157
Figure 4.9 Differences in vdW Interactions 158
Figure 4.10 Comparison of Water Networks in the S2′ Subsite. 161
Figure 5.1 Chemical Structures of Synthesized Compounds 176
Figure 5.2 Binding Mode of Parent Compounds..... 189
Figure 5.3 Comparison of Inhibitor Binding Conformations. 191
Figure 5.4 H-Bonding of Inhibitors Bound to WT HIV-1 Protease..... 192
Figure 5.5 Comparison of Inhibitor Binding Conformations. 193
Figure 5.6 Comparison of Per Residue vdW Contacts of the Phe-Phe Core.... 197
Figure 5.7 Comparison of the Binding Conformations of the bis-THF Moiety... 199
Figure 5.8 Comparison of Per Residue vdW Contacts of the bis-THF Moiety .. 201
Figure 5.9 Comparison of H-Bond Interactions in the S2/S2′ Subsites..... 203
Figure 5.10 Comparison of Per Residue vdW Contacts in the S2/S2′ Subsite . 206
Figure 5.11 Comparison of vdW Packing in the S2′ Subsite..... 207
Figure 5.12 RMSD of Protein C_α Atoms During MD Simulations 208
Figure 5.13 Distance Distribution of C_α Atoms During MD Simulation 210
Figure 5.14 RMSF of Non-Hydrogen Atom Mapped onto Crystal Structures. .. 212
Figure 5.15 Dihedral Angle (φ5) vs. Ile50–Ile50′ Distance..... 214
Figure 5.16 Intra-Protease van der Waals Interactions. 214
Figure 5.17 Protein RMSF) of the Distal and Proximal Compounds..... 216
Figure 5.18 Protease-Inhibitor H-Bond Frequencies from MD Simulations. 218
Figure 5.19 Protease-Inhibitor H-Bond Frequencies from MD Simulations 219
Figure 6.1 Inhibitors Tested Against HTLV-1 Protease Inhibitors Tested Against HTLV-1 Protease..... 235
Figure 6.2 HTLV-1 Gag Cleavage by HTLV-1 Protease in VPLs..... 241
Figure 6.3 Comparison of DRV bound to HTLV-1 vs. HIV-1 protease..... 246
Figure 6.4 Comparison of Inhibitors Bound to HTLV-1 protease 249
Figure 6.5 Protease-Inhibitor Per-Residue vdW Contacts 250

Figure 6.6 Comparison of PU6 Bound to HTLV-1 and HIV-1 Protease. 252
Figure 6.7 RMSF of HTLV-1 Protease C_α atoms from MD Simulations. 254
Figure 6.8 π-π Stacking Interactions 254
Figure 7.1 HIV-1 Protease and Inhibitors Investigated. 266
Figure 7.2 HIV-1 Protease-Like Structures in the PDB by Space Group. 267
Figure 7.3 Electron Density Maps Surrounding the Inhibitor. 272
Figure 7.4 Internal Distance Differences 273
Figure 7.5 Electron Density Maps Surrounding the Inhibitor. 276
Figure 7.6 Internal Distance Differences 279
Figure 7.7 Electron Density Maps Surrounding the Inhibitor. 281
Figure 7.8 Internal Distance Differences 283
Figure 8.1 TMP Resistant Variants of DHFR 293
Figure 8.2 BCR-ABL and EGFR Drug Resistance Mutations 295

LIST OF SCHEMES

Page #

Scheme 4.1 Synthesis of protease inhibitors incorporating (*S*)- and (*R*)-4-(1-hydroxyethyl)benzene as P2′ ligands. 130
Scheme 4.2 Synthesis of protease inhibitors incorporating (S)- and (R)-4-(1,2-dihydroxyethyl)benzene as P2′ ligands. 132
Scheme 5.1 Synthesis of Protease Inhibitors Incorporating *Bis*-Tetrahydrofuran in a Pseudo-Symmetric Phe-Phe Dipeptide Isostere 181

LIST OF THIRD-PARTY COPYRIGHTED MATERIAL

Figure 1.1: Adapted from Trends in Microbiology, volume 24, issue 7, pages 547-557, Kurt Yilmaz, N.; Swanstrom, R.; Schiffer, C. A. Improving Viral Protease Inhibitors to Counter Drug Resistance. *Copyright,* 2016. Reprinted with permission from Elsevier (License Number: 4817220163749)

Figure 1.2: Adapted from Viruses, volume 2, issue 11, pages 2509-35, Ali, A.; Bandaranayake, R. M.; Cai, Y.; King, N. M.; Kolli, M.; Mittal, S.; Murzycki, J. F.; Nalam, M. N.; Nalivaika, E. A.; Ozen, A.; Prabu-Jeyabalan, M. M.; Thayer, K.; Schiffer, C. A., Molecular Basis for Drug Resistance in HIV-1 Protease. *Copyright,* 2010. Reprinted with permission from MDPI.

Figure 1.3: Adapted from Viruses, volume 2, issue 11, pages 2509-35, Ali, A.; Bandaranayake, R. M.; Cai, Y.; King, N. M.; Kolli, M.; Mittal, S.; Murzycki, J. F.; Nalam, M. N.; Nalivaika, E. A.; Ozen, A.; Prabu-Jeyabalan, M. M.; Thayer, K.; Schiffer, C. A., Molecular Basis for Drug Resistance in HIV-1 Protease. *Copyright,* 2010. Reprinted with permission from MDPI.

Figure 1.5: Adapted from Antiviral Research, volume 85, issue 1, pages 59-74, Wensing, A. M.; van Maarseveen, N. M.; Nijhuis, M., Fifteen years of HIV Protease Inhibitors: raising the barrier to resistance. *Copyright,* 2010. Reprinted with permission from Elsevier (License Number: 4817220795208)

Figure 1.6: Adapted from Topics in Antiviral Medicine, volume 27, issue 3, pages 111-121, Wensing, A. M.; Calvez, V.; Ceccherini-Silberstein, F.; Charpentier, C.; Gunthard, H. F.; Paredes, R.; Shafer, R. W.; Richman, D. D., 2019 update of the drug resistance mutations in HIV-1. *Copyright,* 2019. Reprinted with permission from IAS-USA.

LIST OF SYMBOLS, ABBREVIATIONS, NOMENCLATURE

Å	Angstrom
AIDS	Acquired Immunodeficiency Syndrome
APV	Amprenavir
ART	Antiretroviral Therapy
ATL	Adult T-cell Leukemia/Lymphoma
ATV	Atazanavir
Bis-THF	Bis-Tetrahydrofuran
C_α	Alpha Carbon
cART	Combination Antiretroviral Therapy
cLogP	Partition Coefficient
DAA	Direct Acting Antivirals
DMSO	Dimethyl Sulfoxide
DRV	Darunavir
EC50	Half Maximal Effective Concentration
IC50	Half Maximal Inhibitory Concentration
FDA	Food and Drug Administration
FRET	Forster Resonance Energy Transfer
ΔG	Change in Gibbs Free Energy
HIV	Human Immunodeficiency Virus
HTLV	Human T-cell Lymphotropic Virus
IDV	Indinavir
k_{cat}	Enzyme Turnover Rate
K_i	Inhibition Constant
K_M	Michaelis-Menten Constant
LPV	Lopinavir
μM	Micromolar
MD	Molecular Dynamics
MDR	Multidrug Resistant
nM	Nanomolar
NMR	Nuclear Magnetic Resonance
PC	Principal Component
PCA	Principal Component Analysis
PDB	Protein Data Bank
PI	Protease Inhibitor
pM	Picomolar
RMSD	Root Mean Square Deviation
RMSF	Root Mean Square Fluctuation
RTV	Ritonavir
SAR	Structure Activity Relationship
SBDD	Structure–Based Drug Design
TPV	Tipranavir
vdW	van der Waals
WT	Wildtype

PREFACE

Chapter II *is a collaborative study that has been previously published as:*

Lockbaum, G.J.; Leidner F.; Rusere, L.N.; Henes, M.; Kosovrasti, K.; Nachum G.S.; Nalivaika, E.A.; Ali, A.; Yilmaz, N.K.; Schiffer, C.A. "Structural Adaptation of Darunavir Analogues Against Primary Mutations in HIV-1 Protease." ACS Infectious Diseases. **2018**, *5*, 316-325.

Contributions from Gordon J. Lockbaum:

I performed all protein expression and purification of the HIV-1 protease variants in this study. I performed or oversaw all enzyme inhibition studies and analysis of the biochemical data. I determined all crystal structures and created the structural figures and tables. I analyzed and wrote the manuscript with Florian Leidner with guidance from Celia A. Schiffer and Nese Kurt Yilmaz.

Chapter III *is a collaborative study that has been previously published as:*

Henes, M.*; **Lockbaum, G.J.***; Kosovrasti, K.; Leidner, F.; Nachum, G.S.; Nalivaika, E.A.; Lee, S.K.; Spielvogel, E.; Zhou, S.; Swanstrom, R.; Bolon, D.N.A.; Yilmaz, N.K.; Schiffer, C.A. "pM to μM–Elucidating the Role of Distal Mutations in HIV-1 Protease in Conferring Drug Resistance." *ACS Chemical Biology* **2019**, *14*, 2441-2452. **joint first author*

Contributions from Gordon J. Lockbaum:

I performed all protein expression and purification of the HIV-1 protease variants in this study. I determined all crystal structures and created the structural figures and tables. I analyzed and wrote the manuscript with Mina Henes with guidance from Celia A. Schiffer and Nese Kurt Yilmaz.

Chapter IV *is a collaborative study that has been previously published as:*

Rusere, L.N.*; **Lockbaum, G.J.***; Lee, S.K.; Henes, M.; Kosovrasti, K.; Spielvogel, E.; Nalivaika, E.A.; Swanstrom, R.; Yilmaz, N.K.; Schiffer, C.A.; Ali, A. "HIV-1 Protease Inhibitors Incorporating Stereochemically Defined P2′ Ligands to Optimize Hydrogen Bonding in the Substrate Envelope." *Journal of Medicinal Chemistry*. **2019**, *62*, 8062-8079. **joint first author*

Contributions from Gordon J. Lockbaum:

I performed all protein expression and purification of the HIV-1 protease variants in this study. I determined all crystal structures and created the structural figures and tables. I analyzed and wrote the manuscript with Linah N. Rusere with guidance from Akbar Ali.

Chapter V *is a collaborative study that has been submitted for publication:*

Rusere, L.N.*; **Lockbaum, G.J.***; Henes, M.; Lee, S.K.; Spielvogel, E.; Rao, D.N.; Kosovrasti, K.; Nalivaika, E.A.; Swanstrom, R.; Yilmaz, N.K.; Schiffer, C.A.; Ali, A. "Structural Analysis of Potent Hybrid HIV-1 Protease Inhibitors Containing Bis-Tetrahydrofuran in a Pseudo-symmetric Dipeptide Isostere." **joint first author*

Contributions from Gordon J. Lockbaum:

I performed all protein expression and purification of the HIV-1 protease variants in this study. I determined all crystal structures and created the structural figures and tables. I analyzed and wrote the manuscript with Mina Henes with guidance from Akbar Ali.

Chapter VI *is a collaborative study that is being prepared for publication:*

Lockbaum, G.J.; Henes, M.; Talledge, N.; Kosovrasti, K.; Nalivaika, E.A.; Somasundaran, M.; Ali, A.; Mansky, L.; Yilmaz, N.K.; Schiffer, C.A. "Inhibiting HTLV-1 Protease: A Viable Target for Potent Antivirals."

Contributions from Gordon J. Lockbaum:

I performed all protein expression and purification of the HTLV-1 protease variants in this study. I determined all crystal structures and created the structural figures and tables. I analyzed and wrote the manuscript with Celia A. Schiffer and Nese Kurt Yilmaz.

Chapter VII *is a collaborative study that has been previously published as:*

Lockbaum, G.J.; Leidner, F.; Royer, W.E., Yilmaz, N.K.; Schiffer, C.A. "Optimizing the Refinement of Merohedrally Twinned $P6_1$ HIV-1 Protease–Inhibitor Cocrystal Structures." *Acta Crystallographica Section D*. **2020**, *76*, 302-310.

Contributions from Gordon J. Lockbaum:

I devised the concept of this manuscript and performed all crystallographic refinements. I created all figures and tables in this study. I wrote the manuscript with guidance from William E. Royer, Celia A. Schiffer, and Nese Kurt Yilmaz.

CHAPTER 1:

Introduction

1.1 Drug Resistance is a Major Biomedical Problem

Antimicrobial resistance is a global threat to public health that requires immediate action and a concerted effort.[1-2] Antimicrobials include antibiotics, antivirals, antifungals, and antiprotozoals. More broadly, drug resistance also includes resistance to pesticides, herbicides, and cancer therapeutics. Drug resistance costs our society billions of dollars annually and causes at least 700,000 deaths globally per year, which could increase to over 10 million by 2050 if no action is taken.[3-4]

Over-prescription and poor adherence contribute to resistance, but resistance ultimately occurs any time evolution exists under the selective pressure of a drug that restricts growth of the wild-type yet does not entirely inhibit the growth of other evolutionarily accessible variants. Under sub-optimal drug concentrations, these heterogeneous populations eventually render effective drugs obsolete. Resistance is partially an inadvertent consequence of drug design where avoiding the probability of resistance is not included as part of the design strategy.

Therapeutics take many forms, but most drugs are small molecules that bind to proteins or enzymes and alter their function in a way that cures or alleviates the disease. Most drug precursors are discovered by screening hundreds or thousands of compounds in a high-throughput inhibition assay against a particular macromolecular therapeutic target. After a lead compound is discovered, the inhibitor goes through iterative rounds of optimization to improve potency and binding affinity. When no structure is available, structure activity relationship (SAR)

studies are used to determine how changes to the inhibitor affect potency. When a protein-ligand structure is available, structure-based drug design (SBDD) is employed, where the inhibitor is designed in three-dimensional space to optimally interact with the target.

These techniques, however, do not focus on the biologically relevant interactions of the target, but rather focus solely on disrupting the target's activity. Disrupting the therapeutic target's activity is necessary but not sufficient for developing a robust drug with a lower probability of resistance. The high rates of target-site drug resistance suggest that our current drug design paradigm is insufficient. Restricting the evolutionary pathways to resistance requires further understanding the underlying molecular mechanisms: the structural and dynamic changes in the enzyme–inhibitor system. Improved drug design strategies would integrate data across many inhibitors and variants to optimize conserved enzyme-inhibitor interactions thereby making inhibitors more robust and reduce rates of resistance.

1.2 HIV is a Good Model of Evolution

In biological research, model systems are used to probe different questions by monitoring genotypic and phenotypic changes in response to different variables. Bacteria, yeast, flies, worms, fish, mice, and viruses can function as a model system. Human Immunodeficiency Virus (HIV) is a great model system to test evolution and adaptation. HIV is known for its large amount of heterogeneity, even within a single human. In 1996, it was discovered that one person living with HIV

for six years had the same viral diversity as influenza across the entire planet.[5] HIV's genetic diversity comes from a high rate of replication,[6-8] error prone reverse transcriptase,[9-10] propensity for homologous recombination,[11] and interactions with host mRNA editing enzymes like APOBECs.[12]

HIV's diversity has been illuminated by the decreasing cost of genetic sequencing and the standard practice of viral sequencing in clinics.[13-14] One repository for these sequences is the Stanford University HIV Drug Resistance Database, which has over one hundred thousand clinical isolates obtained from persons infected with HIV-1.[15-16] This database highlights HIV's ability to tolerate many mutations throughout its genome while retaining viral fitness. Databases like these help scientists to determine which treatments cause mutations in the viral genome. Specific drugs can induce trademark mutations in different viral enzymes and monitoring these mutations allow doctors to better prescribe more effective treatments.[17]

Viral sequences derived from patient isolates are incredibly valuable for research, but they represent a snapshot in time, often with limited context like patient history, including adherence and drug regimens. Generating resistant variants in a laboratory is commonly used to determine relative rates of resistance. Challenging a heterogenous pool of virus with a sub-optimal dose of inhibitor selects for resistant variants fit enough to pass their genes on to the next generation. Slowly increasing the inhibitor concentration with each viral passage generates extremely resistant proteins over time. This longitudinal experiment with

dosage and temporal control allows determination of the order of mutations against specific drugs at specific inhibitor concentrations.

1.3 HIV Protease

1.3.1 HIV Protease: The Enzyme

HIV protease is a necessary enzyme in the viral life cycle, responsible for maturation of the virion into an infectious particle.[18] HIV-1 protease consists of two 99 amino acid monomers held together by a dimerization domain to form a functional dimer [**Figure 1.1A**].[18-19] The protease has two flaps which are required to open to allow binding and close around amino acid substrates [**Figure 1.1B**].[18-21] In the closed conformation, the active site resembles a tunnel which orients substrate cleavage sites near the protease's catalytic aspartates for hydrolysis to occur.

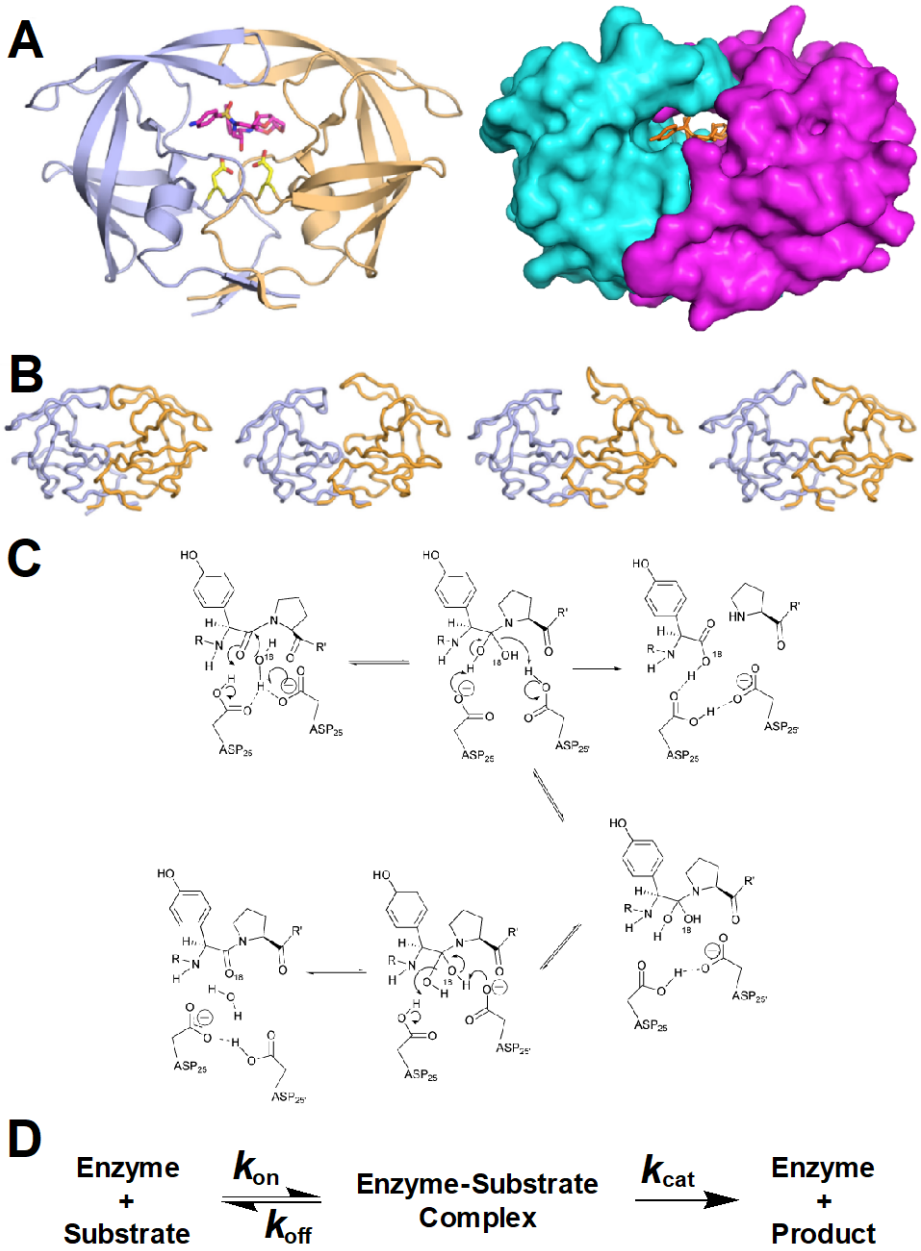


Figure 1.1 HIV-1 protease structure and dynamics. (A) HIV-1 protease dimer shown as cartoon with inhibitor bound and catalytic aspartates shown as yellow sticks. HIV-1 protease dimer shown in surface representation.[21] (B) Frames taken from a molecular dynamics simulation showing HIV-1 protease in various conformations, including closed, partially opened, and fully open.[21] (C) The reaction mechanism for HIV-1 protease catalyzed proteolysis .[22] (D) Michaelis-Menten one-substrate reaction.

The reaction mechanism for peptide bond cleavage is shown in **Figure 1.1C**,[22] where the catalytic aspartates perform a coordinated, water-mediated backside attack of the carbonyl carbon which creates an unstable geminal diol tetrahedral intermediate. The intermediate can either collapse in a way that completes hydrolysis or that re-forms the substrate. By monitoring isotope exchange, the reformation of the substrate was found to occur at a rate of 0.01-0.1 times that of the forward reaction.[23] Because a majority of the enzyme-substrate transition state is suscessfully cleaved into product, the reaction can be modeled with a Michaelis-Menten one-substrate mechanism [**Figure 1.1D**]. This comes with a few assumptions such as the steady state approximation, free ligand approximation, and the rapid equilibrium approximation. *In vitro* studies measuring product formation using very low concentrations of protease and increasing concentrations of hydrolysable substrate allow the determination of the initial velocities. The initial velocities as a function of substrate concentration are used to determine V_{max}, k_{cat}, and K_m, where K_m is approximately the ratio of k_{on} and k_{off}. The experiments show the proteolysis reaction can be explained by the Michaelis-Menten equation and fits the assumption that the k_{cat} is the rate limiting step in the reaction. Therefore, in HIV-1 protease, despite the large dynamic movements of the flaps, the rate determining step is the hydrolysis and product release, rather than the binding of substrate.[23-24]

1.3.2 HIV Protease is a Good Drug Target

HIV-1 protease inhibitors (PIs) are considered one of the greatest achievements in SBDD. The first structure of HIV-1 protease was determined in 1989 and by 1996 three drugs were approved by the Food and Drug Administration (FDA) and immediately improved the life expectancies of those treated.[25-26] In total, there have been 9 FDA-approved PIs: Saquinavir (SQV), Indinavir (IDV), Ritonavir (RTV), Nelfinavir (NFV), Amprenavir (APV), Lopinavir (LPV), Atazanavir (ATV), Tipranavir (TPV), and Darunavir (DRV) [**Figure 1.2**].[27] Besides TPV, all inhibitors are linear peptidomimetics. Currently, only LPV, ATV, and DRV are used in clinic and all are taken with low-dose RTV to improve pharmacokinetics.[28-29] These inhibitors have very high potency, with affinity to wild type (WT) protease in the low-picomolar (pM) range.[27]

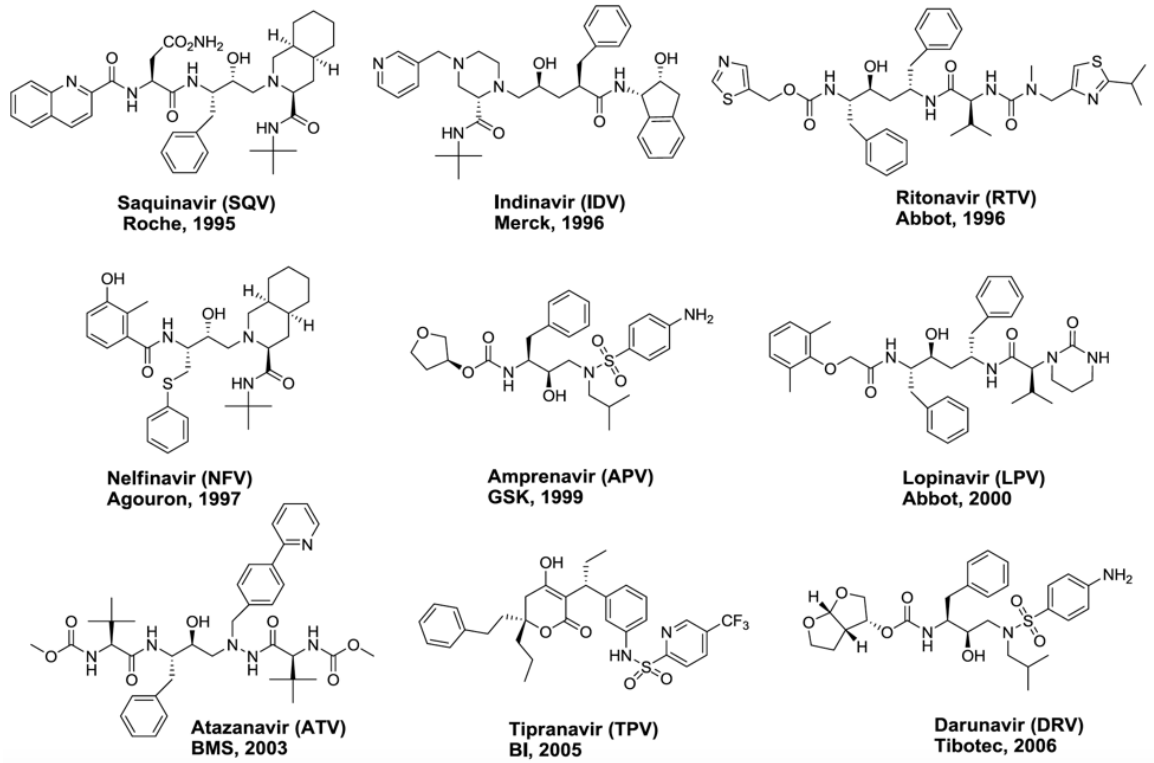


Figure 1.2 Chemical structures of FDA-approved HIV-1 PIs, including their three-letter abbreviation, the drug manufacturer and the year the drug was approved by the FDA.[27]

All nine PIs have similar characteristics that contribute to their high affinity for HIV-1 protease. Each inhibitor has a central hydroxyl moiety which functions as a non-hydrolysable transition state mimetic [**Figure 1.3**] and creates strong hydrogen bond interactions with both catalytic aspartates. All inhibitors, besides TPV, also coordinate a “flap water” between the inhibitor and the flaps which helps to keep the flaps in the closed conformation [**Figure 1.4**]. In addition, each inhibitor makes a number of other direct and water-mediated hydrogen bonds as well as van der Waals (vdW) interactions with hydrophobic side chains in the active site.[27]

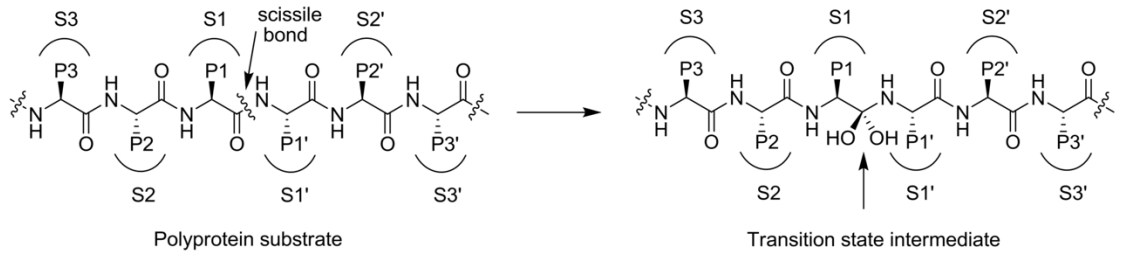


Figure 1.3 Simplified diagram of an amino acid polymer bound to HIV-1 protease and the transition state intermediate formed before hydrolysis.[27]

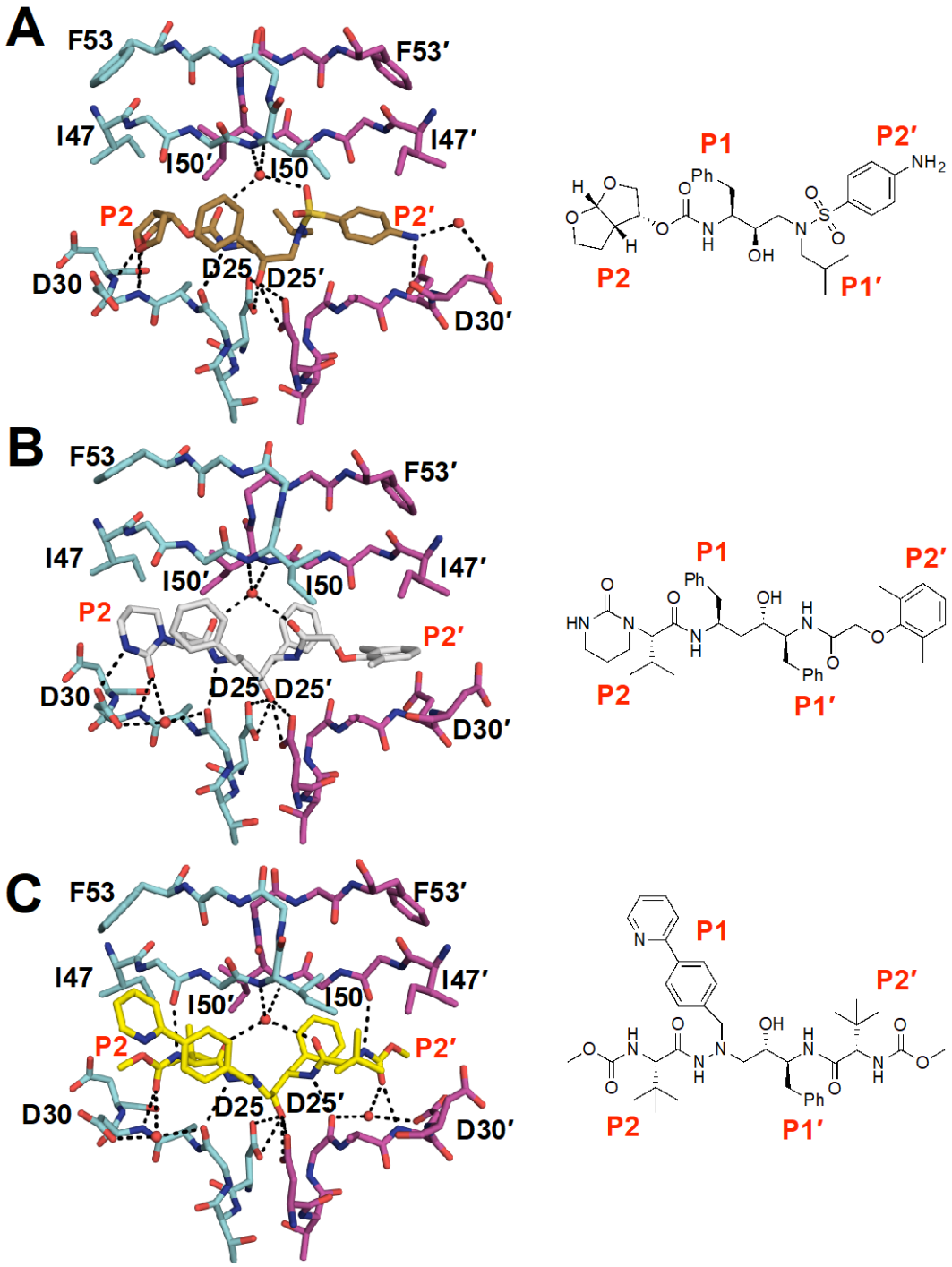


Figure 1.4 Clinically relevant protease inhibitors. (A) DRV, (B) LPV, and (C) ATV bound to HIV-1 protease. All are shown with their respective chemical structures with the peptidomimetic moieties labeled P2 to P2′.

1.3.3 HIV Protease is an Ideal Drug Resistance Model System

HIV's propensity for genetic heterogeneity makes any of its enzymes a good model system to study drug resistance. HIV-1 protease is best suited to elucidate complex mechanisms of resistance and identify enzyme-inhibitor characteristics to optimize, due to 1) thousands of viable protease variants, 2) thousands of published inhibitors with inhibition data, 3) hundreds of high-resolution protein-ligand complexes, and 4) thousands of molecular dynamics simulations.

1.3.4 HIV Protease Mutations

Despite needing to maintain substrate specificity, HIV-1 protease is exceptionally mutable. About two-thirds of its 99 amino acid sequence can tolerate a single site mutation and retain functionality, while about half of its sequence has been directly implicated in drug resistance.[17, 30] Under selective pressure, when challenged by an inhibitor, some patient isolates have replicating viruses with functional proteases that contain over 25 mutations.[31-36]

Resistance mutations can be found throughout the protease and are sometimes classified as either "primary" or "secondary," as depicted in **Figure 1.5**.[30] Primary mutations usually occur first and function to reduce inhibitor binding affinity. Primary mutations are often located in or around the active site, which are referred to as "proximal" mutations. Secondary mutations generally appear later, away from the active site, or "distal", and are thought to be mostly compensatory as they often restore functionality to the enzyme.[30]

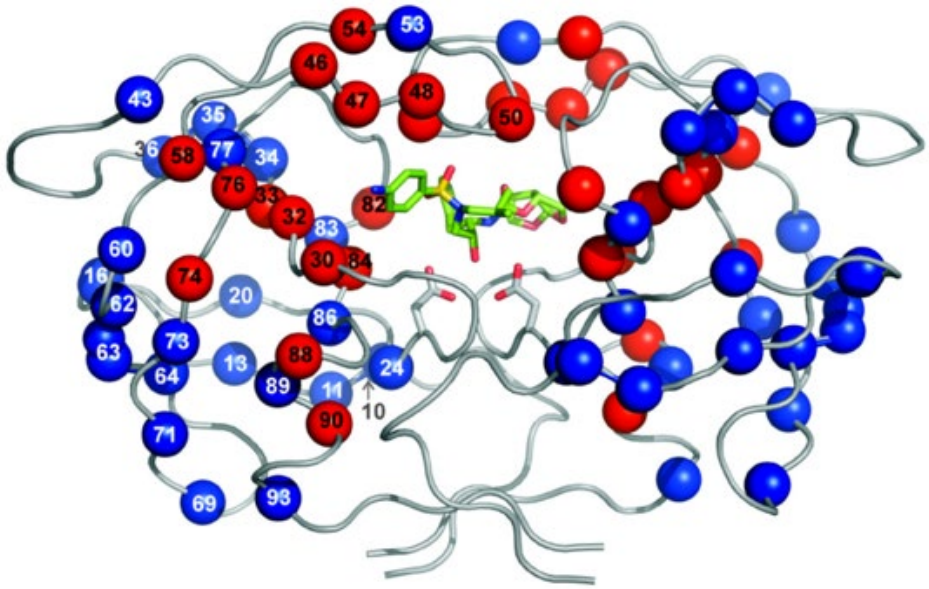


Figure 1.5 Primary (red) and secondary (blue) mutations mapped onto the structure of HIV-1 protease.[30]

Primary mutations tend to directly perturb protease-inhibitor interactions, but if the protease variant has reduced inhibitor binding affinity, then it likely has reduced affinity for natural substrates as well. If the protease's affinity of natural substrates is too low, then the virus will not be viable. Therefore, secondary mutations are necessary to restore affinity for natural substrates or to alter the protease dynamics to improve catalytic efficiency.

Primary/proximal mutations have been extensively investigated in HIV-1 protease.[27, 37-39] Some primary mutations are pan-resistant against all protease inhibitors, but most FDA-approved PIs have signature mutations, especially drugs that are no longer used in clinic [**Figure 1.6**].[17] In general, competitive inhibitors are most affected by proximal mutations that significantly perturb protease-inhibitor

contacts. This highlights the importance of proactively identifying which active site residues are variable and to minimize inhibitor interaction with those residues. 30 years of HIV protease research indicates the impact of proximal mutations on inhibitor binding affinity can be readily understood, while the impact of distal mutations is less obvious.

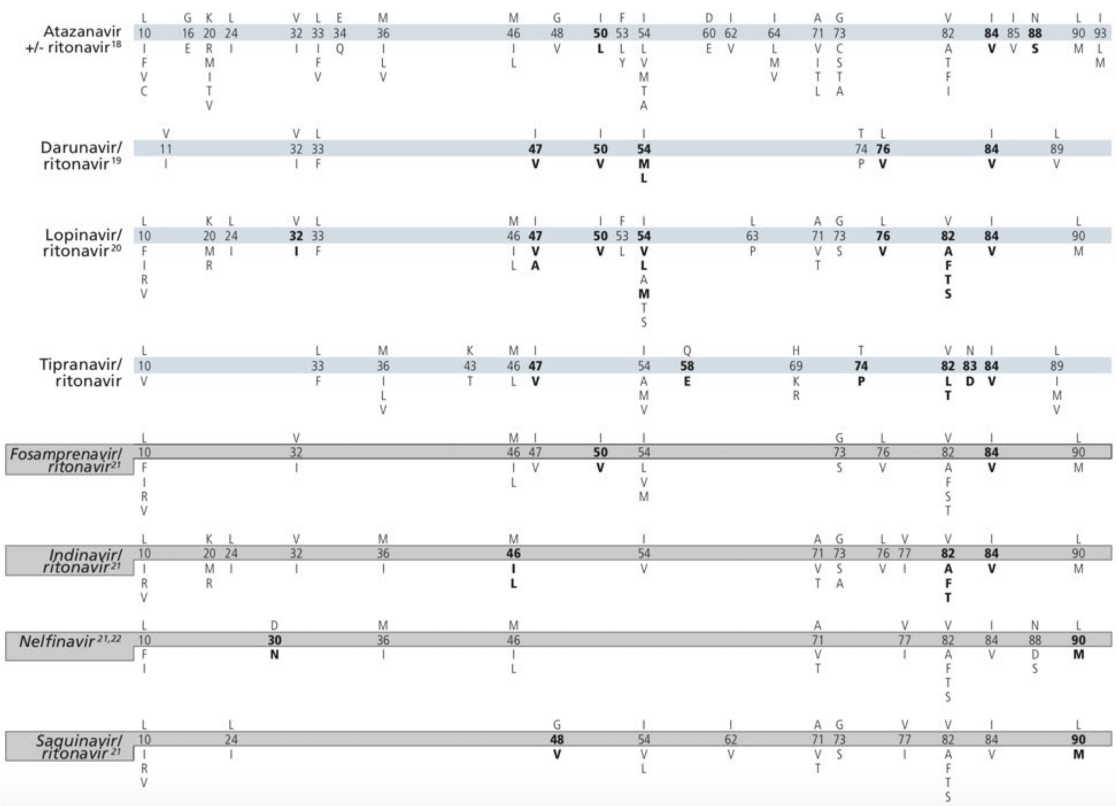


Figure 1.6 Mutations in HIV-1 protease associated with resistance to each protease inhibitor.[17]

By definition, distal mutations cannot directly contact the inhibitor. Distal mutations are generally considered either a naturally-occuring polymorphisms or a compensatory mutation. Compensatory mutations have received a fair amount of research and are known to restore functionality by altering protease dynamics, like in the flaps. A most prevalent example is the I50V/A71V mutation combination.[40] The I50V active site mutation at the tip of the flaps reduces affinity for inhibitors, but also significantly reduces rates of substrate cleavage. The A71V mutation, over 13 Å away from the active site, restores substrate cleavage rates to near WT levels, while maintaining resistance against the inhibitor. There has been very little research regarding if distal mutations contribute to reductions in inhibitor binding.

1.3.5 HIV Protease Inhibitors

HIV-1 protease cleaves the viral polyproteins at 11 different sites.[10] Maintaining relative binding affinity and cleavage rates between the substrates is critical for proper viral maturation.[41-42] Although the substrate sequences are highly diverse, the sequences have a conserved volume when aligned which our lab has termed the “substrate envelope”.[43] Inhibitors that protrude outside of the substrate envelope are more susceptible to resistance mutations that perturb inhibitor binding without affecting substrate binding.[21]

Darunavir (DRV) is a special inhibitor because it is relatively small, extremely potent, and has a good resistance profile as it fits well within the substrate envelope.[44] Despite being the most potent FDA-approved PI with low resistance

rates when properly administered, DRV greatly loses binding affinity against mutated protease variants.[31, 38] Our lab found that small modifications to DRV's P1′ and P2′ moieties improved inhibitor potency against a panel of resistant proteases.[45] Over the years, each of DRV's peptidomimetic positions, P2 to P2′, has undergone extensive SAR studies, resulting in hundreds of DRV analogues with different combinations of modifications. Changing inhibitors in a systematic way and testing them against a panel of resistant protease variants helps elucidate and optimize for the most important protease-inhibitor interactions.

1.3.6 HIV Protease Structures

Since the structure of HIV-1 protease was determined in 1989, it has become one of the most published structures, with over 700 high-resolution (< 2.5 Å) protease-inhibitor complexes deposited into the Protein Data Bank (PDB).[46] Despite being a success story of SBDD, the first published HIV-1 protease structure was infamously incorrect, with residues missing from the C terminus, a lack of a central alpha helix, and a poorly modeled dimerization domain.[20] Luckily, several other labs had crystalized the protease and it was corrected shortly after.[19]

X-ray crystallography relies on the repetitive and symmetrical nature of crystals, but internal symmetry, like the pseudo-C2-symmetric nature of the two monomers of HIV-1 protease, can complicate structural determination. In the apo form, the two monomers are highly symmetric, but binding an inhibitor or substrate induces asymmetry in the protease, especially in the flaps.[47-48] Certain space groups, like $P2_12_12_1$, are ideal because they have two unique monomers in the

asymmetric unit and the electron density of the inhibitor is found in one conformation. Other space groups, such as $P6_1$, retain high levels of symmetry between the monomers, so the data appears merohedrally twined and the inhibitor density is such that the inhibitor is best modeled in two conformations.[49-50] This twinning can average out subtle asymmetric differences between the monomers and inhibitor and lead to a sub-optimal model. Even slight imperfections in the model can mislead SBDD efforts, therefore highly accurate cocrystal structures are critical.

1.3.7 HIV Protease Dynamics

Although so much data can be derived from cocrystal structures alone, proteins exist in a dynamic equilibrium in solution and the conformational dynamics of the enzyme–inhibitor complex should not be overlooked. The dynamic analyses are a necessary component to elucidate mechanisms of resistance and to properly analyze protease-inhibitor interactions, critical for designing robust drugs to inhibit a dynamic target.

Protein structures determined by X-ray crystallography are inherently “packed” together, where each macromolecule interacts with neighboring proteins at several crystal contacts. Although a protein’s size and shape might not change much, dynamic regions can be highlighted through other structural techniques like protein nuclear magnetic resonance (NMR). Protein NMR determines the structure of a protein in solution by detecting the local molecular environment of each atom. A single protein NMR dataset can be used to determine dozens of structures and

indicate which regions are more flexible than others. Using protein NMR, our collaborators have shown how protein dynamics change with the addition of drug resistance mutations and how those mutations reduce inhibitor potency.[51]

Our lab has specialized in utilizing molecular dynamics (MD) simulations to evaluate protease-inhibitor dynamics *in silico*. MD simulations take a structure, put it in a box, solvate that box, and apply force fields around each atom and let each atom bounce around and interact for a period of time. 3D information about every atom's position over time is recorded for further analysis. The quality of a simulation is corelated with the accuracy of the starting model, further highlighting the importance of accurate structures.

Like protein NMR, MD simulations can be used to elucidate mechanisms of resistance, especially distal mutations reduce inhibitor affinity. To affect inhibitor binding from a distance, distal mutations must propagate changes through the protease in a dynamic way. Using MD simulations, we have shown how some mutations alter protease dynamics through changes in hydrophobic packing within the core of the protease.[52-53] We have also shown how certain mutations alter the hydrogen bond network throughout the protease resulting in drug resistance.[54]

MD simulations can also explain how inhibitor modifications improve potency. Our lab has showed that just monitoring inhibitor fluctuations over the course of a simulation was not well correlated with inhibitor potency, but maintaining correlated motion with the protease was a better indicator of binding affinity.[55]

1.4 The Rise of HTLV-1 and HTLV-1 Protease as a Drug Target

The most logical place to transfer over 20 years of knowledge inhibiting HIV was to try to inhibit a related virus, Human T-cell lymphotropic virus (HTLV). HTLV was actually the first human retroviruses described nearly 40 years ago.[56-57] For decades HTLV-1 has been known to be highly carcinogenic and to cause severe paralytic neurologic disease as well as immune disorders that can increase susceptibility to bacterial infections. HIV was thought to be related to HTLV when it was first classified. HTLV-1 is transmitted via the same routes as HIV-1 (sexually, via blood or mother-to-child) with significant HIV/HTLV co-infections reported in Europe, America, and Africa.[58] About 20 million people worldwide are infected with HTLV-1.[59] Unfortunately, there are no effective vaccines or direct acting antivirals (DAAs) against HTLV-1.

Treatment options for HTLV-1 infected patients are scarce. For the estimated 3–5% of the HTLV-1 infected patients who develop adult T-cell leukemia/lymphoma (ATL), chemotherapy is ineffective with very poor survival rates and relapse leading to death.[60-61] In the clinic, a combination of Zidovudine, an HIV-1 reverse transcriptase inhibitor, and IFN-α significantly improved prognosis, indicating active HTLV-1 replication in ATL patients that can be targeted by antiretrovirals. As with HIV-1 infections, development of specific DAAs has a great potential to improve patient/treatment outcomes.

HTLV-1 has a similar viral live cycle as HIV-1 and utilizes similar viral machinery such as reverse transcriptase, integrase, and protease. Success

against HIV-1 protease suggests similar potency can be achieved against HTLV-1 protease. HTLV-1 protease is a 28 kDa homodimeric aspartyl protease that is 28% identical to HIV-1 protease with 45% identity between active site residues,[62-63] yet HTLV-1 protease has considerably distinct substrate specificity [**Figure 1.7**].[63-67] The differences in substrate specificity suggest why the FDA-approved HIV-1 protease drugs bind HTLV-1 protease with less affinity, with the best candidates reaching low nanomolar potency.[68] A number of novel HTLV-1 protease inhibitors have been published, yet none have reached sub-nanomolar potency.

The structure of HTLV-1 protease was determined in 2005,[65] but only 10 WT structures have been deposited into the Protein Data Bank (PDB).[69-70] Some *in silico* drug design has been reported, but no studies have utilized MD simulations on this complex. Also, although there is no data on drug resistant mutations in HTLV, our HIV protease methodology can be used to first determine a scaffold that will ultimately result in a clinically relevant HTLV-1 protease inhibitor. The substrate envelope model can be used to develop robust inhibitors likely to be less susceptible to potential drug resistance.

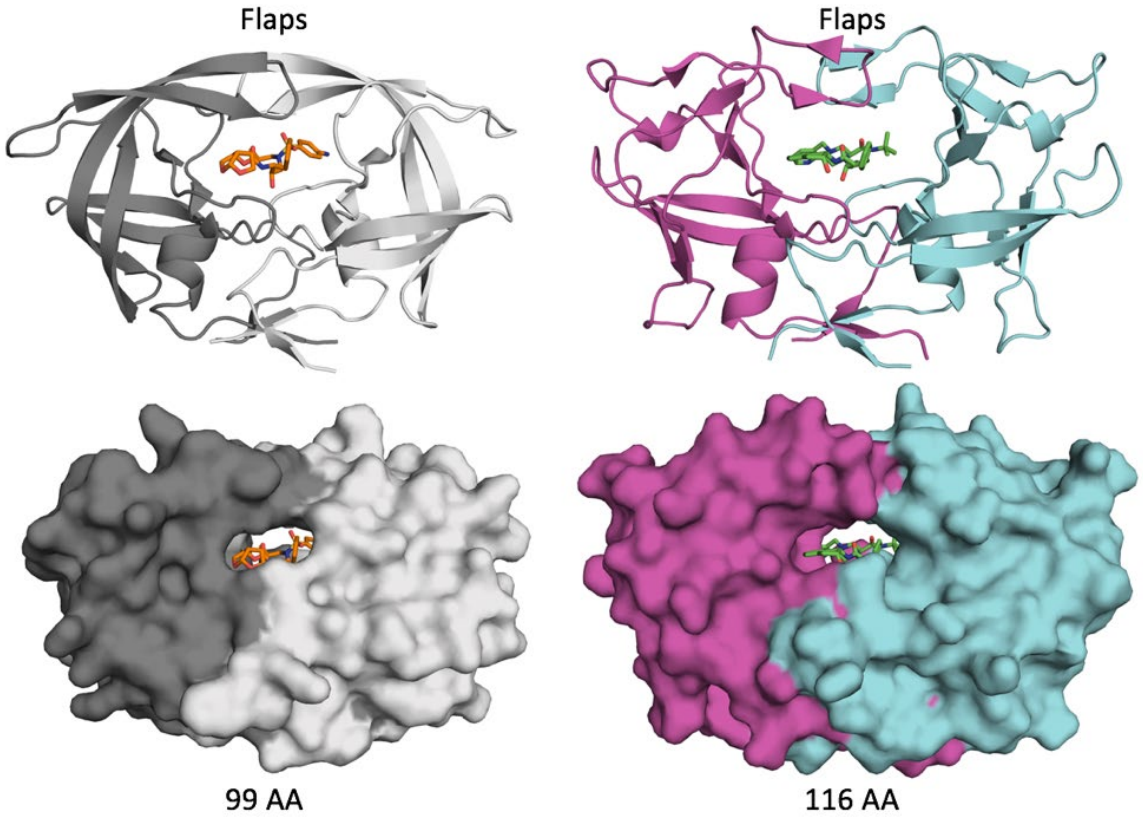


Figure 1.7 Comparison of HIV and HTLV-1 protease. HIV-1 protease dimer shown with cartoon and surface representations (PDB: 1T3R) and HTLV-1 protease dimer shown with cartoon and surface representations (PDB: 3WSJ).

1.5 Scope of Thesis

This thesis attempts to: 1) understand molecular mechanisms of drug resistance against HIV-1 PIs, arising from both proximal and distal mutations, 2) evaluate HIV-1 PI design strategies to improve potency and resistance profiles, and 3) establish a viable/novel inhibitor scaffold against HTLV-1 protease.

Chapter 2 and Chapter 3 explore molecular mechanisms of drug resistance. In Chapter 2, while investigating primary resistance mechanisms, comparing changes at the protease-inhibitor interface showed why certain P1' modifications were more potent against one common primary resistance mutation but more susceptible to another. In Chapter 3, for mutations distal to the active site, molecular dynamics simulations were necessary to elucidate that dynamic changes were propagated throughout the protease resulting in dramatically reduced DRV binding through weakened protease-inhibitor interactions. This research also established a new viable protease variant with 11 mutations that has the highest reported resistance to DRV (760 nM), binding DRV with over 150,000x lower affinity than WT protease.

Chapter 4 and Chapter 5 involve evaluating HIV-1 PI design strategies and modifications. In Chapter 4, we show how modifications to the P2' moiety improved potency against highly resistant variants by optimizing hydrogen bonds with the protease backbone. In Chapter 5, we characterized a series of novel hybrid inhibitors that showed excellent potency, but without a sulfonamide group as a potential allergen.

In Chapter 6, I applied HIV-1 strategies to HTLV-1 protease, a potential drug target. I identified DRV as a viable scaffold and evaluated a small panel of DRV analogues, leading to a lead compound with low-nanomolar potency and potential for further optimization.

Lastly, in Chapter 7, I go into detail about a crystallographic method where I discovered an alternative method to refine hexagonal HIV-1 protease cocrystals resulting in better statistics.

Together these studies highlight the complexity of drug resistance and drug design, but examples such as HTLV-1 protease show that we can quickly apply these methodologies toward the development of novel therapeutics against other rapidly evolving drug targets.

2 CHAPTER 2:

Structural Adaptation of Darunavir Analogs Against Primary Mutations in HIV-1 Protease

2.1 Abstract

HIV-1 protease is one of the prime targets of agents used in antiretroviral therapy against HIV. However, under selective pressure of protease inhibitors, primary mutations at the active site weaken inhibitor binding to confer resistance. Darunavir (DRV) is the most potent HIV-1 protease inhibitor in clinic; resistance is limited, as DRV fits well within the substrate envelope. Nevertheless, resistance is observed due to hydrophobic changes at residues including I50, V82 and I84 that line the S1/S1′ pocket within the active site. Through enzyme inhibition assays and a series of 12 crystal structures, we interrogated susceptibility of DRV and two potent analogs to primary S1′ mutations. The analogs had modifications at the hydrophobic P1′ moiety compared to DRV to better occupy the unexploited space in the S1′ pocket where the primary mutations were located. Considerable losses of potency were observed against protease variants with I84V and I50V mutations for all three inhibitors. The crystal structures revealed an unexpected conformational change in the flap region of I50V protease bound to the analog with the largest P1′ moiety, indicating interdependency between the S1′ subsite and the flap region. Collective analysis of protease-inhibitor interactions in the crystal structures using principal component analysis was able to distinguish inhibitor identity and relative potency solely based on vdW contacts. Our results reveal the complexity of the interplay between inhibitor P1′ moiety and S1′ mutations, and validate principal component analyses as a useful tool for distinguishing resistance and inhibitor potency.

2.2 Introduction

Human Immunodeficiency Virus (HIV) infects roughly 37 million people globally, with over 2 million new infections and over 1 million AIDS-related deaths occurring each year[71]. After 30 years of research, advances in antiretroviral therapies (ARTs), combinations of small molecule inhibitors, greatly extended life expectancy for those who receive treatment[25-26]. ARTs inhibit critical proteins necessary for viral replication and maturation, most commonly targeting HIV-1 reverse transcriptase and protease. Although ARTs are highly effective in most patients, the current therapies are still an imperfect solution to a complex problem, as HIV-1 can evolve to confer drug resistance through accumulation of mutations[9]. Primary resistance mutations, typically occurring proximal to the active site where the inhibitor binds, are selected early under selective pressure of inhibition and directly affect inhibitor binding and allow the accumulation of additional mutations. Secondary mutations can occur distal to the active site but still indirectly affect substrate processing or inhibitor binding[30, 72]. No HIV-1 inhibitor is resistance-proof, and modifying an inhibitor to better tolerate primary resistance mutations may help to prevent the accumulation of additional mutations[73].

HIV-1 protease, a 99-amino acid, homodimeric, aspartyl protease[18, 20] **[Figure 2.1a]**, is essential for viral replication and maturation, making this enzyme an ideal drug target[41-42]. The protease processes twelve unique cleavage sites on viral Gag and Gag-Pol polyproteins to release individual viral proteins required for viral replication and maturation. While the cleavage sites share low amino acid

sequence identity, when bound to HIV-1 protease they occupy a consensus volume termed the substrate envelope[43]. We have previously shown that inhibitors that fit within this volume are less prone to resistance as the protease cannot mutate to reduce inhibitor binding without compromising affinity for natural substrates[21].

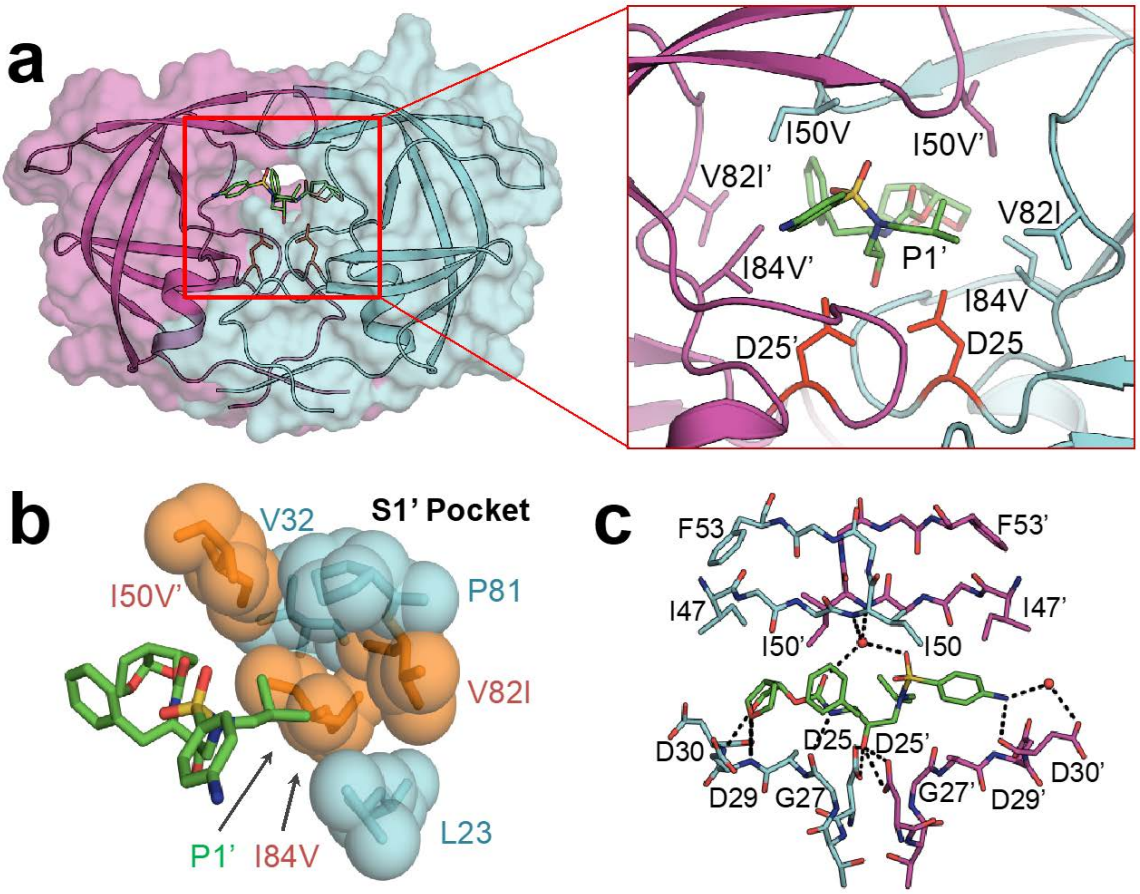


Figure 2.1 a) Co-crystal structure of DRV (green sticks) bound to WT HIV-1 protease. Chain A (cyan) and chain B (magenta) are shown as a cartoon with a transparent surface. D25/D25' catalytic aspartates (red) are displayed as sticks. a-insert) Residues that contribute to primary drug resistance, shown as sticks. b) Spherical representation of residues that make up the S1′ subsite / P1′ pocket. Variable residues are shown in orange. c) Hydrogen bonds (black dashes) between DRV and WT HIV-1 protease. Coordinated waters are shown as red spheres.

The most potent FDA approved protease inhibitor, darunavir (DRV) **[Figure 2.2a]**, fits well inside the substrate envelope, yet resistance still occurs due to accumulation of mutations in HIV-1 protease in patient isolates[31, 38]. DRV is a peptidomimetic inhibitor with four major chemical moieties, denoted as P2, P1, P1′ and P2′ **[Figure 2.2]**. Highly mutated clinical isolates with 18 or more mutations in the protease have been identified, and DRV binding to these resistant proteases studied enzymatically and structurally[31-36]. Most of the patient-derived sequences bear a constellation of secondary mutations as well as primary mutations at the active site that confer DRV resistance, notably including I50V and I84V[15-16, 74-75]. Mutations at I50 are often selected together with A71V mutation, which is distal from the active site but compensates for the loss of enzymatic fitness[76]. Thermodynamics and structural studies of DRV binding to I50V/L and A71V mutations have revealed significant loss of van der Waals (vdW) contacts between the inhibitor and protease underlying loss of binding affinity[77].

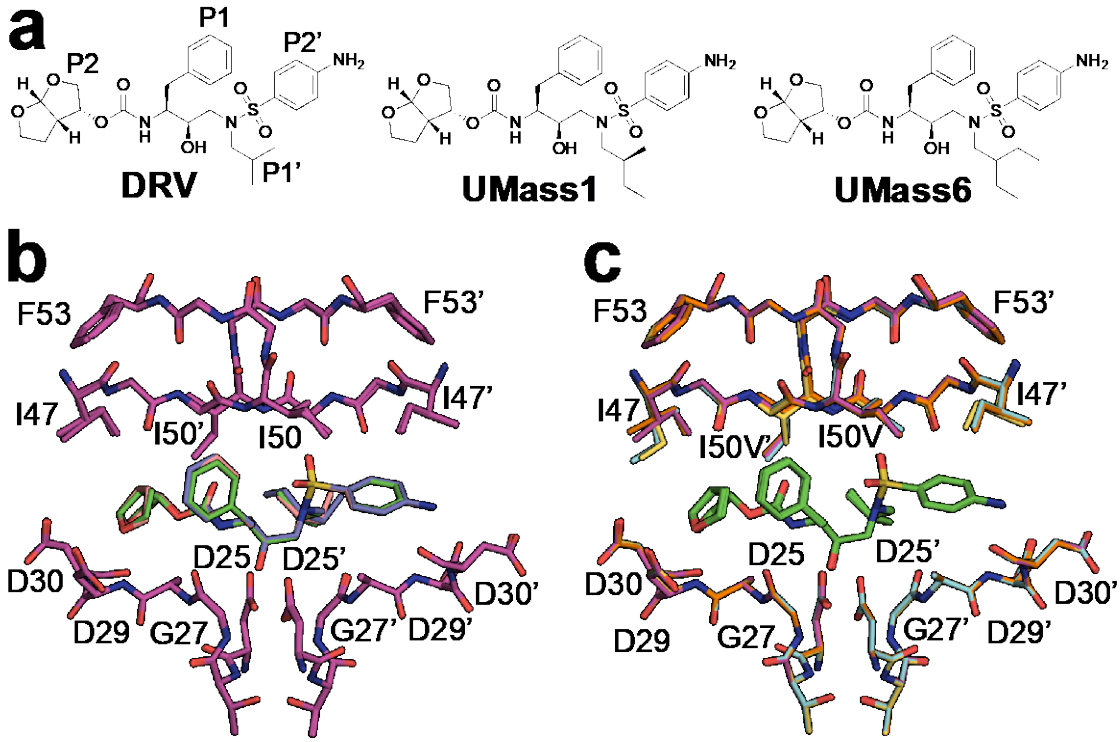


Figure 2.2 a) 2D chemical structure of DRV (with peptidomimetic moieties labeled) and the two DRV analogs (UMass1 and UMass6) with modifications at the P1′ moiety. b) WT protease (magenta) inhibited by DRV (green), UMass1 (pink), and UMass6 (purple) c) DRV (green) bound to WT (magenta), I84V (cyan), V82I (orange), and I50V (gold) protease variants.

DRV and analogs have been and continue to be the subject of chemical, viral, structural and dynamics studies[78]. Modifications to the bis-THF P2 moiety of DRV, including fluorine decorations, expansion to tris-THF and fusion into a tri-cyclic group[33, 79-80] and additional modifications of the P1 and P2' moiety have been shown to improve potency and resistance profiles[81-83]. Our computational studies suggested that P1′ and P2′ moieties of DRV could be further optimized[84] while still staying within the substrate envelope. Accordingly, we previously designed, synthesized and evaluated a panel of 10 DRV analogs with varying P1′ and P2′ moieties, and demonstrated that substrate envelope-guided design can achieve improved inhibition and potency against resistant variants[45]. These inhibitors, named UMass1–10, have subsequently served us as tools to probe subsite interdependency[55], water structure[85], and structural changes in solution by NMR[86]. Among these analogs, UMass1 and UMass6 share the same P2' moiety with DRV but have larger hydrophobic groups at the P1′ position **[Figure 2.2a]** leveraging the unexploited space in the S1′ pocket of the substrate envelope. Thus this set of 3 inhibitors, varying only by the size of the hydrophobic P1' moiety, is ideal to probe the interplay between inhibitor modifications and protease mutations in the S1′ subsite. The underlying mechanism of how P1′ modifications may affect inhibitor response to resistance mutations surrounding the moiety had not been investigated.

In this work, we investigate the interdependency of potency loss for HIV-1 protease inhibitors that differ at the P1′ moiety (DRV, UMass1, and UMass6)

[Figure 2.2a], with WT and 3 variants of HIV-1 protease bearing mutations at residues in the S1/S1′ pocket. The hydrophobic residues I84, V82 and I50 that form the S1′ pocket where the P1′ moiety binds **[Figure 2.1]** (as well as the S1 pocket, as the protease is a homodimer) are all highly variable, and mutations are implicated in many instances of drug resistance, commonly mutating to I84V, V82I, and I50V[38, 87-88] in multi-mutant protease variants. Enzyme inhibition assays were performed and a complete set of 12 crystal structures were determined for the inhibitor–protease combinations. Principal component analysis (PCA) was applied for a collective analysis of the determined structures, which was able to distinguish inhibitor potency and identity based solely on the intermolecular van der Waals contacts. Based on our results, we expect similar applications of PCA can be extended to deconvolute the major determinants underlying resistance/potency in other systems, and to develop predictive models to assess inhibitor potency in drug design.

2.3 Results

To elucidate the mechanisms of resistance and impact on potency of primary S1′ mutations, enzymatic assays and crystal structural analysis were performed with WT HIV-1 protease and 3 variants. The WT protease had the near-consensus clade B sequence of the NL4-3 strain. All three protease variants had a single mutation that altered the shape of the hydrophobic S1/S1′ pocket of HIV-1 protease: I50V, V82I and I84V. Three inhibitors, DRV with an isobutyl P1′ moiety and 2

analogs in which the P1′ moiety was extended to better fill the substrate envelope, UMass1 with a methylbutyl and UMass6 with an ethlybutyl moiety,[45] were analyzed.

2.3.1 Enzymatic Activity of Protease Variants

The enzymatic activity of WT HIV-1 protease and chosen variants (I84V, V82I and I50V) were tested using a natural substrate sequence (MA/CA)[89] **[Table 2.1]**. WT NL4-3 HIV-1 protease had a Michaelis-Menten constant, K_m, of 71 ± 7 μM for cleaving this substrate, which was similar to that of I84V and V82I variants (66 ± 4 and 62 ± 4 μM, respectively). The primary resistance mutation I50V is known to reduce catalytic activity,[77] and while I50V is still catalytically active, the K_m value is beyond the limit of detection for this assay. A71V, a compensatory mutation that is far from the active site and almost always observed with I50V, restores the functionality to WT level (K_m = 73 ± 9 μM), as previously reported.[77] A time-course gel shift assay confirmed the catalytic activity of I50V single mutant, and the rescued activity of I50V/A71V variant in cleaving purified Gag polyprotein **[Figure 2.3]**. The sustained catalytic activity of proteases across all variants indicates these primary mutations can indeed appear early in drug resistance pathways without compromising substrate cleavage.

Table 2.1 K_m of HIV-1 protease variants measured using a natural substrate sequence.

	WT	I84V	V82I	I50V	I50V/A71V
K_m (μM)	71.4 ± 6.8	66.4 ± 4.3	61.7 ± 4.4	Undefined	73.2 ± 9.1

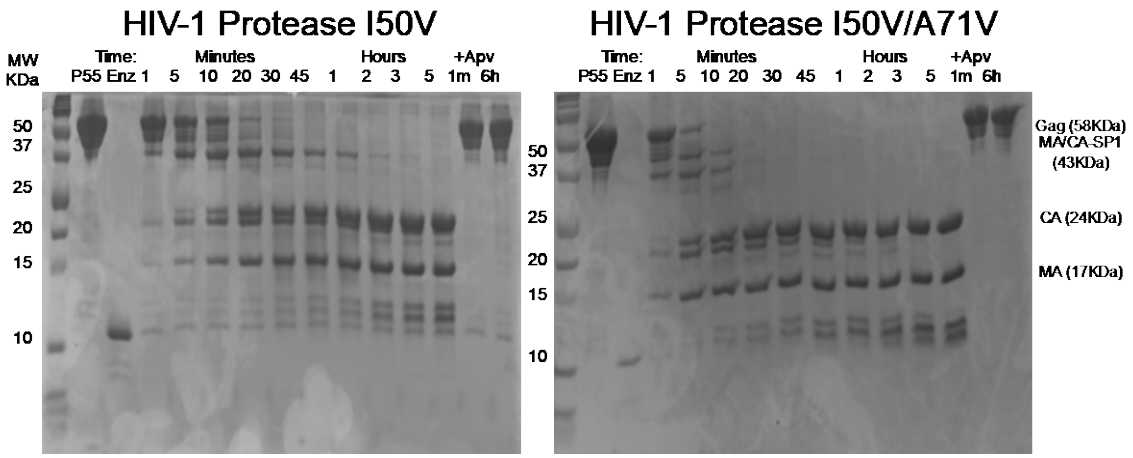


Figure 2.3 Time course assay of I50V and I50V/A71V cleaving HIV-1 Gag polyprotein.

2.3.2 Effect of P1' Modifications on Inhibition of Variants with Active Site Mutations

The enzyme inhibition constant (K_i) was measured to determine the potency of inhibitors against each HIV-1 protease variant **[Table 2.2]** using an optimized assay[90]. The optimized assay enabled K_i determination with smaller errors and lower enzyme concentrations, and gave results that were overall comparable to previously published values[45]. Although the assay is optimized, the lower limit of detection is about 5 pM and all 3 inhibitors were too potent to obtain a reliable value against WT protease. The V82I mutation did not confer any measurable resistance for the inhibitors, with the K_i staying below 5 pM. The I84V mutation caused a reduction in potency, with the K_i increasing to around 25 pM for DRV and UMass1, and approximately half of that for UMass6. Thus, the inhibitor with the largest P1′ moiety performed 2-fold better than the other inhibitors against I84V variant. The I50V mutation was more deleterious with K_i values increasing by two orders of magnitude relative to WT protease. Interestingly, UMass6 performed slightly worse than DRV and UMass1 (K_i = 146 ±11 versus 117 ± 6 and 131 ± 8 pM, respectively). Although both I84V and I50V mutations cause the same reduction in side chain size, these two mutations had the opposite effect on the change in potency of UMass6 compared to DRV and UMass1.

Table 2.2 Enzyme inhibition constant, K_i (pM), of inhibitors against WT protease and variants with primary mutations, measured using an optimized enzymatic assay.

Inhibitor	WT	I84V	V82I	I50V	I50V/A71V
DRV	< 5	25.6 ± 5.6	< 5	117.2 ± 5.8	74.5 ± 5.6
UMass1	< 5	26.1 ± 3.7	< 5	131.3 ± 8.2	110.3 ± 8.8
UMass6	< 5	12.8 ± 3.1	< 5	146.2 ± 10.7	100.0 ± 9.9

2.3.3 Structural Rearrangements Underlying DRV Susceptibility to Primary Resistance Mutations

To determine the structural basis of observed potency changes due to primary resistance mutations around the P1′ moiety, each protease variant was co-crystalized with each inhibitor resulting in a set of 12 crystal structures **[Table 2.3]**. All structures were solved in the same space group with a resolution of 2 Å or better and in the same NL4-3 background, affording detailed investigation of atomic interactions (our previous structures of these inhibitors[45] were determined with SF-2 protease). We determined the crystal structures of DRV, UMass1 and UMass6 bound to WT HIV-1 protease of NL4-3 strain, which was also used in the enzymatic assays above.

Table 2.3 X-Ray Crystallography Statistics

Protein	NL4-3	NL4-3	NL4-3	NL4-3	NL4-3	NL4-3	NL4-3	NL4-3	NL4-3	NL4-3	NL4-3	NL4-3
Mutation	WT	WT	WT	I84V	I84V	I84V	V82I	V82I	V82I	I50V	I50V	I50V
Inhibitor	DRV	UMass1	UMass6	DRV	UMass1	UMass6	DRV	UMass1	UMass6	DRV	UMass1	UMass6
PDB ID	6DGX	6DGY	6DGZ	6DH0	6DH1	6DH2	6DH3	6DH4	6DH5	6DH6	6DH7	6DH8
Resolution (A)	2.00	1.95	2.00	1.90	1.97	1.98	1.91	1.94	2.00	1.97	2.00	1.95
Space Group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
a (A)	51.0	51.2	51.2	51.1	51.3	51.3	51.3	51.3	51.4	51.1	51.1	51.8
b (A)	58.3	58.0	58.2	58.1	57.9	57.8	58.0	57.8	57.9	58.0	57.8	56.7
c (A)	61.9	62.2	62.0	61.8	62.1	62.1	62.2	62.2	62.3	62.1	62.2	61.5
Completeness	98.1	98.0	94.1	99.9	94.6	95.0	96.9	95.6	96.8	98.5	98.5	99.4
Tot. Reflections	84332	89641	55311	181306	73682	92688	68517	69346	88129	141343	58534	95807
Uniq. Reflect.	12758	13687	12406	15103	12875	12775	14540	13520	12472	13416	12857	13663
Avg I/Sig	17.5	12.6	16.0	12.4	8.9	12.6	13.6	14.1	16.9	11.3	19.8	14.8
Redundancy	6.6	6.5	4.5	12.0	5.7	7.3	4.7	5.1	7.1	10.5	4.6	7.0
R-Merge (Lin)	0.057	0.095	0.059	0.057	0.071	0.067	0.063	0.065	0.093	0.068	0.076	0.095
RMSD Bonds*	0.003	0.003	0.005	0.003	0.003	0.004	0.003	0.002	0.002	0.003	0.008	0.004
RMSD Angles*	0.590	0.651	0.752	0.710	0.729	0.757	0.683	0.551	0.577	0.690	0.952	0.648
R-Free* R-Work*	22.8 19.2	22.8 19.4	22.8 18.6	21.2 18.3	23.0 19.2	23.4 19.1	21.0 17.4	22.9 18.5	22.1 18.4	22.2 17.5	21.2 16.8	20.1 16.4

*Based on Phenix Program

When DRV and other peptidomimetic inhibitors bind to HIV-1 protease, the hydroxyl moiety is centered between the catalytic aspartates (D25/D25′) and the inhibitor makes a number of additional hydrogen bonding interactions with the backbone nitrogen and oxygen atoms of residues that line the active site **[Figure 2.1c]**. DRV also makes a key water-mediated interaction with the backbone nitrogen of residue 50, located at the tip of each flap. DRV's P1 and P1′ moieties are hydrophobic and interact with hydrophobic residues/pockets in the hinge of the flaps. Inhibitor potency diminishes if any of these polar or non-polar interactions are perturbed.

DRV-bound crystal structures were compared to reveal any structural rearrangements in response to single mutations at the active site. Overall, the structures were highly similar **[Figure 2.2]**. In agreement with the maintained potency against V82I variant, the hydrogen bonding and packing of DRV at the active site was conserved in the V82I crystal structure **[Figure 2.2b and Table 2.4]**. The additional steric bulk of the isoleucine side chain is not directed towards the inhibitor but instead is solvent exposed, and does not perturb inhibitor binding. As in the V82I variant, the binding mode and hydrogen bonds of DRV were not altered in I84V and I50V structures **[Figure 2.2b]**. However, van der Waals (vdW) interactions with residues 84 and 84' decreased due to I84V mutation, which was in part compensated by increased packing against I47 and V82′ **[Table 2.4]**. In the I84V variant, residues V32′ and L23′ in chain B underwent a side chain conformer change relative to WT protease **[Figure 2.4].** In addition, residue I47 was found in

an alternative rotamer in chain A, causing increased inhibitor interactions with that residue. This asymmetric compensation indicates subtle alterations in the overall packing of the inhibitor at the active site, which reduced the effect of I84V mutation on potency.

A similar repacking was also observed in the I50V–DRV structure, where increased packing against I47 in both chains compensated lost vdW interactions at residue 50. Our previous computational analysis had indicated I47 as a major modulator of DRV-protease interactions[54]. Thus, the structures explain why the V82I single mutation did not confer resistance to DRV, and in agreement with previous studies reveal the rearrangements in vdW packing around DRV underlying susceptibility to I50V[77, 91]. Importantly, intermolecular interactions are not solely altered at the site of mutation but throughout the active site, urging a more detailed structural analysis.

Table 2.4 Change in intermolecular van der Waals (vdW) interactions between inhibitor and protease active site residues relative to WT complex. Only residues with changes greater than 0.40 kcal/mol are listed. Primed residue numbers correspond to chain B. Decrease and increase in per residue contacts with respect to WT protease are colored blue to red, respectively.

Resi #	I84V DRV	I84V UMass1	I84V UMass6	V82I DRV	V82I UMass1	V82I UMass6	I50V DRV	I50V UMass1	I50V UMass6
30	0.28	-0.14	-0.83	0.46	0.05	-0.44	0.00	-0.01	-0.48
47	1.22	1.07	1.12	-0.02	-0.04	1.12	1.19	-0.08	-0.18
48	0.49	0.10	0.22	0.14	0.13	0.26	0.34	0.05	-0.18
49	0.33	-0.02	0.08	0.06	0.10	0.25	0.19	0.27	-0.44
50	-0.50	-0.60	-0.53	-0.11	-0.04	0.13	-1.70	-1.32	-1.94
84	-1.17	-1.10	-1.17	-0.12	0.15	-0.01	-0.05	0.27	-0.31
23'	-0.60	-0.33	-0.25	-0.06	-0.17	-0.11	0.10	-0.09	-0.07
25'	-0.41	-0.31	-0.37	0.10	0.01	-0.11	-0.08	0.04	-0.30
28'	-0.55	-0.34	-0.28	0.06	0.29	0.24	0.02	0.40	-0.47
47'	-0.36	-0.47	-0.42	0.20	0.04	-0.02	0.49	0.09	0.41
48'	-0.16	-0.03	0.03	-0.01	0.16	0.19	0.15	0.15	0.44
49'	-0.09	-0.07	0.07	-0.03	0.09	0.19	0.18	0.16	0.61
50'	-0.08	0.16	-0.03	-0.09	0.11	0.03	-0.84	-0.44	0.17
82'	1.17	1.03	0.83	0.23	0.41	0.27	0.10	0.15	0.21
84'	-1.82	-1.61	-1.54	-0.03	-0.03	-0.02	0.03	-0.23	-0.21

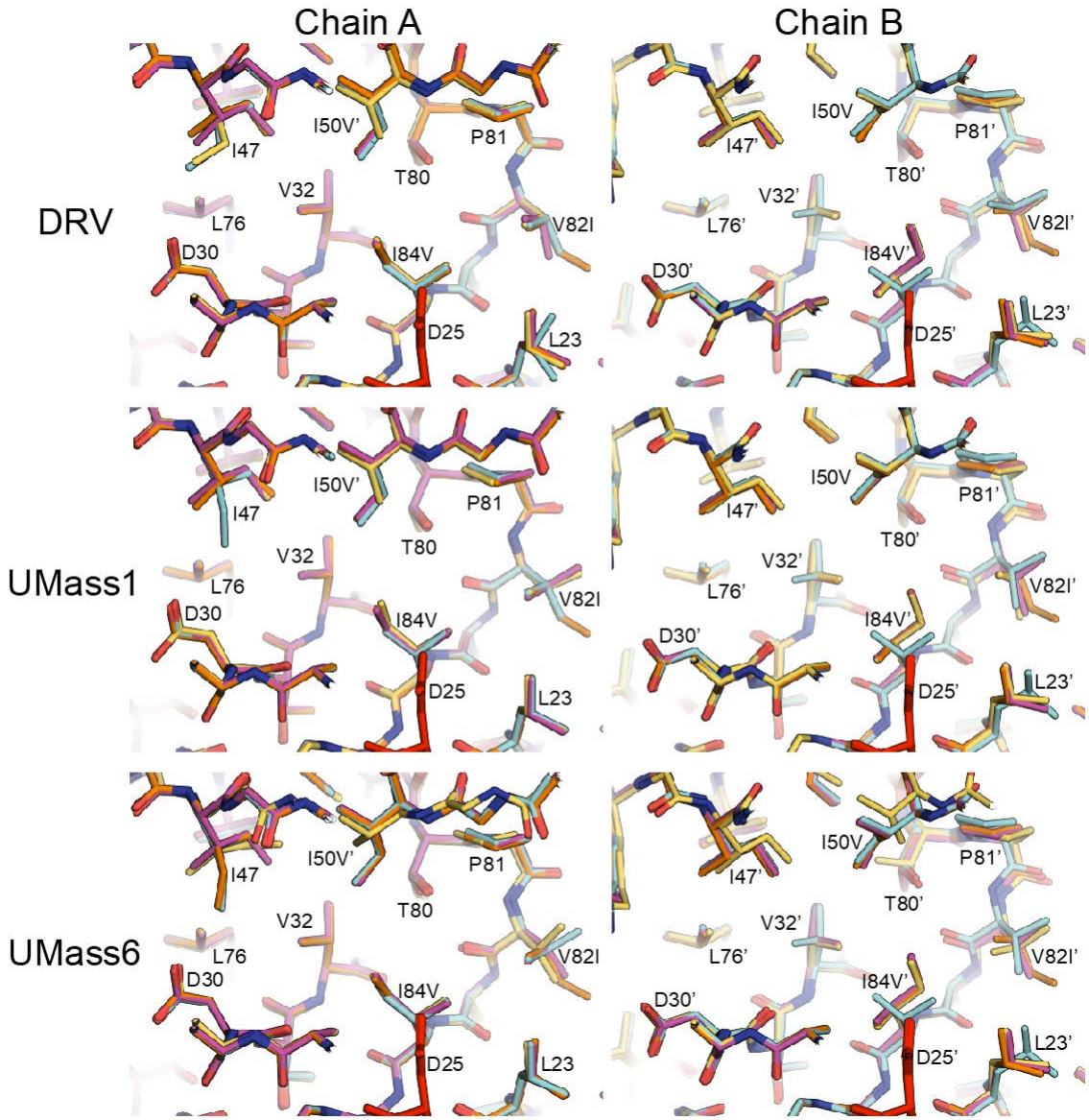


Figure 2.4 Side chains at the protease active site surrounding the bound inhibitor (not shown) in the crystal structures. Alignment of protease active site residues when (Top) DRV, (Middle) UMass1, and (Bottom) UMass6 is bound to WT (magenta), I84V (cyan), V82I (orange), and I50V (gold) protease variants.

2.3.4 Structural Response to Modifying P1′ Moiety Against Primary Mutations

The larger P1′ moieties of UMass1 and UMass6 were designed to improve hydrophobic packing and increase vdW contacts with the protease while still staying within the substrate envelope[45]. Accordingly, the total vdW energy of intermolecular interactions between the inhibitor and protease increased by 1.6 and 4.7 kcal/mol, respectively for UMass1 and UMass6, as the P1′ moiety increased in size **[Table 2.5]**. However, the simple measure of total vdW interaction energy is not sufficient to explain the overall K_i value trends, suggesting a more comprehensive residue-based analysis may be needed to capture inhibitor potency. In the V82I structure, while DRV did not gain significant interactions compared to WT complex, UMass1 and UMass6 gained approximately 2 kcal/mol in vdW energy as their P1′ moieties directly interact, but not clash, with the sterically larger side chain. Thus, although the K_i values are below the limit of detection, structural analysis suggests that V82I mutation might confer hyper-susceptibility to UMass1 and UMass6 inhibitors; thus unlike with DRV, the V82I mutation is not likely to be selected under the pressure of these inhibitors.

Table 2.5 Total van der Waals interactions (kcal/mol) between inhibitor and protease in crystal structures.

Inhibitor \ Protease	WT	I84V	V82I	I50V
DRV	**-83.1**	**-80.6**	**-83.5**	**-82.9**
UMass1	**-84.7**	**-82.5**	**-86.6**	**-85.7**
UMass6	**-87.8**	**-84.7**	**-90.6**	**-85.0**

The co-crystal structures of UMass1 bound to WT protease and variants were very similar to those of DRV **[Figure 2.2c]**, albeit with subtle alterations. Against I84V and I50V mutations, which are located directly at the pocket where P1′ moiety binds, UMass1 experienced a similar reduction in potency as DRV **[Table 2.2]**. The UMass1–I84V structure displayed the same phenomenon of asymmetric inhibitor repacking as with DRV **[Table 2.4]**. The additional methyl group in UMass1 is oriented away from the steric space provided by the I84V mutation but still makes additional contacts with P81 and V82. When UMass1 bound I50V, there was a reduction in vdW interactions at residue 50 and 50′ as seen for DRV, but no compensation at residue 47. Instead, repacking of UMass1 increased against G49 and I84 in chain A, and residues 28–30 in chain B. However, this alternate compensation was more distributed and subtle, which may underlie the slightly worse K_i of UMass1 compared to DRV against I50V protease. Thus, the crystal structures indicate that although the potency loss against I50V is similar for DRV and UMass1, the underlying structural changes and repacking at the active site are distinct.

2.3.5 I50V Mutation Induces Conformational Changes in Both UMass6 and Protease Flaps

UMass6 has an even bulkier P1′ moiety than UMass1 and binds WT protease similar to DRV and UMass1 **[Figure 2.2c]** but with enhanced overall vdW interactions (-87.8 kcal/mol compared to -83.1 and -84.7, respectively; **Table 2.5**). Against I84V mutation, enhanced packing at the P1′ moiety of UMass6 resulted in

overall better interactions and potency. In UMass6, the methyl groups on both branches of the P1′ moiety help UMass6 to better accommodate the steric space provided by the I84V mutation, leading to a 2-fold improvement in K_i compared to the other two inhibitors.

Contrary to the case of I84V mutation, UMass6 bound to the I50V variant with the lowest potency. Unexpectedly, the P1′ moiety of UMass6 in I50V protease structure adopted a completely different conformation compared to the other 11 structures **[Figure 2.5]**. In addition to this inhibitor's unique binding conformation, the flaps in I50V protease underwent rather large-scale changes: In chain A, I47 assumed a rare rotamer to maintain interaction with the sterically smaller I50V **[Figure 2.4]**. In chain B, T80 underwent a conformational change which may affect flap-tip curling and flexibility[92,93]. This conformational change caused the flaps to bow outward **[Figure 2.5c]** away from UMass6, further reducing vdW contacts. Interestingly, the resulting vdW losses of UMass6 with residue 50 due to the I50V mutation were completely asymmetric, with a loss similar to DRV and UMass1 in chain A but almost no loss in chain B **[Table 2.4]**. However, there was no compensation at residue 47 interactions in either chain, due to the different conformer of I47, which resulted in a greater overall loss in vdW interactions, in agreement with the greatest potency loss observed for UMass6 against I50V mutation.

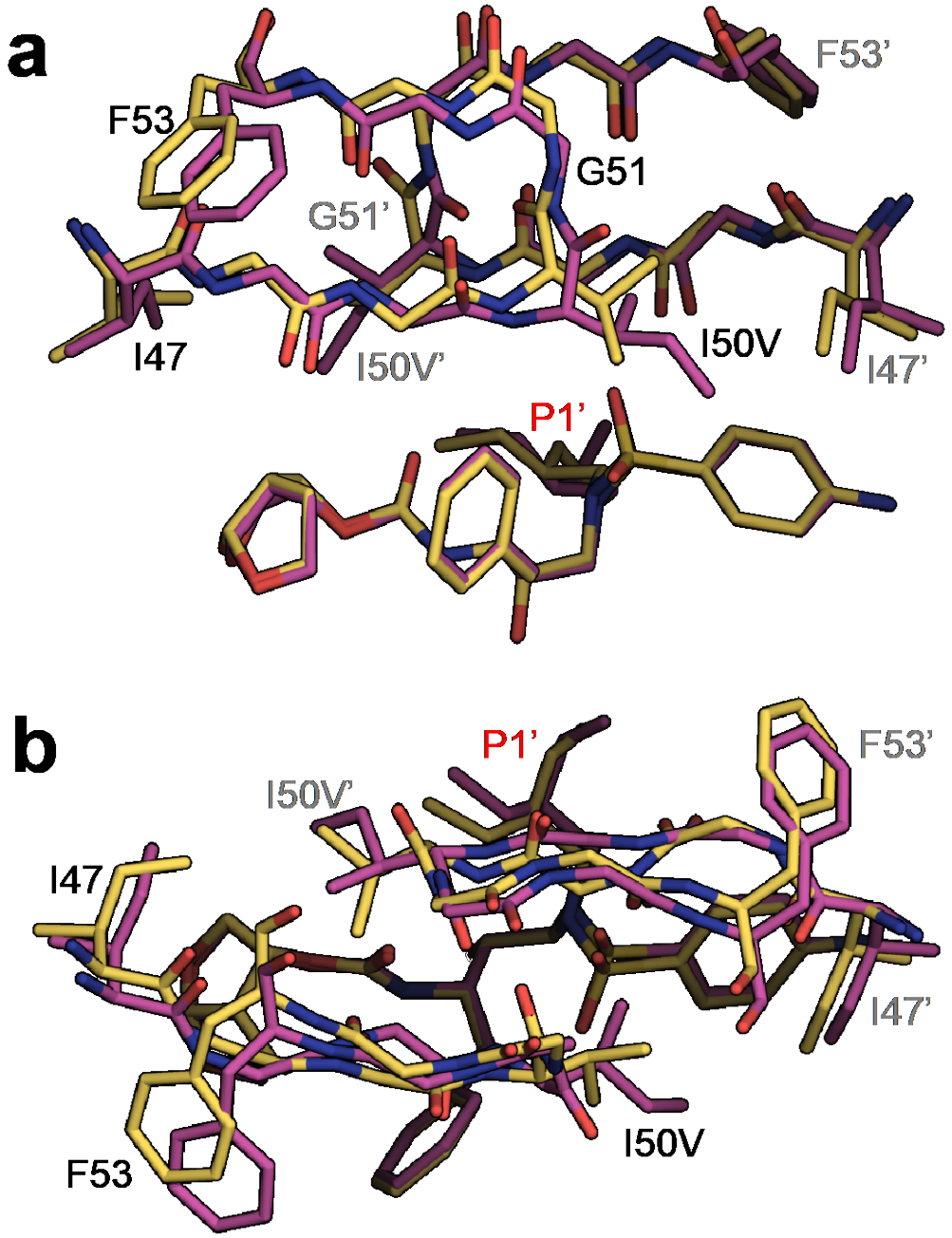


Figure 2.5 The P1′ moiety of UMass6 bound to the I50V variant and the protease flaps exist in an alternate conformation. a) Superposition of WT-UMass6 (magenta) and I50V-UMass6 (gold) structures, displaying the inhibitor and the flap region. Residues 47-53 in both monomers are shown as sticks. b) The same superposition with 90° rotation to show the top view of the flap region.

Although the position of most of the flap residues were altered in the UMass6–I50V structure relative to WT complex, the typical hydrogen bond distances were not significantly perturbed **[Figure 2.6, Table 2.6]**. However, the flipping of the flaps caused a noticeable shift in the location of the so-called flap water **[Figure 2.6]**, which is crucial for flap stabilization[94], and which in turn caused the sulfonyl group to shift toward the flaps to maintain H-bond distances. The shift of the sulfonyl group fused to the aniline P2′ moiety weakened the bond formed by the conjugated water interacting with the side chain of D30 by making the overall distance longer. Weakening of inhibitor hydrogen bonding with D29-D30 has been reported to be coupled with increased flap flexibility in drug resistance due to I50V[95].

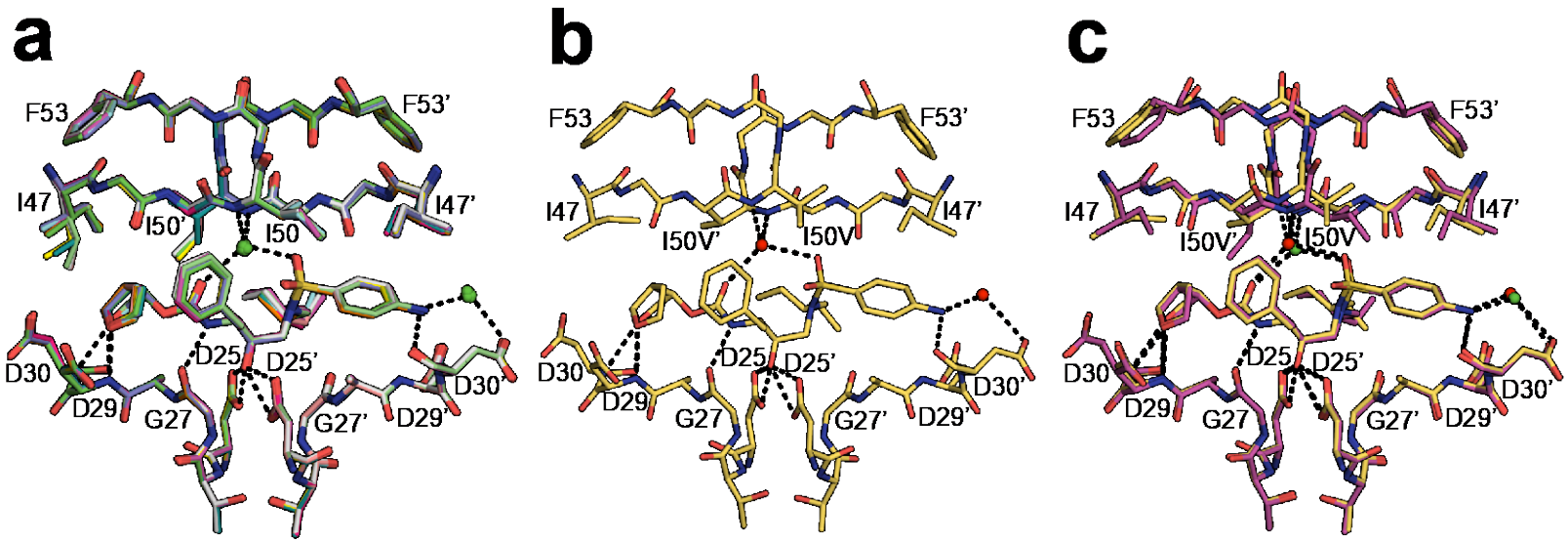
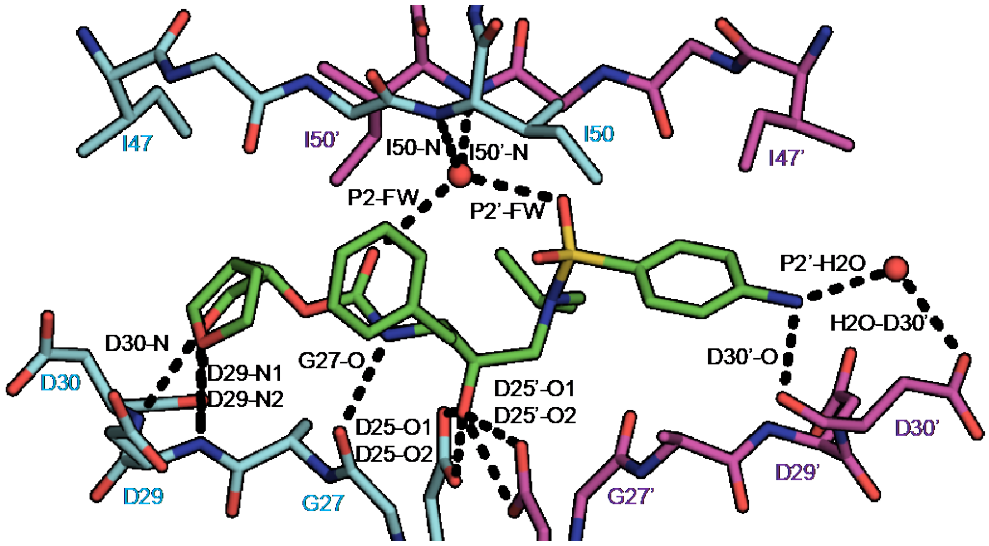


Figure 2.6 a) Hydrogen bonding between inhibitor and protease active site in the 11 of 12 co-crystal structures. Coordinated waters are shown as green spheres. Hydrogen bonds are indicated with black dashes. (b) Hydrogen bonding in I50V-UMass6 crystal structure. Coordinated waters are shown as red spheres. (c) Alignment of WT-UMass6 (magenta) and I50V-UMass6 (gold).



	D30-N	D29-N1	D29-N2	G27-O	D25-O1	D25-O2	D25'-O1	D25'-O2	P2-FW	P2'-FW	I50-N	I50'-N	D30'-O	P2'-H2O	H2O-D30'
WT-DRV	2.9	3.2	2.8	3.0	3.2	2.5	2.9	2.5	3.0	2.8	3.1	2.9	3.1	2.9	2.9
I84V-DRV	2.9	3.1	2.8	3.1	3.3	2.5	3.0	2.6	2.8	2.7	3.0	2.9	3.1	2.8	3.1
V82I-DRV	2.9	3.1	2.8	3.1	3.3	2.5	2.9	2.5	2.8	2.8	3.0	3.0	3.1	2.8	2.9
I50V-DRV	2.9	3.0	2.8	3.1	3.2	2.5	2.9	2.5	2.8	2.7	3.0	3.0	3.0	2.8	3.1
WT-UMass1	3.0	3.1	2.8	3.1	3.2	2.6	2.9	2.5	2.8	2.7	3.0	3.0	3.1	2.8	2.7
I84V-UMass1	3.1	3.2	2.8	3.1	3.3	2.5	2.9	2.4	2.7	2.7	2.9	3.1	3.0	2.8	3.1
V82I-UMass1	3.1	3.2	2.8	3.1	3.2	2.5	2.8	2.5	2.8	2.8	3.0	3.1	2.9	2.8	2.9
I50V-UMass1	3.0	3.1	2.8	3.0	3.3	2.5	2.9	2.5	2.6	2.8	2.9	3.1	2.9	2.8	3.1
WT-UMass6	3.0	3.1	2.8	3.1	3.3	2.6	3.0	2.5	2.8	2.8	3.0	3.0	3.1	2.6	3.0
I84V-UMass6	3.1	3.2	2.8	2.9	3.3	2.5	3.0	2.5	2.8	2.7	2.9	3.1	3.1	2.7	3.2
V82I-UMass6	3.1	3.3	2.7	3.0	3.1	2.5	2.8	2.6	2.9	2.6	2.9	3.1	3.0	2.8	3.2
I50V-UMass6	3.1	3.2	2.8	2.9	3.2	2.6	2.9	2.5	2.8	2.9	3.0	2.9	3.0	2.7	3.4

Table 2.6 Hydrogen bonding distances (in Å) for intermolecular protease-inhibitor hydrogen bonds in the crystal structures. The hydrogen bonds are displayed above as black dashed lines on the WT–DRV crystal structure for reference, where DRV is in green sticks and water molecules are displayed as red spheres.

The structural rearrangements of the protease in the I50V–UMass6 structure were quantified and visualized by distance-difference matrices **[Figure 2.7]**. Distance-difference matrices measure the internal distances between alpha carbon atoms of every residue pair in a structure, then compare each distance to that in a reference structure, which in our case is the inhibitor bound to WT protease. This method detects conformational changes due to a mutation without any structural superposition bias. Comparison of the variant structures to their corresponding WT structures shows the discrepancy between the I50V–UMass6 and all the other 11 structures **[Figure 2.7].** Mapping the distance-differences onto the 3D structure revealed that the structural changes due to I50V in the UMass6-bound protease propagated throughout the protein structure **[Figure 2.8]**.

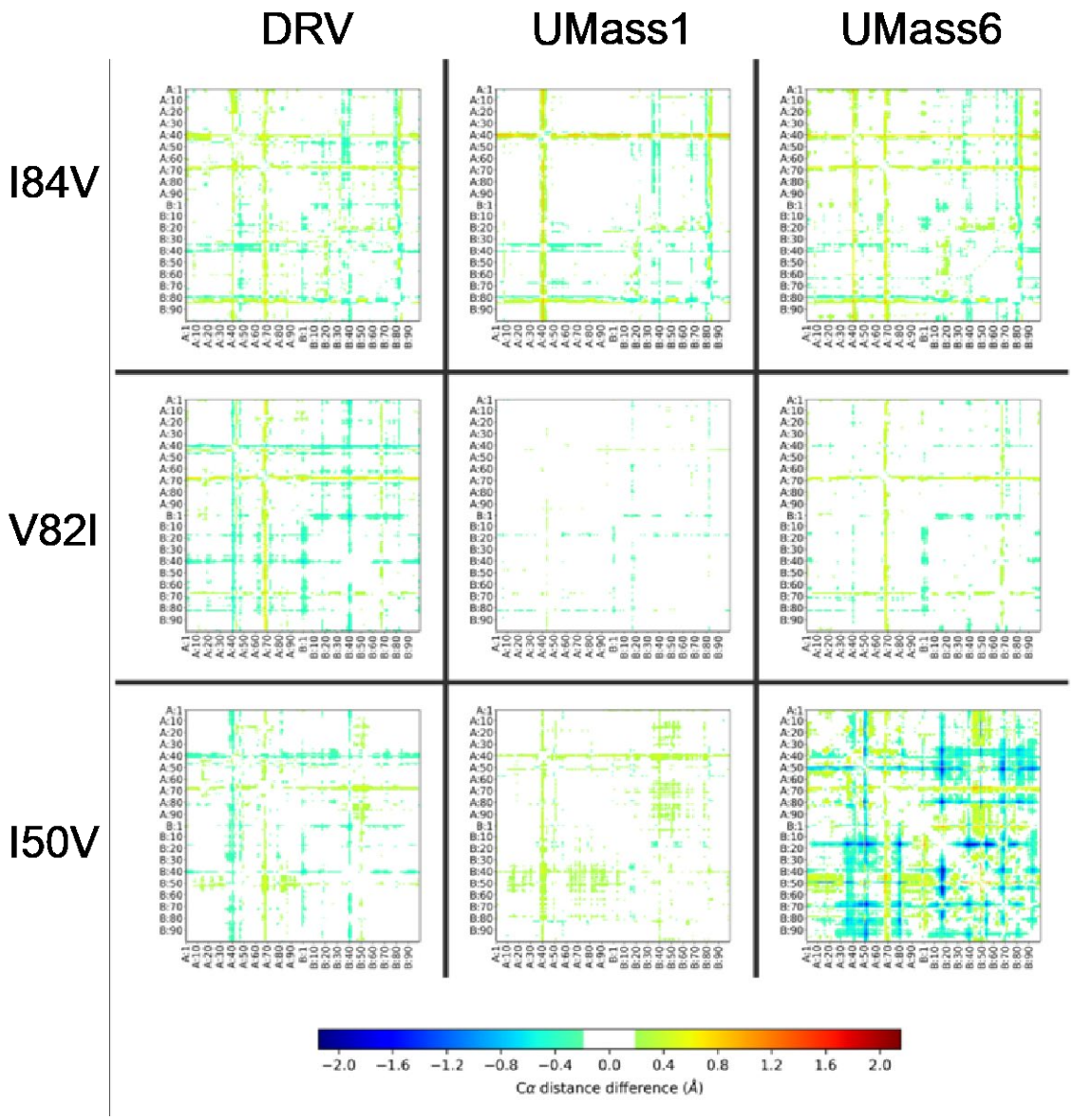


Figure 2.7 Distance-difference matrices comparing HIV-1 protease variant-inhibitor crystal structures to the corresponding WT complex.

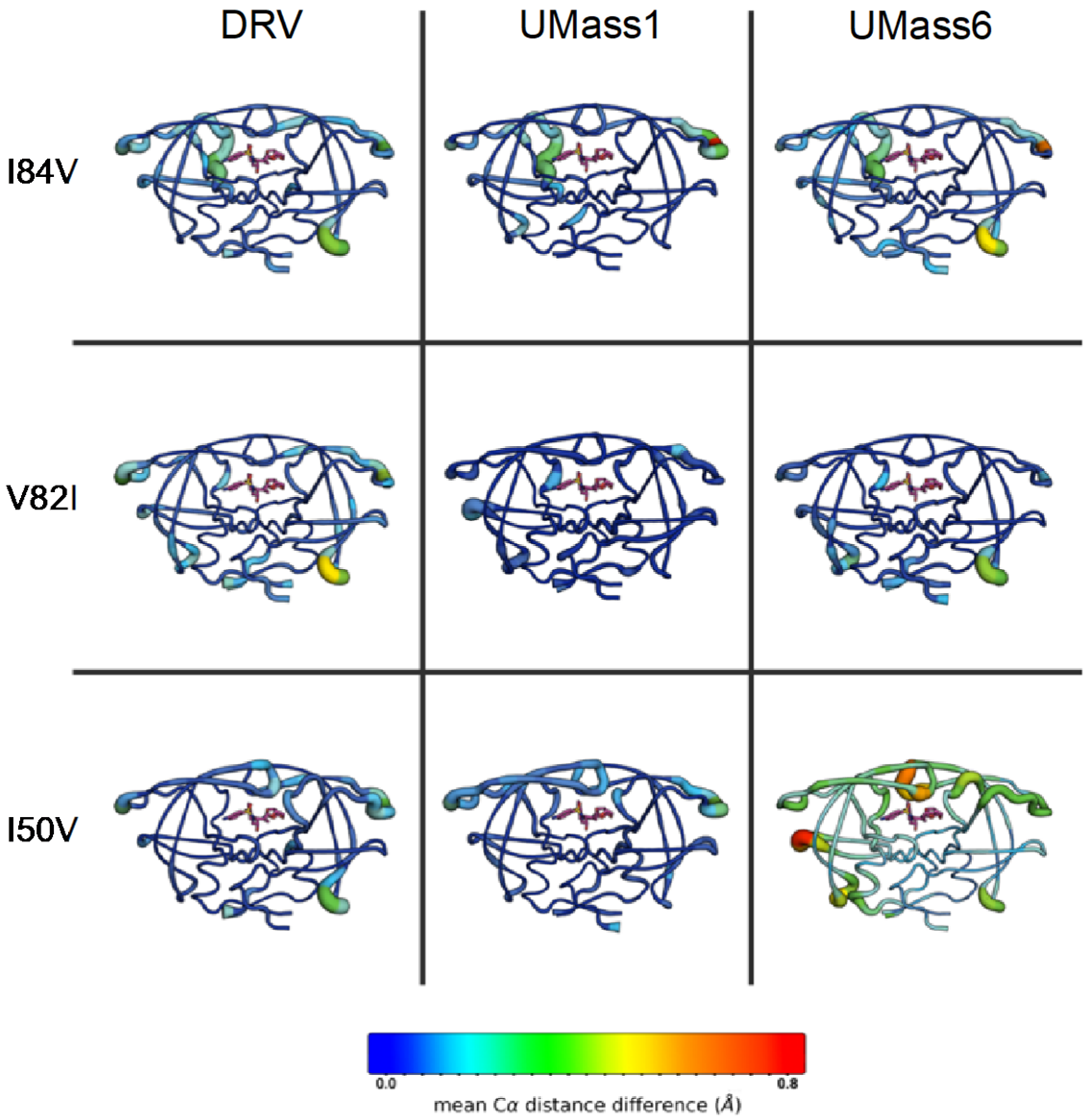


Figure 2.8 Cartoon-putty diagrams depicting the mean changes in C_α distance differences relative to WT protease crystal structure. Tube thickness and warm colors indicate larger differences.

2.3.6 Principal Component Analysis (PCA) of Residue-Specific Interactions

The set of 12 inhibitor–protease crystal structures and the corresponding K_i values enabled a collective analysis of the interplay between inhibitor modifications and protease mutations in the S1/S1′ subsite. First, the intermolecular vdW interaction energy with the inhibitor for each residue in a given structure was calculated for each of the 12 structures to yield a 198x12 matrix. Although this matrix contains rich information on the molecular basis of inhibitor potency, the multidimensionality and complexity of the data precludes deducing what specific properties are responsible for the observed variation. Thus, the matrix of residue-wise vdW contacts was subjected to principal component analysis (PCA) to extract principal components (PC) that best account for the variance in data. The first three principal components (PCs) accounted for 83% of the observed variance, with the first PC (PC1) accounting for approximately half (43%) of the variance **[Figure 2.9].** In addition to reducing the dimensionality of the dataset and, the PCs reveal what causes the variations in inhibitor–protease interactions among these structures.

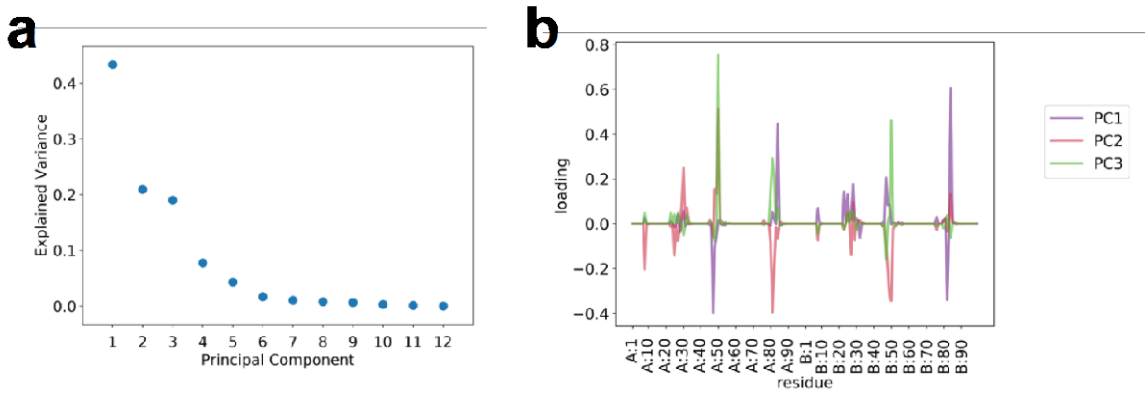


Figure 2.9 Principal component analysis (PCA) of per residue vdW interactions in the 12 crystal structures. (a) Explained variance of the first 12 principal components. **(b)** Loading of the first 3 principal components, showing the contribution of residues to the PCs.

The PC1 separated the structures mainly according to the protease variant, rather than inhibitor identity **[Figure 2.10].** Significantly, structures containing the I84V primary resistance mutation of all 3 inhibitors clustered away from the remaining structures. Inspection of the loading indicated PC1 was largely determined by changes in vdW interactions between the inhibitor and residue 84 **[Figure 2.9]**. This is in agreement with our previous work that showed that mutations at residue 84 significantly alter the pattern of vdW contacts in a panel of drug resistant variants.[96] Additionally, contacts between I47 and the inhibitor contributed significantly to the spread along PC1, which can assume a different side chain conformer and compensate for loss of interactions at residue 84 as explained above.

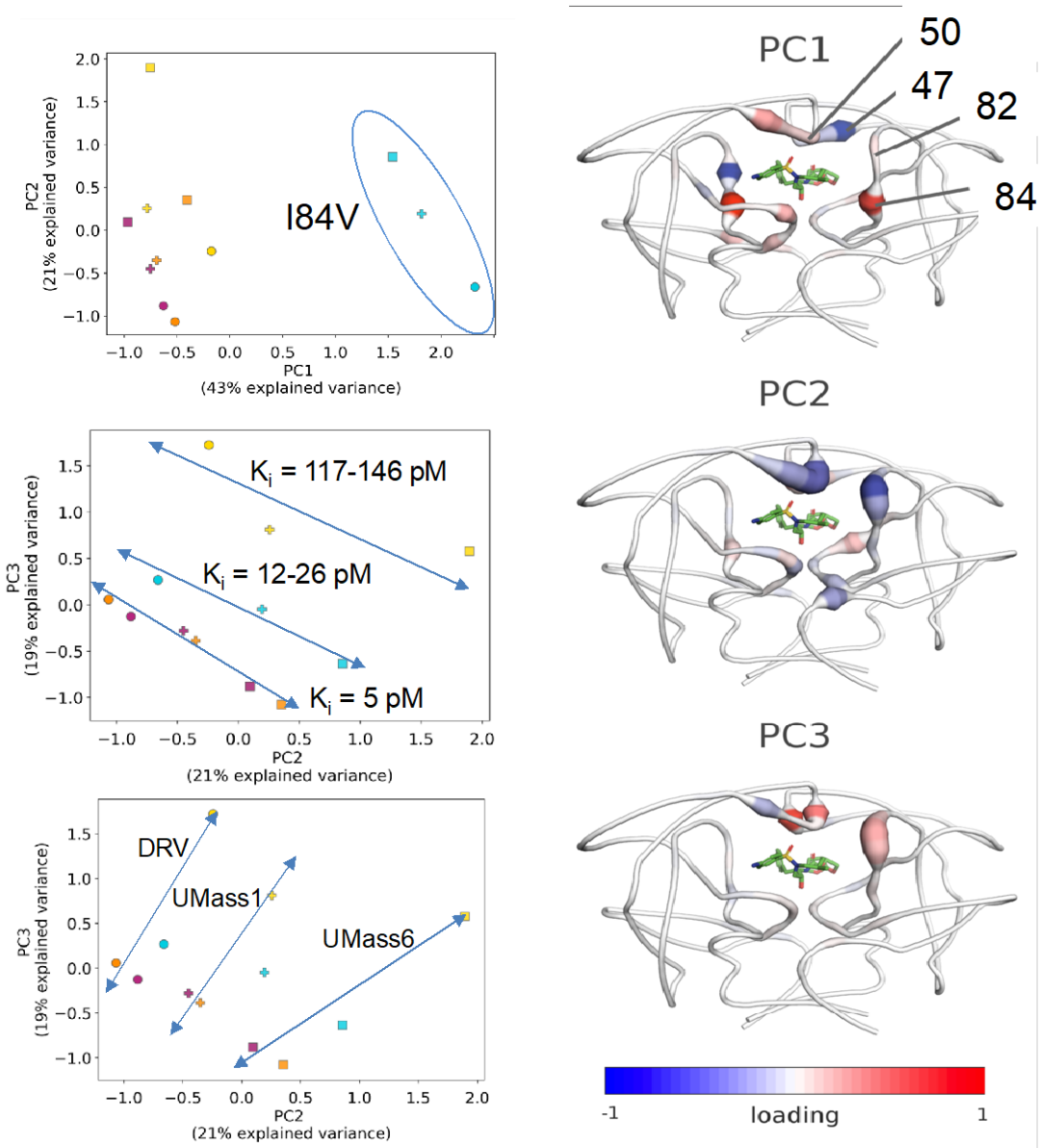


Figure 2.10 Principal component analysis of vdW interactions in the 12 protease-inhibitor complexes. (left) The protease–inhibitor pairs plotted according to first and second principal components, PC1 versus PC2 (upper panel) and PC2 versus PC3 (lower panels). In the plots, circles (DRV), plus signs (UMass1), and squares (UMass 6) indicate different inhibitors and colors (magenta: WT, yellow: I50V, orange: V82I, cyan: I84V) indicate protease variant. The blue oval and straight lines are shown to guide the eye and highlight clusters. (right) The contribution of individual residues to the PCs according to the loading vector are depicted on protease backbone structure using cartoon-putty diagrams, where tube thickness indicates higher weights and warm colors are positive loading and cold colors are negative loading. Residues with the highest weights are labeled on the cartoon-putty for PC1.

The second and third principal components (PC2 and PC3) together accounted for 40% of the observed variance in inhibitor–protease vdW interactions. PC2 separated structures mostly according to inhibitor identity, rather than protease variant **[Figure 2.10]**. Inspection of the loading of PCs showed that variance at residues 84 and 50′ contributed significantly to the ordering of structures along PC2, indicating modification of inhibitor P1′ moiety caused alteration in packing against these residues **[Figure 2.9b]**. The contribution of residue 50 in chain A had the same direction for both PC2 and PC3 whereas the contributions of 84 and 50′ had the opposite direction. Contacts between the inhibitor and residues 84 and 50′ changed significantly in the presence of a small P1’ moiety but were less effected when the P1′ moiety was larger. Overall, the PCA captured the asymmetric compensation of vdW interactions observed in the crystal structures, and quantified the contribution of residue-specific interactions to the overall variance in the dataset.

Strikingly, plotting PC2 and PC3 against each other enabled separating the 12 crystal structures both according to inhibitor identity and potency **[Figure 2.10]**. All 6 inhibitor-protease pairs with <5 pM affinity clustered on a line defined by a linear combination of PC2 and PC3. Intriguingly, the I84V and I50V complexes fell on separate but parallel lines, together defining a set of 3 roughly parallel lines on the PC2-PC3 plane. These 3 lines were ordered according to experimentally measured potency, with the I84V complexes nearer to the high affinity line, followed by the I50V complexes. Additionally, 3 other roughly perpendicular lines

separated the inhibitor-protease pairs according to inhibitor identity. The inhibitors were also ordered, according to the chemical similarity of their P1′ moieties as DRV-UMass1-UMass6. Thus, the collective analysis of the set of 12 crystal structures using PCA enabled clustering inhibitor identity and potency based solely on per residue vdW contacts, indicating this information is sufficient to discern inhibitors.

2.4 Discussion

In this study the interplay between modifications of an inhibitor's functional group and modifications known to confer drug resistance that surround that functional group. Specifically, we have investigated HIV-1 protease inhibitor DRV and two analogs and assessed modifications at P1′ and how these analogs interactions by three single site drug resistant variants. A set of 12 high-resolution crystal structures comprising three analogous inhibitors bound to WT protease and three variants allowed a collective analysis of inhibitor-protease interactions in modulating potency and resistance. The overall intermolecular vdW energy or hydrogen bonding was not sufficient to explain the observed inhibitor potency; thus we employed a more comprehensive and residue-based analysis. The residue-based vdW contact energies in all 12 structures was subjected to PCA, which not only identified the key residues that determined the variance in inhibitor packing but also captured the asymmetry in this variance. Rather than visual inspection or manual comparison, PCA enabled an unbiased quantitative assessment of intermolecular interactions collectively in all the 12 structures.

More significantly, clustering of the structures along major PCs was able to categorize the inhibitor-protease pairs according to inhibitor type and potency. This suggests residue-specific vdW interactions may serve as “fingerprints” to categorize and classify inhibitors bound to protease, particularly for inhibitors similar to those used here. PCA has recently been applied to overall structures of HIV-1 protease deposited in the PDB to cluster and categorize protease conformations[97-98]. As with any mathematical model, the training set (or the input data) needs to be appropriate for the purpose and expanded when possible to improve the model and enable assessing more diverse inhibitors. In addition, analyzing mutations that involve polar or charged residues will likely require analyzing intermolecular electrostatic interactions in addition to vdW contacts. Nevertheless, we found here that certain simple and linear combinations of PCs, derived from complex and multidimensional structural data, might be indicators or even predictors of inhibitor potency.

DRV and the analogs used differ only at the P1’ moiety. The larger, flexible P1′ groups of UMass1 and UMass6 were previously shown to increase potency compared to DRV in cellular assays, specifically against a panel of WT clades and drug resistant variants with many mutations[45]. Our results validated that larger P1′ moieties were effective against I84V and V82I mutations, but an unexpected alternative mechanism of resistance was uncovered against the I50V mutation. The I50V mutation is commonly observed in variants that are resistant to DRV, and has previously been investigated[34, 75]. Unlike DRV and UMass1, the P1’ group of

UMass6 is large enough to be able to contribute a methyl group to interact with the destabilized hydrophobic interactions involving the flap tips. However, this interaction affects the orientation of V50, which to avoid an unfavorable side chain rotamer, induces flipping of the flaps and a rather unforeseen conformational rearrangement. A similar but distinct flap rearrangement was previously observed in response to a coevolution mutation in Gag substrate in the I50V/A71V variant, which enhances vdW interactions with the substrate[40]. Hence the relative flexibility of flaps in this variant might be able to enhance interactions with substrates while weakening inhibitor binding, thus contributing to conferring resistance.

Unlike the other two mutations, V82I did not confer measurable resistance against the inhibitors tested. However, crystal structures suggested that the intermolecular interactions are improved for UMass1 and UMass6 relative to WT protease, unlike for DRV. Thus, the bulkier P1′ groups in UMass inhibitors might cause hyper-susceptibility to V82I and prevent selection of this mutation on resistance pathways. However, groups larger than that in UMass1 may act as a selective pressure forcing I50V to become a dominant variant. Thus, despite high similarity, inhibitors with P1′ modifications may select distinct mutational patterns of resistance.

The analysis on three inhibitor analogs here revealed that the basic idea that a larger P1′ moiety would help the inhibitor to retain better interactions upon shortening of a side chain is not necessarily correct. Rather than a simple alteration in interactions at this subsite, there was an overall and asymmetric rearrangement

of the vdW interactions throughout the active site pocket due to either I84V or I50V mutation. As intended, the bulkier P1′ moiety of UMass6 helped to retain better potency against I84V, but the reverse was the case for I50V mutation. Thus, although both I84V and I50V are primary resistance mutations with the same side chain change in the S1′ pocket, the interplay between P1′ moiety and rearrangements of inhibitor interactions were distinct, and moreover contrasting for these two mutations. These unexpected and distributed alterations in intermolecular interactions prompted the application of PCA for a collective and comprehensive analysis of the crystal structures. The analysis revealed the interplay between inhibitor modifications and structural response of the target with primary mutations underlying resistance, indicating PCA may be a useful tool to extract determinants or even predictors of inhibitor potency from complex structural information.

2.5 Materials and Methods

2.5.1 Protease gene construction

Protease gene construction was carried out as previously described[40, 99]. The NL4-3 strain has four naturally occurring polymorphisms in the protease relative to the SF2 strain[78, 100]. In short, the protease variant genes (I50V, V82I, I84V) were constructed using QuikChange site-directed mutagenesis (Genewiz) onto NL4-3 WT protease on a pET11a plasmid with codon optimization for protein expression in Escherichia coli. A Q7K mutation was included to prevent autoproteolysis[101].

2.5.2 Protein expression and purification

The expression, isolation, and purification of WT and mutant HIV-1 proteases used for the kinetic assays and crystallization were carried out as previously described[40, 99]. Briefly, the gene encoding the HIV protease was subcloned into the heat-inducible pXC35 expression vector (ATCC) and transformed into E. coli TAP-106 cells. Cells grown in 6L of Terrific Broth were lysed with a cell disruptor and the protein was purified from inclusion bodies[102]. The inclusion body centrifugation pellet was dissolved in 50% acetic acid followed by another round of centrifugation to remove impurities. Size exclusion chromatography was used to separate high molecular weight proteins from the desired protease. This was carried out on a 2.1-L Sephadex G-75 superfine (Sigma Chemical) column equilibrated with 50% acetic acid. The cleanest fractions of HIV protease were refolded into a 10-fold dilution of 0.05 M sodium acetate at pH 5.5, 5% ethylene glycol, 10% glycerol, and 5 mM DTT. Folded protein was concentrated down to 0.5-3mg/mL and stored. This stored protease was used in K_m and K_i assays. For crystallography, a final purification was performed with a Pharmacia Superdex 75 FPLC column equilibrated with 0.05 M sodium acetate at pH 5.5, 5% ethylene glycol, 10% glycerol, and 5 mM DTT. Protease fractions purified from the size exclusion column was concentrated to 1-2 mg/mL using an Amicon Ultra-15 10-kDa device (Millipore) for crystallization.

2.5.3 Enzymatic Assays to Determine K_m and K_i

The K_m and K_i Assays were carried out as previously described[89-90]. In the K_m assay, a 10-amino acid substrate containing the natural MA/CA cut site with an EDANS/DABCYL FRET pair was dissolved in 8% DMSO at 40nM and 6% DMSO at 30 nM. The 30 nM of substrate was 4/5th serially diluted from 30 nM to 6 nM, including a 0 nM control. HIV protease was diluted to 120 nM and, using a PerkinElmer Envision plate reader, 5 μL was added to the 96-well plate to obtain a final concentration of 10 nM. K_m assays were conducted in triplicate. The fluorescence was observed with an excitation at 340 nm and emission at 492 nm and monitored for 200 counts, for approximately 23 minutes. FRET inner filter effect correction was applied as previously described[103]. Data corrected for the inner filter effect was analyzed with Prism7. Fluorescence intensity as a function of time was used to calculate initial velocities by fitting the curves as a single exponential increase. Initial velocities were then fit by non-linear regression to the Michaelis-Menten equation to determine K_m and the standard error of the mean (SEM) was calculated.

To determine the K_i, in a 96-well plate, each inhibitor was 2/3 serially diluted from 3000 pM to 52 pM, including a 0 pM control, and incubated with 0.35 nM protein for 1 hour. A 10-amino acid substrate containing an optimized protease cut site with an EDANS/DABCYL FRET pair was dissolved in 4% DMSO at 120 μM. Using the Envision plate reader, 5 μL of the 120 μM substrate was added to the 96-well plate to a final concentration of 10 μM. K_i assays were conducted in

triplicate. The fluorescence was observed with an excitation at 340 nm and emission at 492 nm and monitored for 200 counts, for approximately 60 minutes. Data was analyzed with Prism7. Fluorescence intensity as a function of time was used to calculate initial velocities by non-linear regression using the one-phase association equation (single exponential increase). The initial velocities as a function of inhibitor concentration were fit by non-linear regression to the Morrison Equation to obtain K_i values and the standard error of the mean (SEM) was calucated from the global fit of all data points for triplicates.

2.5.4 Gag Polyprotein Cleavage Assay

The pET28a plasmid containing full length $Pr55^{Gag}$-TEV-His construct was a kind gift from Maria Bewley and John Flanagan. Protein expression and purification was done as described by Bewley et al. (*Protein Expr Purif*, 2017; 130:137-145) and consists of the removal of DNA and ammonium sulfate precipitation. On the day of the assay, ammonium sulfate pellets were dissolved in resuspension buffer (10 mM HEPES pH 7.5, 500 mM NaCl, 0.1 mM TCEP and 0.1 mM EDTA) and diluted 1:5 with the same buffer without NaCl. The final solution was centrifuged at 20k x *g* for 20 min. Cleavage of $Pr55^{Gag}$ polyprotein by HIV-1 protease was monitored by SDS-PAGE of cleavage products visualized by Coomassie staining. Samples were taken from the reaction mixture at designated time points, and the cleavage reaction was quenched by adding gel running buffer containing SDS and boiling for 2 min. Samples without any added protease (P55) and the protease itself (Enz) were also visualized on the gels. Amprenavir (APV),

a potent HIV-1 protease inhibitor, was used in the negative controls for non-cleavage. Gag cleavage by I50V protease was detectable even at early time points, indicated by disappearing of the high-molecular weight band corresponding to full-length Gag, and appearance of lower molecular weight cleavage products.

2.5.5 Protein Crystallization

Discovery of the condition producing reproducible co-crystals of DRV with NL4-3 WT protease was achieved using the JCSG+ screen (Molecular Dimensions), in well G11, consisting of 2 M Ammonium Sulfate with 0.1 M Bis-Tris-Methane Buffer at pH 5.5 with a protease concentration of 1.9 mg/mL with 3-fold molar excess of DRV and mixed with the precipitant solution at a 1:2 ratio. After optimization, all subsequent combinations of co-crystals were grown at room temperature by hanging drop vapor diffusion method in a 24-well VDX hanging-drop trays (Hampton Research) with protease concentrations between 1.0 to 2.4 mg/mL with 3-fold molar excess of inhibitors set the crystallization drops with the reservoir solution consisting of 23-24% (w/v) Ammonium sulfate with 0.1 M bis-Tris-methane buffer at pH 5.5 set with 2 μL of well solution and 1 μL protein and microseeded with a cat whisker. Diffraction quality crystals were obtained within 1 week. As data was collected at 100 K, cryogenic conditions contained the precipitant solution supplemented with 25% glycerol.

2.5.6 Data Collection and Structure Solution

Diffraction data were collected and solved as previously described[40, 104]. Diffraction quality crystals were flash frozen under a cryostream when mounting

the crystal at our in-house Rigaku_Saturn944 X-ray system. The co-crystal diffraction intensities were indexed, integrated, and scaled using HKL3000[105]. Structures were solved using molecular replacement with PHASER[106]. Model building and refinement were performed using Coot[107] and Phenix[108]. Ligands were designed in Maestro and the output sdf file was used in the Phenix program eLBOW[109] to generate the cif file containing atomic positions and constraints necessary for ligand refinement. Iterative rounds of crystallographic refinement were carried out until convergence was achieved. To limit bias throughout the refinement process, five percent of the data were reserved for the free R-value calculation[110]. MolProbity[111] was applied to evaluate the final structures before deposition in the PDB. Structure analysis, superposition and figure generation was done using PyMOL[112]. X-ray data collection and crystallographic refinement statistics are presented in the Supporting Information **[Table S2]**.

2.5.7 Structural Analysis and PCA

Distance-difference matrices were generated for each inhibitor-mutant protease pair to reveal structural changes relative to that inhibitor bound to wild-type protease, as previously described[113]. Briefly, distances between all Cα pairs in the mutant structure were calculated as an *N*x*N* matrix (N = 198 residues for HIV-1 protease), and then those corresponding distances in the wild-type structure were subtracted to construct the distance difference matrix. The mean deviation from the WT structure for each residue was then calculated by taking the average of the absolute value of all the *N* distance differences involving that residue, and

the backbone structure was represented as a cartoon-putty with increasing thickness and warmer color for increasing deviation.

To calculate the intermolecular van der Waals (vdW) interaction energies the crystal structures were prepared using the Schrodinger protein preparation wizard[114]. Hydrogen atoms were added, protonation states determined and the structures were minimized. The protease active site was monoprotonated at D25. Subsequently forcefield parameters were assigned using the OPLS2005 forcefield[115]. Interaction energies between the inhibitor and protease were estimated using a simplified Lennard-Jones potential, as previously described[116]. For each protease residue, the change in vdW interactions relative to WT complex was also calculated for the mutant structures. PCA of the data matrix was performed as described earlier[43]. For the calculation of the principal components, the implementation of PCA in scikit-learn was used[117]. Briefly, the intermolecular vdW interaction energy with the inhibitor for each residue in a given structure was calculated for each of the 12 structures to yield a 198x12 matrix. The dimensionality of the data set was then reduced using PCA to identify orthogonal linear combinations of variables, or principal components (PCs), that best account for the variance in the data. The PCs are ordered starting from first PC according to the greatest variance represented in the data, and contribution of the original variables to a given PC is represented by the loading vector.

2.6 Acknowledgements

This research was supported by NIH P01 GM109767. We would like to thank W. Royer, B. Kelch and B. Hilbert for helpful discussions.

CHAPTER 3:

pM to μM–Elucidating the Role of Distal Mutations in HIV-1 Protease in Conferring Drug Resistance

3.1 Abstract

Drug resistance continues to be a growing, global problem. The efficacy of small molecule inhibitors is threatened by pools of genetic diversity in all systems, including antibacterials, antifungals, cancer therapeutics, and antivirals. Resistant variants often include combinations of active site mutations, and distal “secondary” mutations which are thought to compensate for losses in enzymatic activity. HIV-1 protease is the ideal model system to investigate these combinations and underlying molecular mechanisms of resistance. Darunavir (DRV) binds WT HIV-1 protease with a potency of < 5 pM, but we have identified a protease variant that loses potency to DRV 150,000-fold, with 11 mutations in and outside the active site. To elucidate the roles of these mutations in DRV resistance, we used a multidisciplinary approach, combining enzymatic assays, crystallography, and molecular dynamics simulations. Analysis of protease variants with 1, 2, 4, 8, 9, 10 and 11 mutations showed that the primary active site mutations caused ~50-fold loss in potency (2 mutations), while distal mutations outside the active site further decreased DRV potency from 13 nM (8 mutations) to 0.76 μM (11 mutations). Crystal structures and simulations revealed that distal mutations induce subtle changes which are dynamically propagated through the protease. Our results reveal that changes remote from the active site directly and dramatically impact the potency of the inhibitor. Moreover, we find interdependent effects of mutations in conferring high levels of resistance. These mechanisms of resistance are likely applicable to many other quickly evolving drug targets, and the insights may have implications for the design of more robust inhibitors.

3.2 Introduction

Drug resistance is a growing global problem[2-3]. The efficacy of current drugs is threatened by pools of genetic diversity in many therapeutic targets, including antibacterials, antifungals, cancer therapeutics, and antivirals[1]. Specifically, rapid evolution in viruses can cause emergence of variants that are not effectively inhibited by small molecule inhibitors. Human immunodeficiency virus (HIV) is capable of accumulating a large number of mutations in response to antiretroviral therapy[118-119].

The viral enzymes coded by the HIV genome, including HIV-1 protease, have been targets of extensive drug development efforts. HIV-1 entry, nucleoside reverse transcriptase, non-nucleoside reverse transcriptase, integrase, and protease inhibitors are used in combination to curb the emergence of drug resistance. Resistance can develop quickly against a single agent, with mutations resulting from the high infidelity of the reverse transcriptase[9, 120-121] and the high recombination rate of HIV[11, 122]. Mutations may also arise from patient non-compliance due to a number of factors, such as adverse side effects and high pill burden[123]. HIV-1 protease can tolerate a large number of mutations, with 45 out of 99 amino acid positions able to mutate while retaining adequate enzymatic activity to allow replication[124]. Viral variants from patients who failed therapy and in vitro selection experiments often contain complex combinations of multiple mutations. Especially with potent inhibitors, a single mutation is not sufficient and many mutations need to accumulate to confer appreciable levels of resistance. Under drug pressure, resistance and compensatory mutations accumulate in the gag and gag-pol viral polyproteins, resulting in often unpredictable and complex mechanisms of resistance[125-126]. Given the prevalence and complexity of these mutations and underlying

mechanisms, HIV-1 has served as a model system to identify, analyze, and characterize drug resistance.

Darunavir (DRV) is the latest and most potent FDA-approved HIV-1 protease inhibitor[78]. The potency of DRV is due to several factors, including a number of crucial hydrogen bonds with the protease backbone, conservation of the water-mediated hydrogen bonding network with the flaps, and extensive van der Waals (vdW) contacts with active site residues[84, 127]. Moreover, DRV fits well within the substrate envelope (the consensus volume occupied by the natural substrates[43]), which explains its low susceptibility to resistance[27, 45, 128]. DRV can effectively inhibit variants with common single primary resistance mutations such as I84V and I50V/A71V[37]. However, high levels of resistance are reached as mutations accumulate, both within and outside the active site of HIV-1 protease.

The combinations of mutations in HIV-1 protease can utilize complex mechanisms of resistance to reach significant levels of resistance. The primary mutations proximal to the active site are thought to directly confer drug resistance while distal mutations are considered compensatory and restore catalytic function[30]. However distal mutations can also weaken inhibitor binding, as has been reported not only for HIV-1 protease but other drug targets as well[129-132]. Molecular mechanisms such as reducing the coupled motions of the protease and inhibitor[55] and altering protein intra-molecular hydrogen bonds[54] have all been suggested in resistance mechanisms involving mutations distal to the inhibitor binding site. We have previously shown that cross-correlations between protein and inhibitor fluctuations distinguish tight binders from weak binders in both HIV and HCV protease[54, 104, 133]. In addition to weakening inhibitor binding, mutations can alter the

dynamic ensemble of the protease–inhibitor complex and the balance between inhibition and catalytic activity. However, changes to the structure and dynamic ensemble along a resistance pathway, molecular mechanisms of resistance due to combinations of mutations, and the role of distal mutations in conferring resistance versus restoring catalytic activity are not well understood.

Here we investigate highly mutated and highly DRV resistant variants of HIV-1 protease to understand the role of distal mutations and uncover underlying mechanisms of resistance. While DRV inhibits wild type HIV-1 protease with picomolar inhibition constant, a combination of 11 mutations (3 active site and 8 distal) confers near μM resistance to DRV. All of these mutations have been observed in patient isolates, although not in this specific combination[15-16]. Our inhibition assays show that the active site mutations alone do not account for this severe loss in affinity to DRV. We generated protease variants with subsets of these 11 mutations (1, 2, 4, 8, and 10-mutant variants) and found increased resistance with accumulation of mutations. We also tease out the effect of two individual distal mutations on the viral evolution pathway from 8 to 10-mutant variants, through double-mutant cycle analysis of the two possible intermediate 9-mutant variants. Using high-resolution crystal structures, enzymatic and inhibition assays, and molecular dynamics simulations, our findings support the role of distal mutations in conferring drug resistance. In addition to restoring catalytic activity, distal mutations contribute to drug resistance through interdependent, dynamic processes propagated throughout the protease.

3.3 Results

3.3.1 Viral passaging under drug pressure selected for highly mutated variants

Viral passaging experiments were performed to elucidate drug resistance in HIV-1 protease under increasing DRV selective pressure, and identified 2 highly abundant variants with 8 and 10 mutations[96]. At the highest drug concentration, an additional mutation was selected, leading to a variant with 11 total mutations (11Mut) [**Figure 3.1A,B**]. Sequencing from early time points indicated that I84V (1Mut) was the first fixed mutation. A variant with both V82F and I84V (2Mut) was generated to investigate primary active site mutations. A midpoint variant between 2Mut and 8Mut was chosen to include 2 active site and 2 flap mutations (I84V, V82F, K45I, and M46I; 4Mut variant). Addition of 4 more mutations (I13V, G16E, V32I, and I33F) to 4Mut yielded the 8Mut variant that was highly abundant in the viral passaging at higher DRV concentrations. With increased DRV pressure, the next highly abundant variant included additional A71V and L76V mutations (10Mut). The possible 9-mutant intermediate variants en route to 10Mut from 8Mut were not observed in viral passaging. To understand the role of the additional A71V and L76V mutations individually, 9Mut-A71V and 9Mut-L76V variants were generated. Lastly, the addition of I54L mutation onto 10Mut created the 11Mut variant. For all these 8 variants [**Figure 3.1A**] and wild type protease, the catalytic activity, DRV inhibition constant, and high-resolution crystal structure were determined and MD simulations performed to elucidate the molecular mechanisms of resistance.

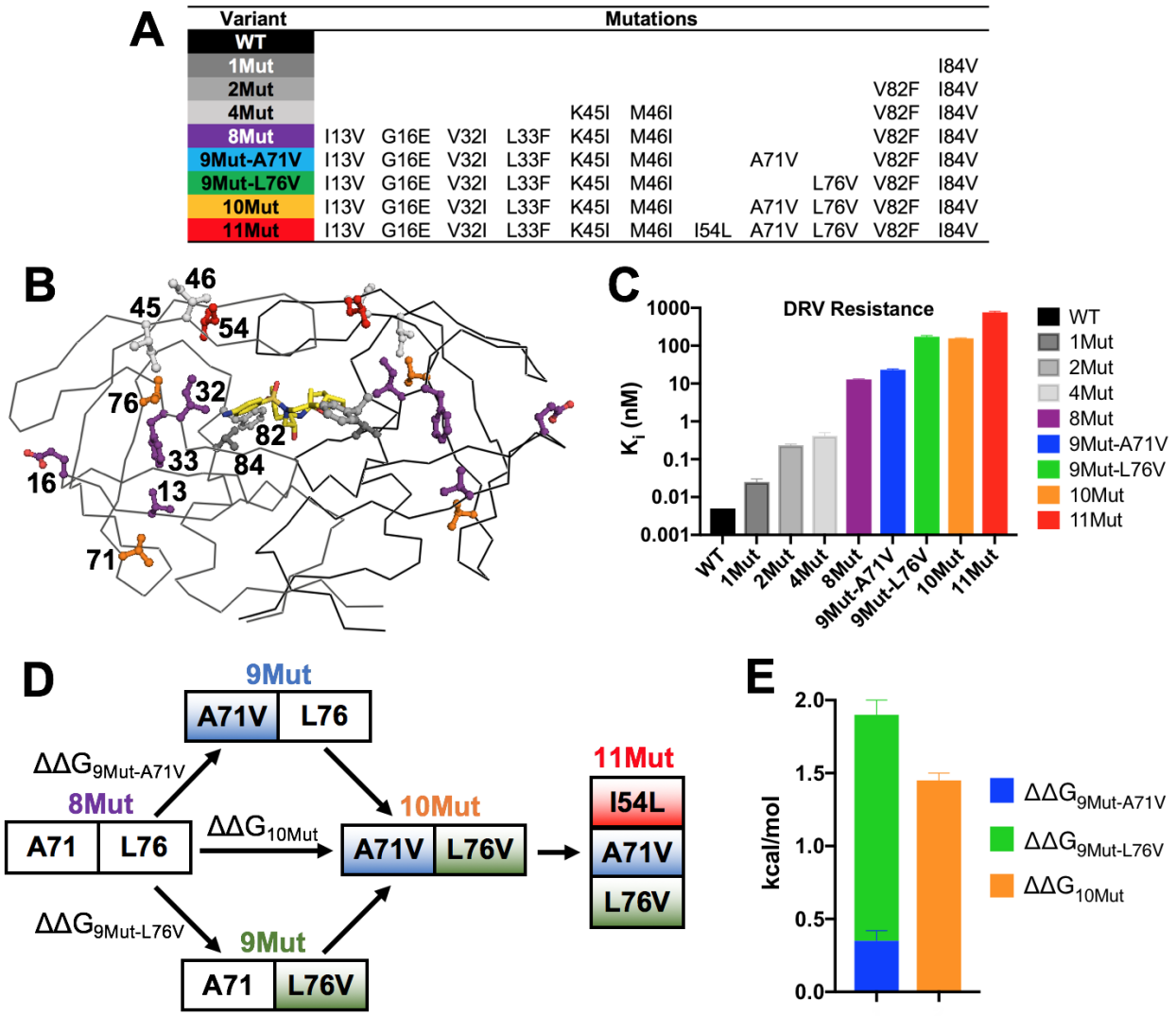


Figure 3.1 A-B) Location of resistance mutations in the variants displayed on HIV-1 protease structure, with side chain of residues mutated shown as ball and stick and colored to match the variant added to in panel B. DRV depicted as yellow sticks. Mutations V32I, V82F, and I84V are classified as proximal while all other mutations are distal. C) Inhibition constant, K_i, of each variant against DRV. Resistance increases as mutations accumulate. D) Double mutant cycle of distal mutations A71V and L76V going from 8Mut to 10Mut variant with 12-fold higher resistance. The additional I54L mutation (11Mut) further increases resistance 5-fold. E) Mutations are interdependent ($\Delta\Delta G_{9Mut\text{-}A71V} + \Delta\Delta G_{9Mut\text{-}L76V} \neq \Delta\Delta G_{10Mut}$).

3.3.2 Resistance mutations are selected to maintain catalytic activity

To determine the effects of mutation accumulation on protease catalytic activity, an established enzyme assay for cleaving the natural MA/CA site was used to measure the Michaelis-Menten constant (K_M) and turnover rate (k_{cat}) [**Table 3.1**]. The ratio of the two measured values was used to calculate the catalytic efficiency (k_{cat}/K_M) for all protease variants. Wild type protease had a K_M of 71.4 μM and k_{cat} of 1282.7 s^{-1}, yielding a catalytic efficiency of 17.1 $\mu M^{-1}s^{-1}$. 1Mut, 2Mut, and 4Mut variants had WT-like K_M values. However, k_{cat} of these variants were 2–7 fold lower than WT protease, resulting in decreased catalytic efficiency. 2Mut variant had a 5-fold decrease in k_{cat}/K_M compared to WT (k_{cat}/K_M = 3.4 ± 0.2 $\mu M^{-1}s^{-1}$).

Relative to WT protease, K_M values of the highly mutated 8Mut, 9Mut-A71V, 9Mut-L76V, and 10Mut variants increased 1.4 to 2.4-fold. The k_{cat} value of the 8Mut and 10Mut decreased 2–3 fold, and the k_{cat}/K_M 3–5-fold relative to WT. The most striking loss was in the turnover rates of the 9Mut variants, with 2 orders of magnitude slowing in the cleavage rate (k_{cat} = 16.9 ± 0.3 and 17.7 ± 0.3, respectively for 9Mut-A71V and 9Mut-L76V). Combined with the increase in K_M, these variants were extremely inefficient with k_{cat}/K_M values 171-fold lower relative to WT, representing a severe enzymatic penalty for these intermediate variants that were undetectable in viral passaging. Strikingly, co-occurrence of the A71V and L76V mutations in the 10Mut variant restored catalytic efficiency (k_{cat}/K_M = 5.3 ± 0.1 $\mu M^{-1}s^{-1}$). The 11Mut exhibited restored, WT-like K_M, but was 13-fold less catalytically efficient compared to WT protease (k_{cat}/K_M = 1.3 ± 0.4 $\mu M^{-1}s^{-1}$).

Table 3.1 Enzyme kinetics of HIV-1 protease variants measured using a natural substrate sequence: Michaelis-Menten constant (K_M), enzyme turnover number (k_{cat}), and catalytic efficiency (k_{cat}/K_M).

	K_M (μM)	k_{cat} (s^{-1})	k_{cat} / K_M ($\mu M^{-1}s^{-1}$)
WT	71.4 ± 6.8	1282.7 ± 0.1	17.1 ± 0.1
I84V	66.4 ± 4.3	782.3 ± 0.1	11.8 ± 0.1
2Mut	55.1 ± 12.4	11.8 ± 0.1	3.4 ± 0.2
4Mut	71.0 ± 4.6	690 ± 0.1	9.7 ± 0.1
8Mut	123.8 ± 30.1	430.7 ± 0.2	3.5 ± 0.2
9Mut-A71V	139.2 ± 35.3	16.9 ± 0.3	0.1 ± 0.2
9Mut-L76V	174.1 ± 44.8	17.7 ± 0.3	0.1 ± 0.3
10Mut	101.4 ± 14.8	532.8 ± 0.1	5.3 ± 0.1
11Mut	63.9 ± 22.6	81.9 ± 0.2	1.3 ± 0.4

This enzymatic assay uses a single cleavage site for a relatively short amount of time. To examine how the highly mutated variants process multiple cleavage sites in full length Gag polyprotein (p55) over a longer time scale, a gel cleavage assay was utilized[134-135] [**Figure 3.2**]. Wild-type protease was extremely efficient cleaving 100% of the p55 polyprotein after 90 minutes. All the mutants were able to cleave the initial p2-NC cleavage site at varying rates. The intermediate 9Mut variants were much slower but were still able to process the multiple cleavage sites in the polyprotein. Overall, all protease variants were able to process substrates, with the highly mutated and abundant 10Mut and 11Mut variants having restored catalytic activity.

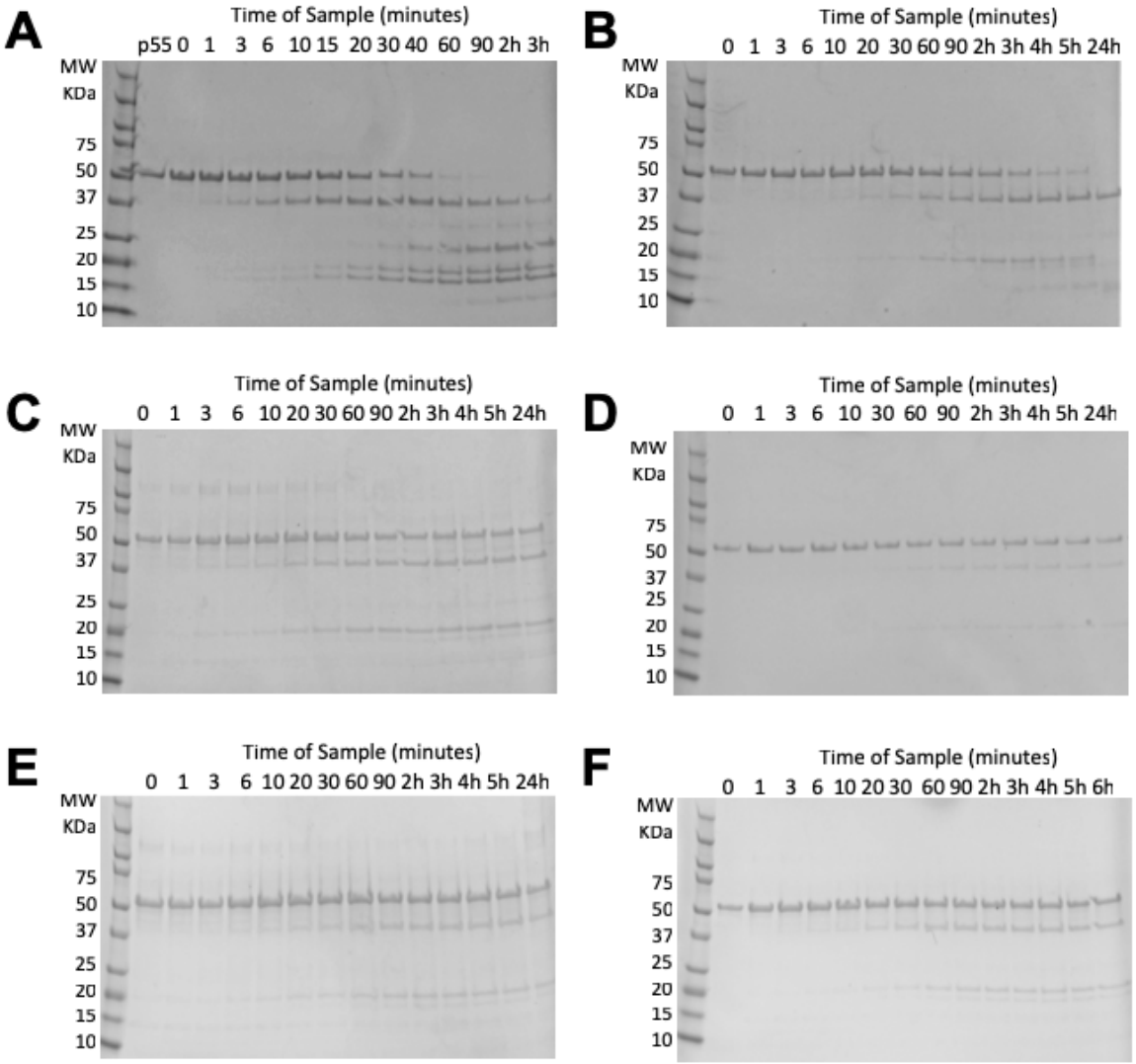


Figure 3.2 Time course cleavage of full-length gag polyprotein, p55, by HIV-1 protease variants. (A) WT, (B) 8Mut, (C) 9Mut-A71V, (D) 9Mut-L76V, (E) 10Mut, and (F) 11Mut.

3.3.3 Accumulation of mutations progressively increases DRV resistance

The enzyme inhibition constant, K_i, of DRV was measured against WT and the selected protease variants using an established assay with an optimized fast-cleaving substrate[90]. Wild type protease is highly susceptible to DRV with a K_i in single digit pM range below the assay limit of reliable detection (K_i < 0.005 nM), as we previously reported[136]. Against 1Mut, 2Mut, and 4Mut, DRV maintained picomolar inhibition, but with K_i increasing progressively 6–80 fold relative to WT [**Figure 3.1C**]. The 8Mut variant was significantly more resistant to DRV, with a 2,560-fold increase in resistance compared to WT (K_i = 12.8 nM). Addition of A71V mutation (9Mut-A71V) resulted in a ~2-fold further increase in resistance (K_i = 23.2 nM). In contrast, addition of L76V (9Mut-L76V) caused an order of magnitude increase in DRV resistance (K_i = 172.7 nM), indicating that the distal L76V mutation contributes a significant level of resistance. The 10Mut variant harboring both distal mutations exhibited 9Mut-L76V-like inhibition (K_i = 156.4 nM). Addition of I54L in 11Mut resulted in dramatic DRV resistance, with a 152,000-fold decrease in inhibition compared to WT (K_i = 759.2 nM). Active site mutations alone do not account for the high levels of resistance seen in the highly mutated variants, as only with the addition of distal mutations the protease was able to exhibit nanomolar-level inhibition. The relationship between DRV inhibition and catalytic efficiency is shown in **Figure 3.3**.

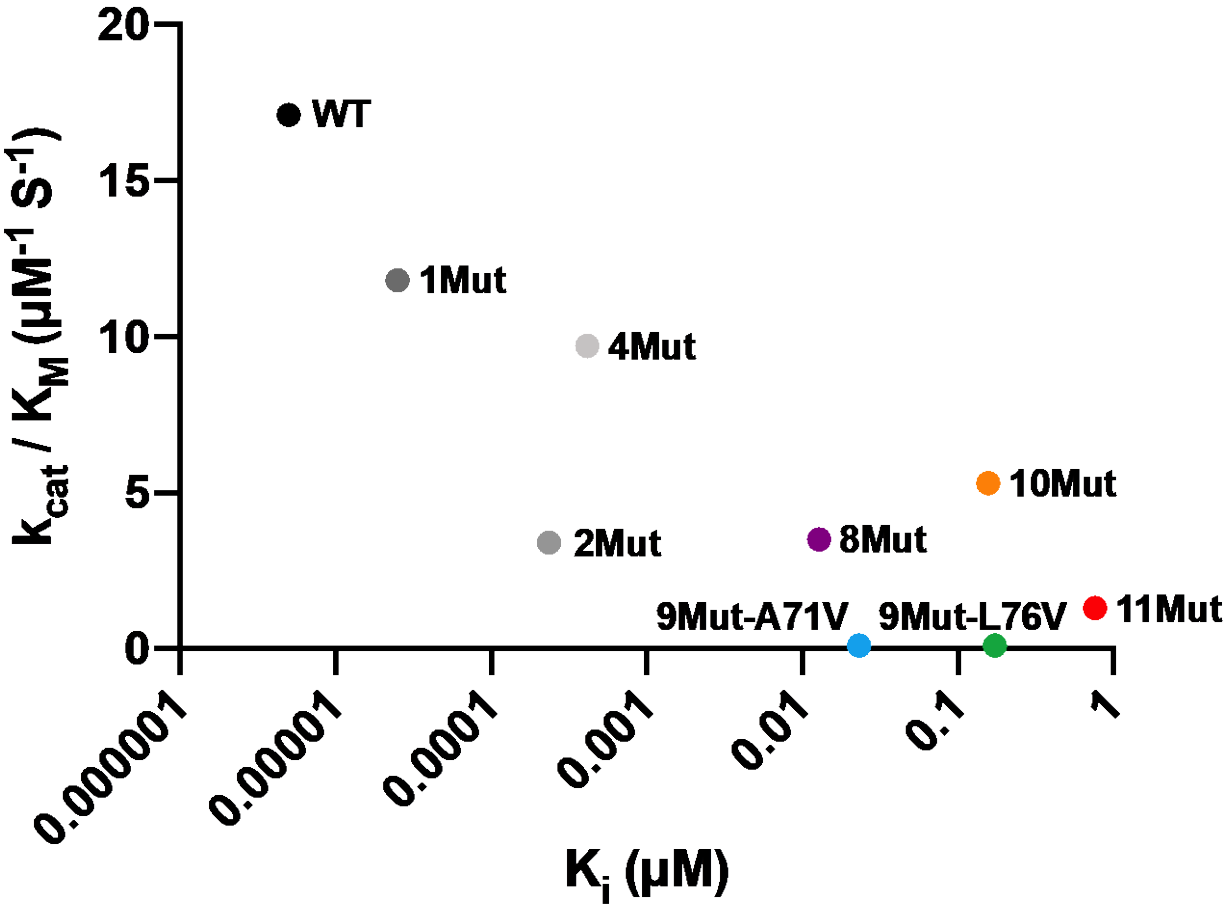


Figure 3.3 The relationship between DRV inhibition and enzyme catalytic efficiency of HIV-1 protease variants with increasing number of mutations.

3.3.4 Double mutant cycle reveals the interdependency of A71V and L76V

Only the 8Mut to 10Mut variants were observed in viral passaging, thus the addition of A71V and L76V mutations individually were examined as well as the double-mutant cycle analysis [**Figure 3.1D**] to elucidate the interdependency of these two distal mutations which cause 12-fold increase in resistance. The change in free energy of inhibitor binding (estimated from $\Delta G = RT\ln[K_i]$)[137] between the reference variant 8Mut and 10Mut ($\Delta\Delta G_{10Mut}$) was calculated. This was then compared to the sum of individual changes for 9Mut-A71V and 9Mut-L76V relative to 8Mut ($\Delta\Delta G_{9Mut\text{-}A71V} + \Delta\Delta G_{9Mut\text{-}L76V}$). When the two mutations are independent, the two values are equal ($\Delta\Delta G_{10Mut} = \Delta\Delta G_{9Mut\text{-}A71V} + \Delta\Delta G_{9Mut\text{-}L76V}$). The sum of free energy changes for the 9-Mut intermediates were more than that for 10Mut [**Figure 3.1D**]**.** $\Delta\Delta G_{9Mut\text{-}A71V}$ was 0.35 ± 0.07 kcal/mol and $\Delta\Delta G_{9Mut\text{-}L76V}$ was 1.55 ± 0.10 kcal/mol, summing up to 1.90 ± 0.12 kcal/mol; however, $\Delta\Delta G_{10Mut}$ was 1.45 ± 0.05 kcal/mol. Thus the double mutant cycle analysis shows that A71V and L76V are interdependent[137] in conferring DRV resistance.

3.3.5 Crystal structures show resistance mutations reduce vdW contacts with DRV

To elucidate the changes in inhibitor binding and molecular mechanisms of resistance, we determined cocrystal structures of DRV bound to the wild type HIV-1 protease of the NL4-3 strain, as well as all eight mutated variants at high resolution (1.9–2.2 Å). Structure of DRV bound to WT protease and the I84V variant were previously determined[37] and the 2Mut crystallized in the same space group ($P2_12_12_1$), which contained one homodimer in the asymmetric unit and one inhibitor bound in a single orientation, allowing direct comparison. The variants containing four or more mutations

had a different preferred crystallographic lattice and were solved in the $P6_1$ space group. For reliable and direct comparison with these variants, WT protease structure was also determined in the hexagonal space group. These structures were also solved containing one homodimer in the asymmetric unit with one inhibitor bound in a single orientation [**Table 3.2**]. Structures of protease variants were overall similar to that of WT protease, with 0.3–0.4 Å overall RMSD [**Figure 3.4**]. Following established convention, throughout this analysis the B chain that contacts the aniline moiety of DRV will be denoted the prime (') chain, while the A chain will remain non-prime. The crystal structures determined provided insights into changes in the overall structure and protein–inhibitor interactions due to the accumulating resistance mutations.

Table 3.2 X-Ray Crystallography Statistics

	WT (NL4-3)	1MUT	2MUT	WT-HEX	4MUT	8MUT	9MUT-A71V	9MUT-L76V	10MUT	11MUT
INHIBITOR	DRV	DRV	DRV	DRV	DRV	DRV	DRV	DRV	DRV	DRV
PDB ID	6DGX	6DH0	6OPT	6OPS	6OPU	6OPV	6OPW	6OPX	6OPY	6OPZ
RESOLUTION (Å)	2.00	1.90	1.96	2.08	1.94	1.91	2.10	2.03	2.13	2.20
SPACE GROUP	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P6_1$	$P6_1$	$P6_1$	$P6_1$	$P6_1$	$P6_1$	$P6_1$
TWIN LAW				h,-h-k,-l	h,-h-k,-l	h,-h-k,-l	h,-h-k,-l	h,-h-k,-l	h,-h-k,-l	h,-h-k,-l
A (Å)	51.0	51.1	50.8	62.3	61.5	61.7	62.1	61.9	62.1	61.8
B (Å)	58.3	58.1	57.9	62.3	61.5	61.7	62.1	61.9	62.1	61.8
C (Å)	61.9	61.8	61.5	82.3	81.3	81.2	82.3	81.5	82.6	82.4
COMPLETENESS	98.1	99.9	96.5	97.0	97.9	99.7	99.9	99.5	97.0	99.8
TOT. REFLECTIONS	84332	181306	252365	104283	126710	135511	110972	127376	53059	92665
UNIQ. REFLECT.	12758	15103	13103	10660	12722	13672	10571	11470	9912	9120
AVG. I/SIGMA	31.1	41.1	53.2	24.0	48.5	43.7	37.1	33.8	29.7	37.2
REDUNDANCY	6.6	12.0	19.3	9.8	10.0	9.9	10.5	11.1	5.4	10.2
R-MERGE (LINEAR)	0.057	0.057	0.056	0.086	0.045	0.052	0.064	0.077	0.048	0.059
RMSD BONDS	0.003	0.003	0.003	0.003	0.002	0.006	0.004	0.003	0.006	0.004
RMSD ANGLES	0.590	0.710	0.663	0.499	0.616	0.888	0.820	0.678	0.783	0.606
R-FREE **R-WORK**	22.8 19.2	21.2 18.3	22.8 18.2	25.1 22.5	23.2 20.5	22.3 17.6	28.5 26.0	25.6 23.1	22.6 19.1	26.8 22.8

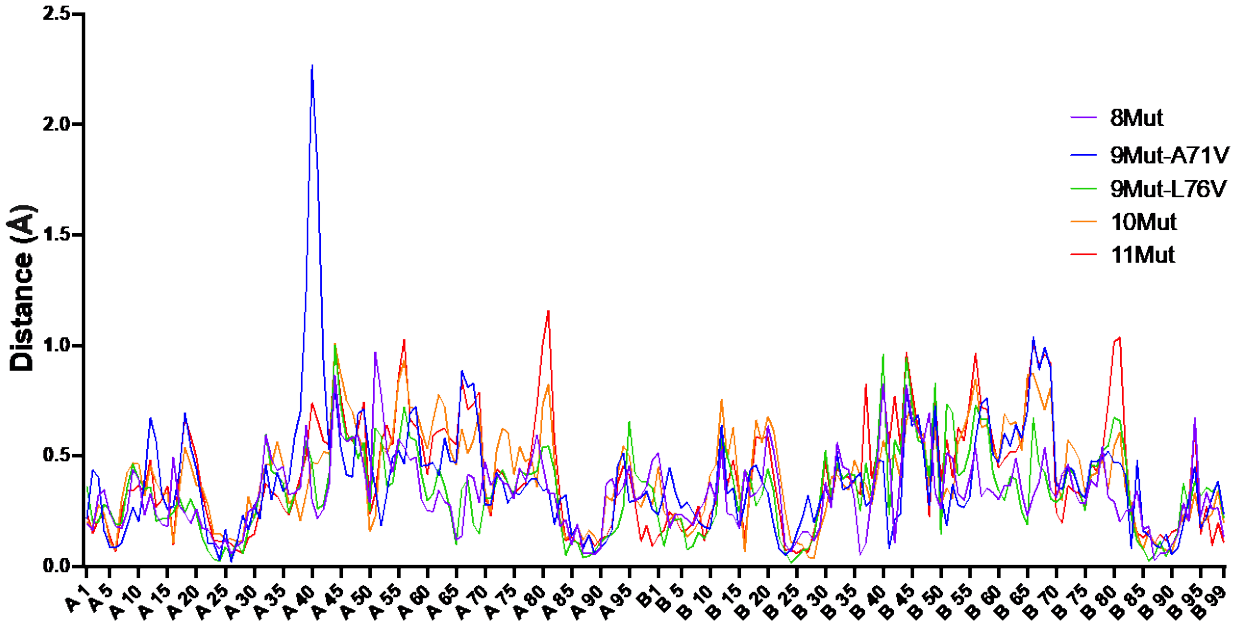


Figure 3.4 Deviation of internal C_α distances in protease variant crystal structures relative to WT–DRV complex.

Resistance mutations selected under DRV selective pressure are expected to directly or indirectly perturb inhibitor binding and result in reduced potency. To elucidate alterations in inhibitor packing at the active site, we calculated vdW contacts of DRV with the protease residues from the crystal structures [**Table 3.3**]. Compared to WT protease, the common multidrug resistance mutation I84V resulted in a loss of 1.2 and 1.8 kcal/mol in vdW contacts of residue 84 and 84', respectively [**Figure 3.5A**, **Table 3.3**]. Similarly, all remaining variants, which also contain this mutation, had 1–2.5 kcal/mol losses in vdW contacts at these two residues. In the presence of single I84V mutation, the location of the other active site mutation, V82, had a compensatory gain in vdW contacts in chain B, which was progressively alleviated as the mutations were added. Overall, V82F exhibits a minor gain in vdW contacts (0.2 kcal/mol on average), but the larger nonpolar side chain significantly shifts Arg8' away from DRV resulting in 0.6–1.1 kcal/mol loss in vdW contacts. Addition of flap mutations K45I and M46I caused reduced vdW contacts of I50 in 4Mut variant, and a net loss in total vdW contacts which correlate with reduced potency [**Figure 3.5B**, **Table 3.3**]. The highly resistant 8Mut variant structure continued the trend, exhibiting a net reduction in vdW contacts. Notably, although the V32I mutation results in a larger amino acid at the edge of the active site, vdW contacts with DRV were reduced 0.4–0.6 and 0.5–1.0 kcal/mol due to the orientation of the additional methyl group away from the inhibitor [**Figure 3.5A**, **Table 3.3**]. The combination of I13V, L33F and V32I exacerbated losses of vdW contacts at residues I84V and I84V' by 0.2 and 0.7 kcal/mol, respectively. Each mutation, both proximal and distal, either directly or indirectly reduced vdW contacts with the inhibitor, thereby destabilizing inhibitor binding.

Table 3.3 A) Protease–DRV per residue vdW contacts for all crystal structures. Darker red indicates more contacts. B) Difference in per residue protease–DRV vdW contacts relative to WT cocrystal structure in the same space group. Positive numbers and cooler colors indicate reduced vdW contacts and negative numbers and warmer colors indicate increased vdW contacts, relative to WT structure.

A

	WT (P2)	I84V	2Mut	WT (P6)	4Mut	8Mut	9Mut A71V	9Mut L76V	10Mut	11Mut
A23	-1.0	-1.0	-1.1	-1.1	-0.9	-1.0	-0.8	-0.9	-1.2	-1.2
A27	-3.8	-3.6	-3.8	-4.3	-3.9	-3.9	-4.0	-3.9	-4.0	-3.7
A28	-5.4	-5.6	-5.6	-5.4	-5.2	-5.2	-5.0	-5.1	-5.2	-5.5
A29	-4.2	-4.2	-4.2	-4.1	-3.9	-3.9	-4.1	-3.6	-4.1	-4.0
A30	-3.8	-3.8	-3.0	-3.0	-3.0	-3.1	-2.7	-3.0	-3.0	-3.5
A32	-1.2	-1.2	-1.1	-1.2	-1.1	-0.6	-0.7	-0.5	-0.7	-0.8
A47	-1.5	-2.7	-2.6	-2.0	-1.7	-2.1	-1.7	-1.7	-1.8	-1.9
A48	-2.1	-2.7	-2.5	-2.6	-2.3	-2.3	-2.7	-2.1	-2.3	-2.3
A49	-2.9	-3.3	-3.0	-3.2	-3.2	-2.9	-3.1	-2.7	-2.9	-2.4
A50	-6.1	-5.8	-5.7	-5.7	-6.3	-5.5	-5.6	-6.1	-5.3	-4.9
A81	-0.5	-0.5	-0.7	-0.6	-0.7	-0.9	-1.3	-0.9	-0.9	-0.9
A82	-1.1	-1.2	-0.6	-1.4	-1.7	-1.6	-1.6	-1.7	-1.6	-1.6
A84	-3.1	-2.0	-2.0	-2.8	-1.6	-1.4	-1.3	-1.3	-1.5	-1.9
B8	-1.4	-1.2	-0.7	-1.6	-0.9	-0.6	-1.0	-0.9	-0.7	-0.9
B23	-1.5	-0.8	-1.3	-1.5	-1.2	-1.3	-1.2	-1.3	-1.3	-1.3
B27	-2.8	-2.8	-2.8	-2.7	-2.7	-3.1	-3.3	-2.9	-2.7	-2.9
B28	-4.4	-3.5	-3.9	-4.8	-4.3	-4.3	-4.2	-4.2	-4.6	-4.6
B29	-1.6	-1.4	-1.6	-1.8	-1.5	-1.6	-2.4	-2.0	-2.1	-1.8
B30	-2.1	-2.1	-2.1	-2.5	-2.3	-2.5	-2.5	-2.7	-2.8	-2.7
B31	-0.5	-0.4	-0.4	-0.5	-0.5	-0.4	-0.3	-0.3	-0.4	-0.3
B32	-1.7	-2.1	-2.0	-1.7	-1.8	-1.0	-0.8	-0.8	-1.2	-1.0
B47	-3.2	-2.9	-3.1	-2.2	-2.0	-1.9	-1.4	-1.6	-1.7	-1.7
B48	-1.8	-1.7	-1.8	-1.7	-1.5	-1.8	-1.7	-1.9	-1.7	-1.7
B49	-1.8	-1.8	-1.8	-2.1	-2.4	-2.5	-2.6	-2.6	-2.1	-2.0
B50	-3.6	-3.5	-3.4	-4.3	-4.4	-4.5	-5.1	-4.7	-4.7	-4.4
B76	-0.6	-0.5	-0.6	-0.5	-0.5	-0.4	-0.3	-0.1	-0.2	-0.1
B81	-1.6	-1.6	-1.6	-1.6	-1.8	-2.0	-1.9	-1.9	-2.0	-1.7
B82	-2.2	-3.2	-2.9	-2.4	-2.8	-2.4	-2.0	-2.7	-2.5	-2.3
B84	-4.0	-2.3	-2.4	-4.2	-2.7	-2.1	-1.8	-2.3	-2.1	-1.9
Total	-83.1	-80.5	-79.4	-84.9	-79.9	-77.8	-78.3	-77.4	-77.9	-76.7

B

WT-P2		WT-P6					
I84V	2Mut	4Mut	8Mut	9Mut A71V	9Mut L76V	10Mut	11Mut
0.1	0.0	0.2	0.2	0.3	0.2	-0.1	-0.1
0.2	0.1	0.3	0.3	0.2	0.3	0.3	0.5
-0.2	-0.2	0.2	0.2	0.4	0.3	0.2	-0.1
0.0	0.0	0.2	0.1	-0.1	0.5	0.0	0.0
0.0	0.8	-0.1	-0.1	0.2	0.0	-0.1	-0.6
0.0	0.1	0.1	0.5	0.5	0.6	0.5	0.4
-1.2	-1.1	0.3	-0.1	0.2	0.3	0.2	0.1
-0.6	-0.4	0.3	0.2	-0.1	0.5	0.3	0.3
-0.4	-0.1	0.0	0.4	0.1	0.5	0.3	0.9
0.3	0.4	-0.6	0.2	0.1	-0.4	0.4	0.8
0.0	-0.1	-0.2	-0.3	-0.8	-0.3	-0.3	-0.4
-0.1	0.5	-0.3	-0.2	-0.2	-0.3	-0.2	-0.2
1.2	1.1	1.2	1.4	1.5	1.5	1.4	0.9
0.2	0.7	0.8	1.1	0.6	0.7	1.0	0.7
0.7	0.2	0.2	0.2	0.3	0.1	0.1	0.2
0.0	0.0	0.0	-0.3	-0.6	-0.2	0.0	-0.2
0.9	0.4	0.4	0.5	0.6	0.5	0.1	0.2
0.1	-0.1	0.3	0.2	-0.6	-0.2	-0.3	0.0
0.1	0.1	0.2	0.0	0.0	-0.2	-0.3	-0.3
0.0	0.0	-0.1	0.1	0.2	0.1	0.0	0.1
-0.3	-0.3	-0.1	0.7	1.0	0.9	0.5	0.7
0.3	0.1	0.2	0.3	0.8	0.6	0.5	0.5
0.1	0.0	0.2	-0.1	0.0	-0.2	0.0	0.0
0.0	0.0	-0.3	-0.4	-0.6	-0.5	-0.1	0.1
0.2	0.2	-0.1	-0.2	-0.8	-0.4	-0.3	0.0
0.1	0.0	0.0	0.1	0.2	0.4	0.3	0.4
0.0	0.0	-0.2	-0.4	-0.3	-0.3	-0.4	-0.1
-1.0	-0.7	-0.4	-0.1	0.4	-0.3	-0.1	0.0
1.8	1.6	1.5	2.2	2.5	1.9	2.1	2.4
2.6	3.7	5.1	7.2	6.7	7.5	7.0	8.2

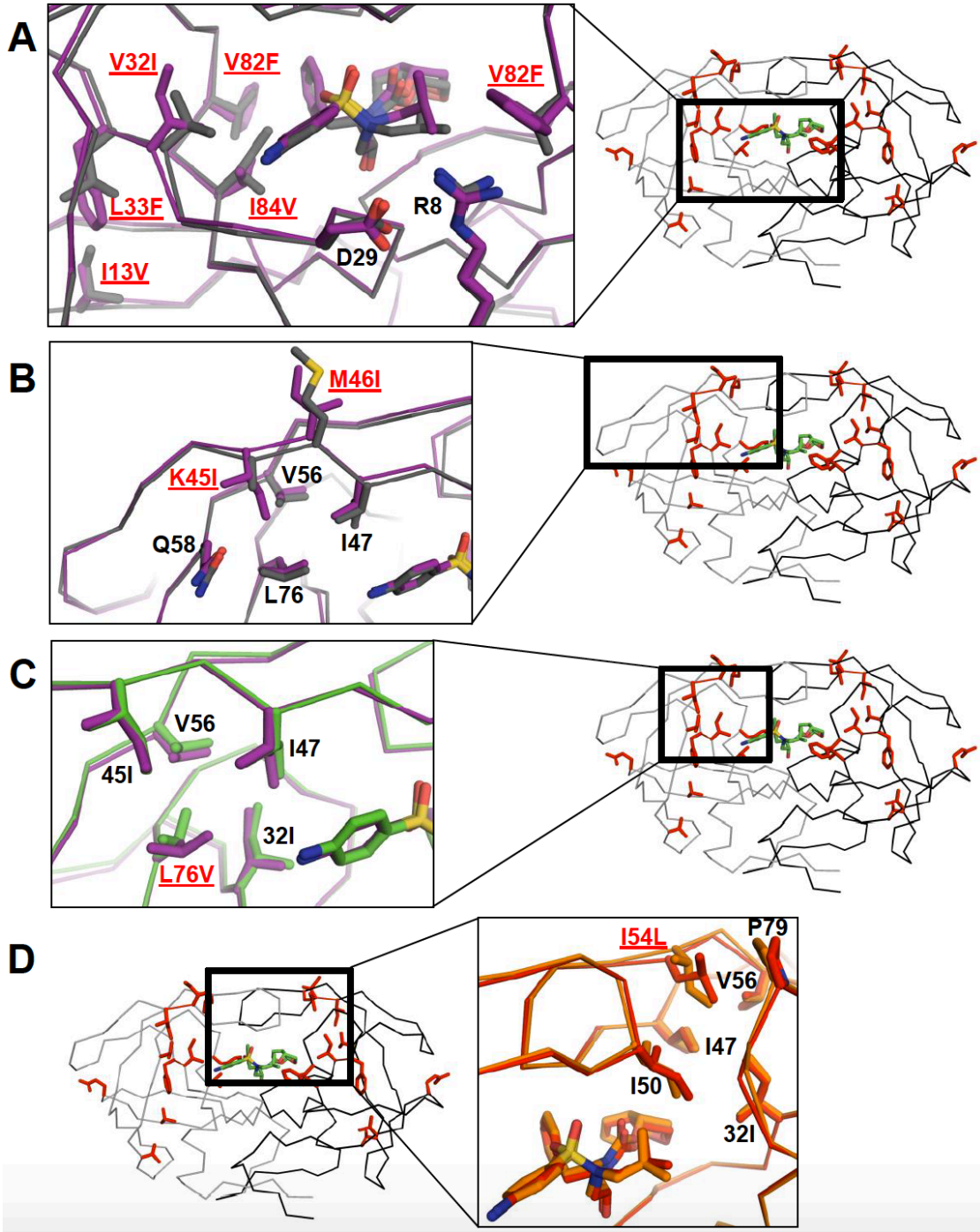


Figure 3.5 A,B) Comparison of WT (black) and 8Mut (purple) variants, focusing on the active site, core, and flaps. C) 8Mut (purple) and 9Mut-L76V (green) variants, focusing on the hydrophobic packing at base of the flaps. D) 10Mut (orange) and 11Mut (red) variants, which differ at I54L mutation, focusing on the hydrophobic packing upon P79.

Although the 9Mut variants were not observed in viral passaging, determining the structures of these complexes highlighted the interdependent structural changes of the mutations at residue 71 and 76 connecting the 8Mut and 10Mut variants. The A71V mutation, located on the distal “70’s loop”, is a well-known compensatory mutation that restores catalytic activity of protease variants containing I50V mutation. The mean differences in internal C_α distances [**Figure 3.6**], also known as distance difference matrices, were calculated to quantify the structural alterations. Comparing the 8Mut and 9Mut-A71V structures, the A71V mutation pushes the flexible “70s loop” outward up to 0.61 and 0.63 Å for residues 67 and 68', respectively. This outward shift can perturb the beta strand (residues 69-76) that directly interacts with the neighboring beta strand extending up to the flaps (residues 53-65). Although there are very few changes in the internal C_α distances of 8Mut and 9Mut-L76V, when comparing 8Mut and 10Mut the changes in distances reflect an additivity of the differences between the 8Mut versus the two 9Mut structures [**Figure 3.6**].

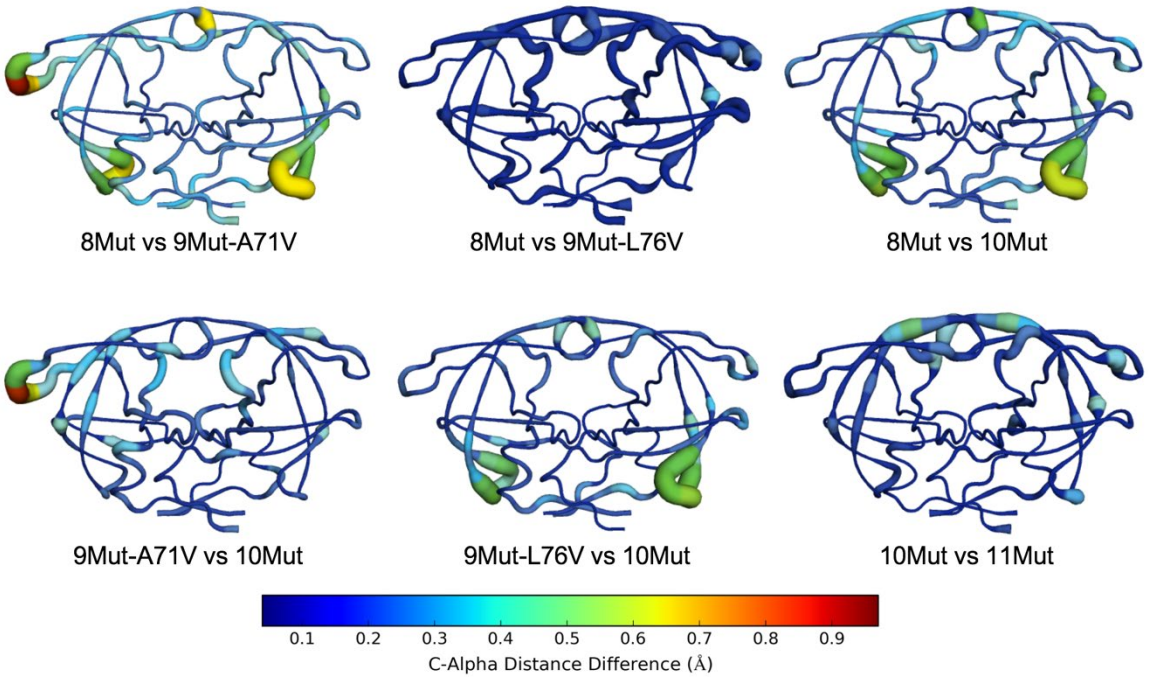


Figure 3.6 Internal C_α distance differences in crystal structures mapped onto HIV-1 protease structure. Warmer colors and larger putty thickness indicate larger differences between the variants compared.

Comparing the complexes of 10Mut and 11Mut helps explain the 5-fold reduction in potency. The longer side chain resulting from the I54L mutation pushed against Pro79 creating a mean difference of 0.36 Å at residues 54 and 54' and 0.20 Å at residues 79 and 79' **[Figure 3.6]**. This widening at the active site may destabilize flap closing, and most notably resulted in reduced vdW contacts at residues 49 and 50 of 0.9 and 0.8 kcal/mol, respectively. 11Mut–DRV cocrystal structure had the lowest total intermolecular vdW contacts, correlating with the highest DRV resistance.

3.3.6 Resistance mutations perturb packing in the hydrophobic core

Many of the mutations observed were changes from one hydrophobic side chain to another, located in the hydrophobic core of the protein, that considerably change the packing between hydrophobic side chains. K45I results in a side chain rotamer that increases hydrophobic packing with residues I47, V56, L76 at the base of the flaps [**Figure 3.5B**]. While I84V may only bridge the gap between V82F and V32I [**Figure 3.5A**], V32I is involved in both the L76V cluster (also including K45I, I47, and V56) [**Figure 3.5C**] and the I54L cluster (also including I47, I50, and V56) [**Figure 3.5D**]. Mutations that perturb the packing of these clusters may also affect "hydrophobic sliding" and thus conformational dynamics of the protease[52, 138].

3.3.7 Resistance mutations increase protease fluctuations

Despite 5 orders of magnitude increase in the inhibition constant (K_i) of the 11Mut variant relative to WT protease, DRV displayed only minor conformational differences in the cocrystal structures, and MD simulations were performed to better interrogate these structures. Starting from the cocrystal structures of the DRV complexes with WT protease and the resistant variants, three replicates of fully hydrated 100 ns MD simulations were performed for each complex. All simulations reached convergence [**Figure 3.7**]. The protein backbone dynamics were compared via root-mean-square fluctuations (RMSF) of C_α atoms [**Figure 3.8**]. In WT–DRV complex, the entire backbone was stable, including the flaps (50s region) that close upon the bound inhibitor. In I84V variant, the flaps are slightly

more mobile but the active site remains stable. Relative to I84V, the 2Mut (I84V/V82F) protease has reduced flaps fluctuations [**Figure 3.9**]. The addition of flap mutations K45I and M46I in the 4Mut resulted in higher fluctuations at chain A flap and residues 79-82. Chain B flap and residues 79'-82' had WT-like fluctuations, remaining stable throughout the MD simulations. Overall, the backbone flexibility of the highly mutated 8Mut, 9Mut-A71V, 9Mut-L76V and 10Mut variants were similar to WT complex, except increased fluctuations at residue 65' in 10Mut. 11Mut variant, bearing flap mutation I54L, displayed extremely large fluctuations at both flaps (residues 42-55 and 45'-52'), increasing more than 0.6 and 0.5 Å in chain A and B, respectively, relative to WT protease. In addition, 11Mut protease had increased fluctuations at residues 79-82 and 78'-80' as well as residues 26-28 in the active site. Thus, the addition of I54L mutation resulted in considerable increase in protease backbone fluctuations at the active site and flaps.

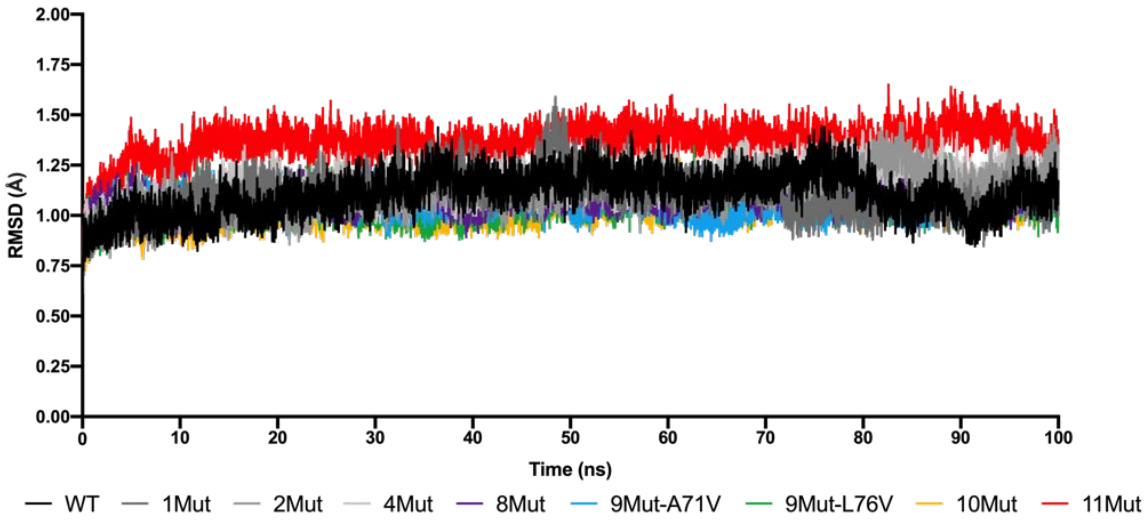


Figure 3.7 Root-Mean-Square Deviation (RMSD) of protein C_α atoms during MD simulations.

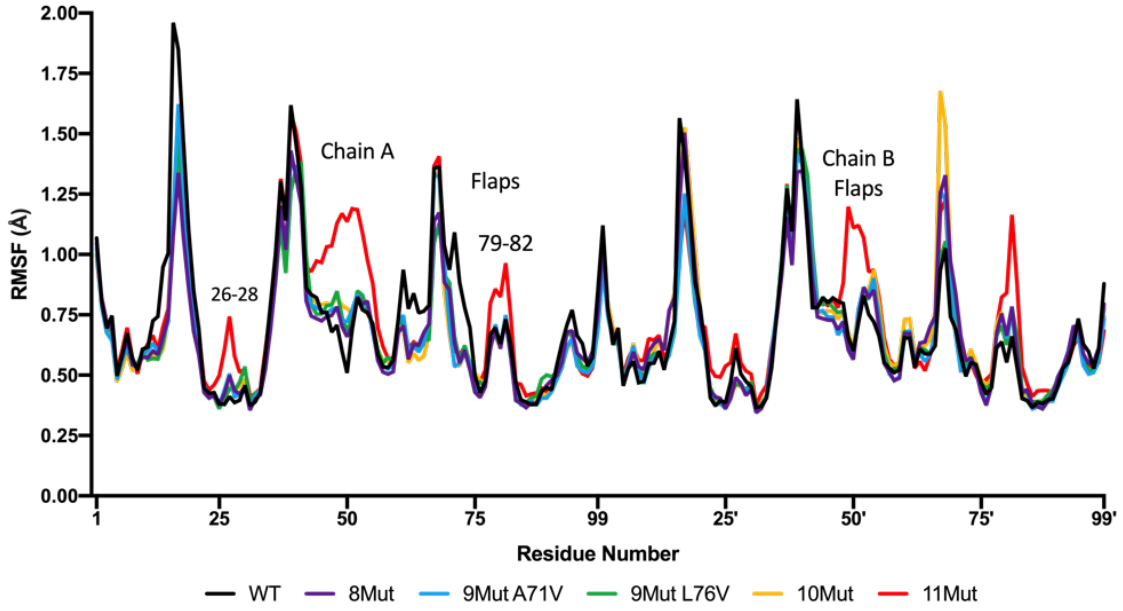


Figure 3.8 Root Mean Square Fluctuation (RMSF) of C_α atoms for WT, 8Mut, 9Mut-A71V, 9Mut-L76V, 10Mut, and 11Mut from MD simulations.

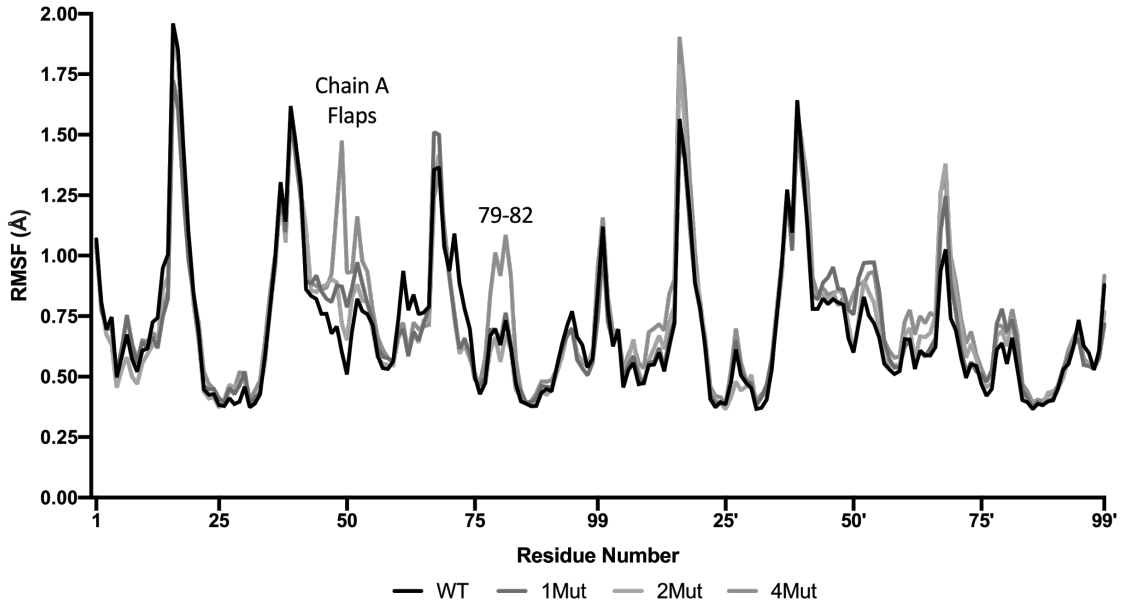


Figure 3.9 Root Mean Square Fluctuation (RMSF) of C_α atoms for WT, 1Mut, 2Mut, and 4Mut from MD simulations.

To compare the dynamic ensemble of DRV-bound variants, mean internal C_α distances from MD simulations were compared, similar to the comparison of crystal structures [**Figure 3.10**]. In this analysis, the internal distances between all C_α carbons were measured for every frame of the trajectory and an average was taken. Comparing 8Mut to 9Mut-A71V, changes were mostly localized to the 70s beta-sheets and residues 15–20 in both chains, with small changes at the flap elbows [**Figure 3.10A**]. Comparing 8Mut to 9Mut-L76V, changes were located in both flaps and the flap elbows. The 80s loop (residues 78 to 82) also had slight changes. In addition, the B chain 70s beta-sheet was affected while the A chain beta-sheet showed very little change [**Figure 3.10B**]. As was observed with the crystal structures, alterations in the two 9-Mut variants were additive when 10Mut was compared to 8Mut. The exception was that changes at residues 78-82 were attenuated when A71V and L76V were simultaneously present [**Figure 3.10C**]. To complete the cycle, the two 9Muts were compared with 10Mut. Similar to the effects noted earlier, A71V induced changes at the 70s beta-sheets and residues 15–20, while L76V distally affected the flaps and flap elbows [**Figure 3.10D-E**]. These alterations in the dynamic ensemble were consistent with the comparison of crystal structures [**Figure 3.6**]. However, alterations in the 11Mut variant were accentuated in the conformational dynamics in the MD simulations, with large structural changes at both flaps and the active site [**Figure 3.10F**]. Compared to other variants, the flaps of 11Mut were more open as indicated by the increasing distance between the C_α atoms of I50-V84' and I50'-V84 [**Figure 3.11**]. Thus, I54V

mutation in 11Mut resulted in destabilization and increased fluctuations of the protease, particularly in the flaps.

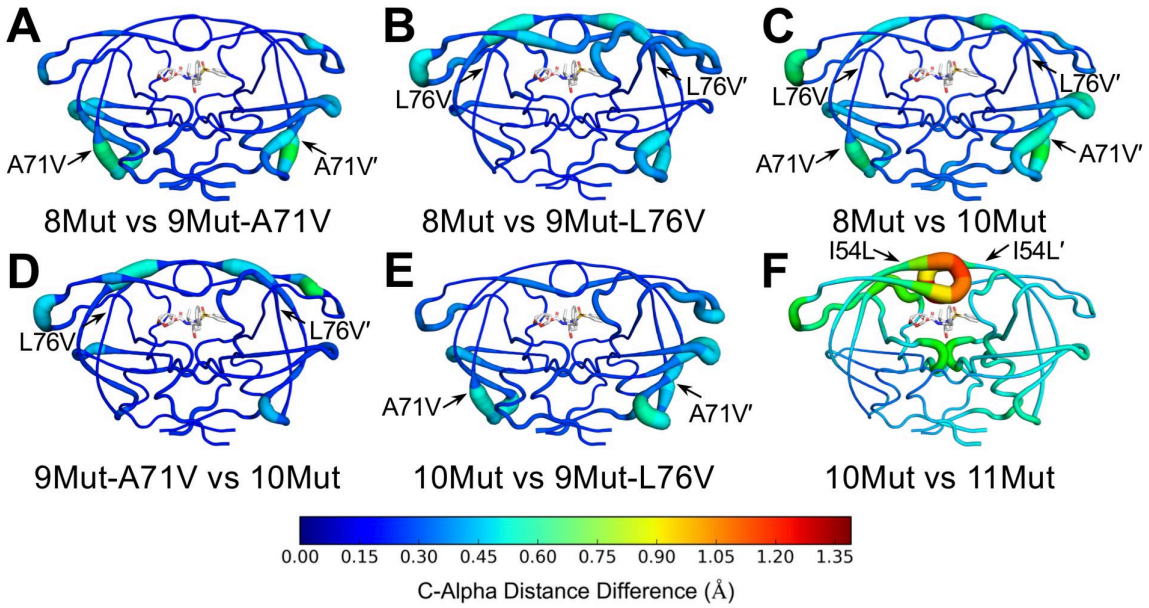


Figure 3.10 Mean internal C_α distance differences between (A) 8Mut and 9Mut-A71V, (B) 8Mut and 9Mut-L76V, (C) 8Mut and 10Mut, (D) 9Mut-A71V and 10Mut, (E) 10Mut and 9Mut-L76V, and (F) 10Mut and 11Mut protease variants from MD simulations plotted onto protease structure. In all panels, DRV is shown as sticks in the active site. Putty thickness and color indicate distance difference where hot colors indicate large changes.

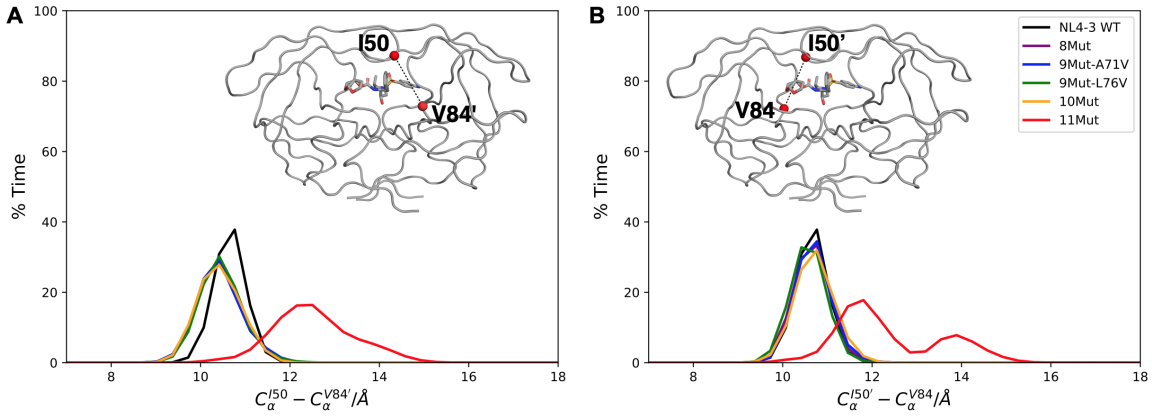


Figure 3.11 Distance between the C_α atoms of A) I50-V84' and B) I50'-V84 from MD simulations indicate that the flaps of the 11Mut are in a semi-open conformation.

3.3.8 Inhibitor fluctuations increase and interactions are destabilized due to resistance mutations

In the cocrystal structures, conformation of DRV was highly similar with the most notable alterations occurring at the P1' moiety [**Figure 3.12**]. To investigate how the dynamics of DRV may vary, the RMSF of every DRV heavy atom from the MD simulations was calculated and grouped according to the four DRV moieties [**Figure 3.13**]. In complex with WT protease DRV was stable, with RMSF below 0.75 Å except for the P1' moiety. In 1Mut, 2Mut and 4Mut variants fluctuations of DRV moieties were similar to those in WT complex, with few minor exceptions at the P1' and P1 moieties (data not shown). When bound to all the highly mutated variants, DRV had increased flexibility, especially at the P1 and P2 moieties [**Figure 3.13**]. In 10Mut complex, the P2 bis-THF moiety had high RMSF compared to WT and all other variants. This moiety makes several hydrogen bonds to protease active site residues and contributes significantly to the high potency of DRV. In complex with the 11Mut variant, the entirety of DRV experienced significant flexibility with the P1 moiety showing the greatest RMSF. Overall, resistance mutations in the protease variants resulted in increased flexibility and higher fluctuations of DRV, suggesting the inhibitor binding and intermolecular interactions are destabilized.

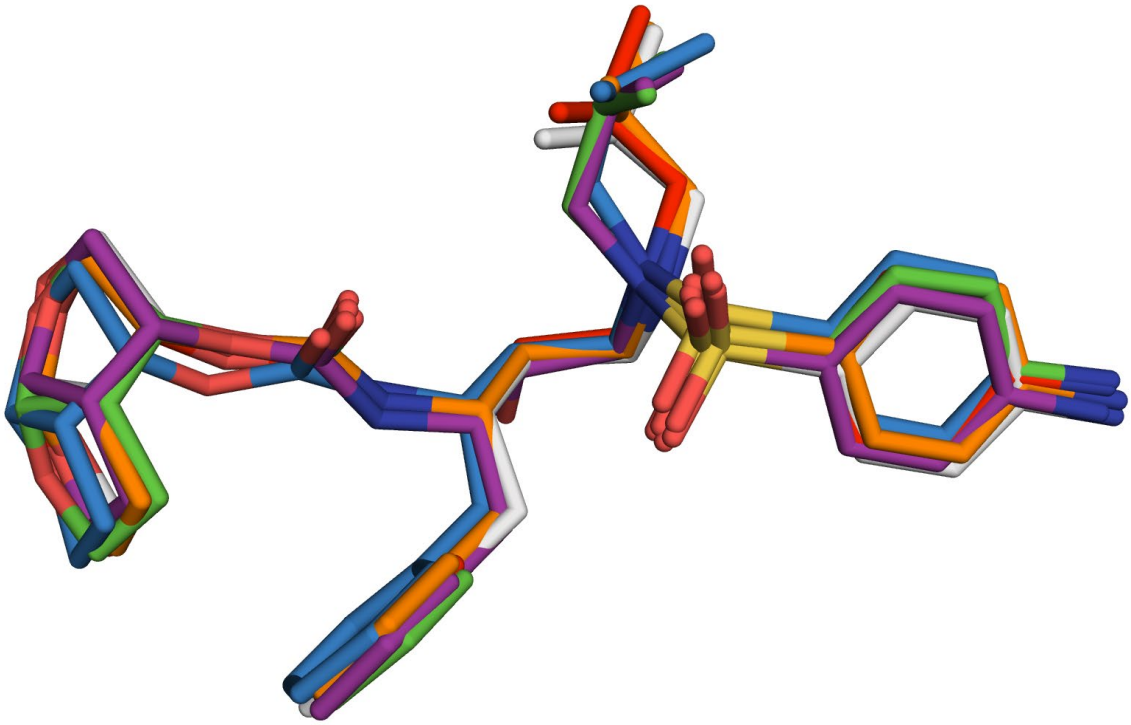


Figure 3.12 Alignment of DRV from crystal structures bound to WT (grey), 8Mut (purple), 9Mut-A71V (blue), 9Mut-L76V (green), 10Mut (orange) and 11Mut (red) variants.

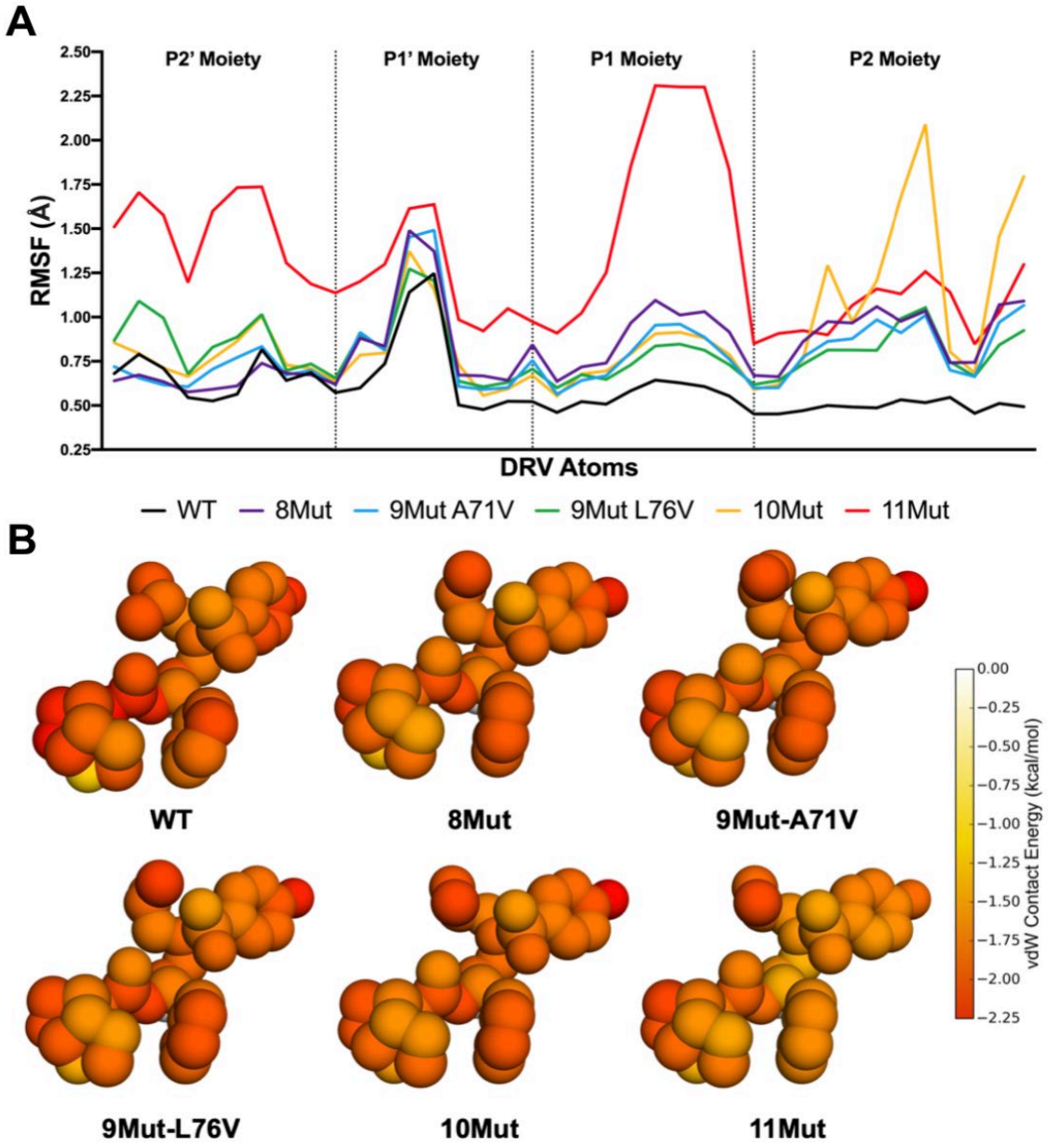


Figure 3.13 A) Root-mean-square fluctuation (RMSF) of DRV atoms grouped by moiety monitored during MD simulations bound to WT and resistant HIV-1 protease variants (color scheme as in Figure 1). B) Packing around DRV in complex with WT protease and resistant variants. Total per atom protease–DRV vdW contact energies mapped onto the respective DRV crystal structure, with red indicating more contacts.

Underlying the high potency of DRV against WT HIV-1 protease is an extensive hydrogen bonding network and strong vdW interactions with active site residues. In highly mutated variants these interactions may be perturbed, as suggested by increased DRV fluctuations. The vdW packing around DRV in complex with 8Mut, 9Mut-A71V, 9Mut-L76V, and 10Mut decreased slightly around the P2 bis-THF moiety compared to WT [**Figure 3.13B**]. In 11Mut variant, packing around DRV was significantly reduced at the P1 phenyl, P1' isobutyl, and P2' aniline moieties. In the cocrystal structure bound to WT protease, DRV makes a number of hydrogen bonds with the protease backbone and a water-mediated hydrogen-bonding network with the flaps [**Figure 3.14A**]. All these hydrogen bonds were stable and highly conserved throughout the MD simulations [**Figure 3.14B**]. In the highly mutated variants, frequency of hydrogen bonds during the MD simulations decreased relative to WT complex, with a severe loss in the hydrogen bond between D25' and the central hydroxyl of DRV. In 8Mut and 9Mut-A71V variants, DRV hydrogen bonds with P2 bis-THF and flaps destabilized 8–17%, and frequency of the hydrogen bond between P2' aniline and backbone nitrogen of D30' decreased 18%. In 9Mut-L76V and 10Mut variants there was a further decrease in the frequency of these hydrogen bonds. In 11Mut variant, water-mediated hydrogen bonding network with the flaps was severely impacted, where two hydrogen bonds were completely lost and the other two maintained only 25% during the MD simulations. The hydrogen bond between D25 and the central hydroxyl was reduced to 62% where it was previously maintained at 83–99% for

WT and all other resistant variants. Additionally, the P2' aniline hydrogen bond with D30' was significantly reduced to only 9%, an 85% decrease compared to WT. Thus, the impact of distal mutations in these highly mutated and resistant variants propagated to the active site to destabilize inhibitor vdW and hydrogen bonding interactions with the protease.

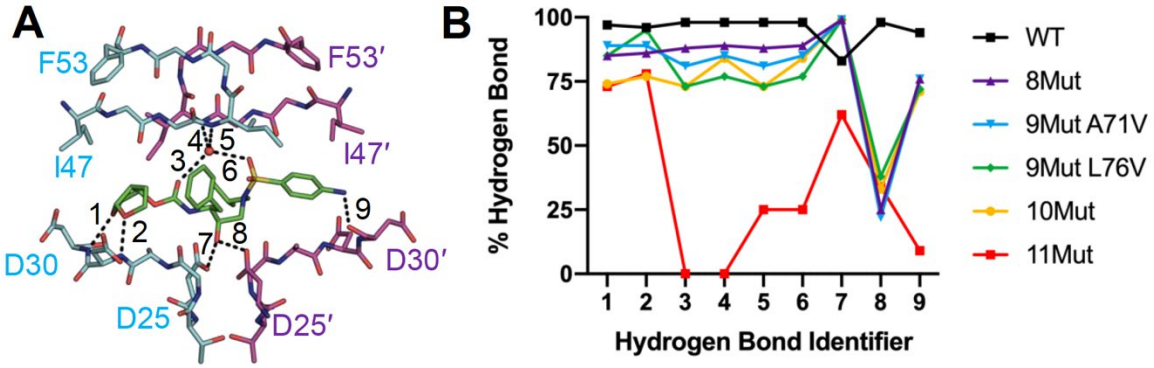


Figure 3.14 Resistant HIV-1 protease variants lose hydrogen bonding with inhibitor DRV. A) Hydrogen bonds with DRV (green sticks) indicated as black dashed lines on the WT protease crystal structure. The “flap water” molecule is depicted as a red sphere. Numbers correspond to hydrogen bonds shown in panel B. B) Hydrogen bonding frequencies from MD simulations. The water-mediated hydrogen-bonding network between the protease flaps and DRV is severely disrupted in 11Mut.

3.3.9 Protease–DRV cross-correlations of fluctuations are lost with increasing resistance

Cross-correlation of collective motions between a ligand and target protein can distinguish tight and weak binders, as we previously reported[55, 104, 139]. For HIV-1 protease, tight binding inhibitors have greater correlation of fluctuations with proximal protease residues, namely the active site and flaps. Contrary to tight binders, weak binding inhibitors lose correlations with these residues corresponding to disrupted intermolecular interactions[104]. Cross-correlations between the fluctuations of DRV and protease residues were calculated from the MD simulation trajectories. As expected, DRV fluctuations were highly correlated with catalytic residues especially at chain A, flaps tips (49-51/49'-51'), and residues 26-31 adjacent to the catalytic D25 [**Figure 3.15A**]. DRV also had moderate correlations with residues 81–87, while the flap elbows and 70's beta-sheets were anti-correlated. In 8Mut, DRV maintained correlation with residues 49'-50' but decoupled from chain A flap tip and the active site residues [**Figure 3.16**]. While 9Mut-A71V restored the lost correlations of DRV with the catalytic residues, the flap tips became further decoupled. In 9Mut-L76V and 10Mut, DRV fluctuations were further decoupled from the flap tips, but in both variants DRV somewhat maintained correlation with catalytic residues [**Figure 5**]. In 11Mut variant, positive correlations with DRV were almost completely lost, and the inhibitor established anti-correlations with various regions of the protein (chain B flap, S2' subsite, chain B 70's beta-sheet and 80's loop). Disruptions in the correlations between DRV and

11Mut fluctuations are consistent with increased resistance and weaker intermolecular interactions.

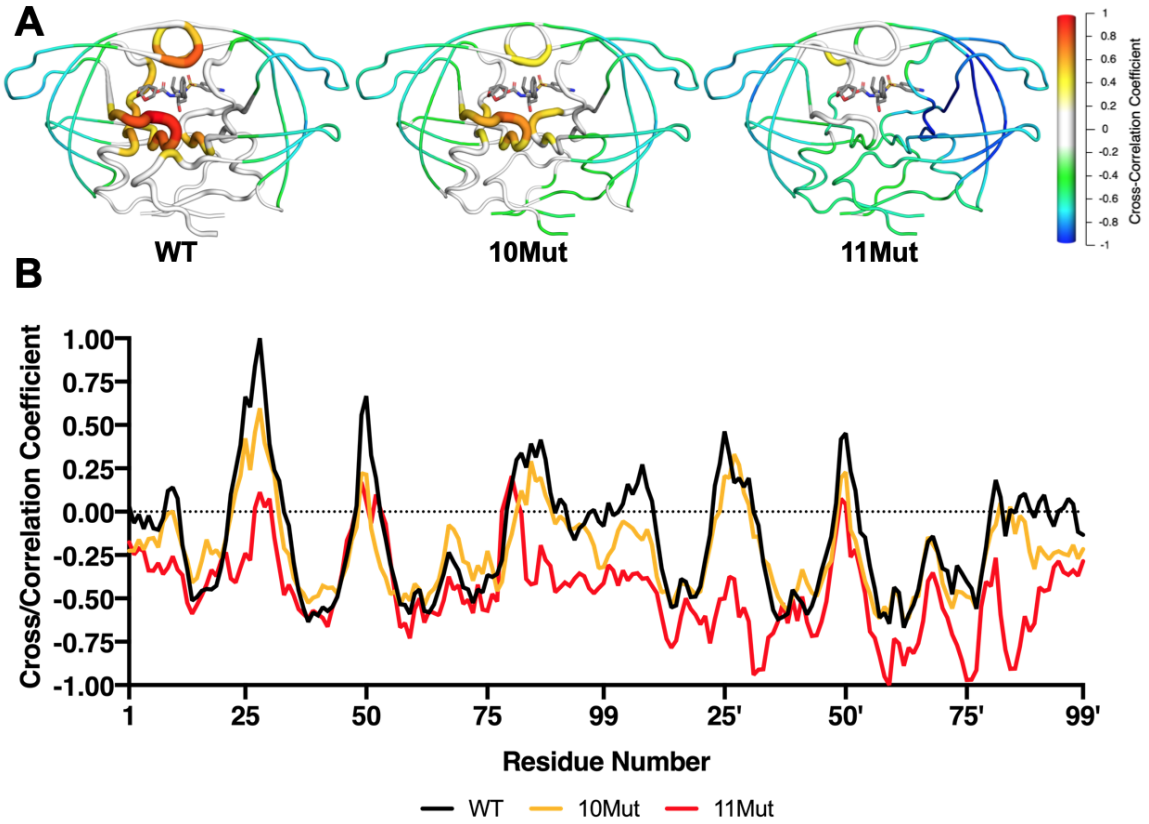


Figure 3.15 Cross-correlation of fluctuations between HIV-1 protease and DRV from MD simulations. A) Mean cross-correlation coefficient for each residue mapped onto protease structure, shown as cartoon putty for WT, 10Mut, and 11Mut variants. B) Cross-correlations for WT, 10Mut, and 11Mut plotted. As resistance mutations accumulate, positive correlations decrease while negative correlations increase.

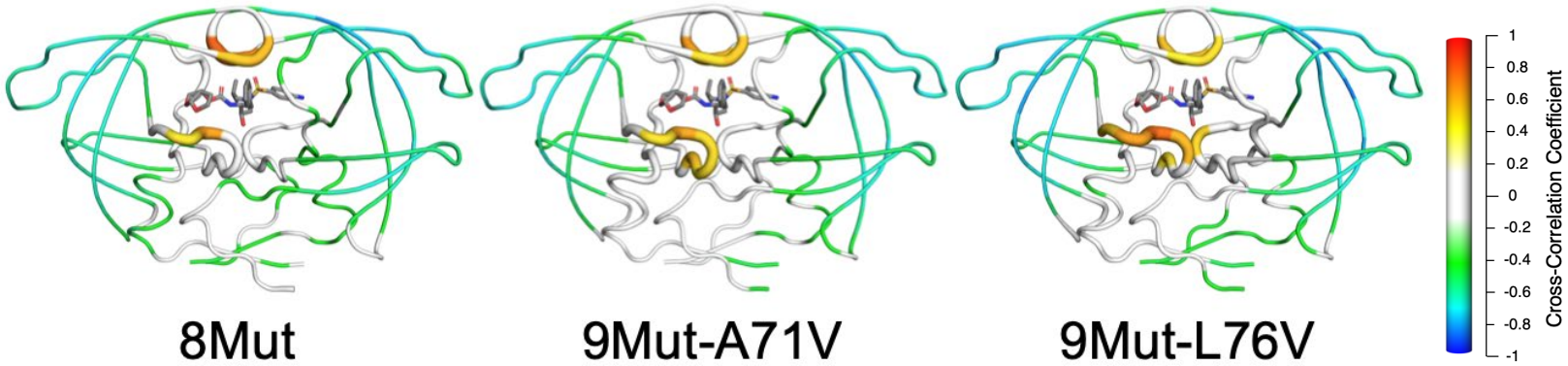


Figure 3.16 Cross-correlation of DRV fluctuations with 8Mut, 9Mut-A71V, or 9Mut-L76V from MD simulations. Mean cross-correlation coefficient for each residue mapped onto protease structure, shown as cartoon putty.

3.3.10 Distal mutations impact the dynamic ensemble to confer resistance

The existing paradigm reasons that distal mutations play a compensatory role in HIV-1 protease, acting to restore catalytic efficiency lost by the accumulation of mutations. However, while certain distal mutations indeed play a compensatory role, they also confer resistance as demonstrated above. We reported previously that distal mutations may propagate their effects to the active site via altering the conformational dynamics of the inhibitor–protease complex, via hydrophobic sliding[52, 138] and the internal hydrogen bonding network[54]. Here we used Jensen-Shannon divergence (JSD) to analyze perturbations in the dynamic ensemble in response to mutations, where protein backbone and side chain dihedral angles were compared between variants during MD simulations. We focused on the impact of 3 additional non-active site mutations, from 8Mut to 11Mut, which results in almost 60-fold increase in resistance.

Consistent with the C_α distance difference analysis, A71V caused only local perturbations in the dynamic ensemble in the 70's beta-sheets [**Figure 3.17A**]. L76V also perturbed the dynamics of the 70's loop (chain B) but also had distal effects on the dihedrals of residues in both flaps [**Figure 3.17B**]. In combination the effects of A71V and L76V were mostly additive, but together they also perturbed the side chain dihedrals of catalytic D25/25' residues (ϕ of D25 and χ^2 of D25') [**Figure 3.17C**]. However, the alterations in D25/D25' were subtle and did not significantly change the hydrogen bonds with DRV [**Figure 3.14**]. Comparison of 9Mut variants with 10Mut confirmed the earlier observation that addition of L76V

mostly impacted the flaps [**Figure 3.17D**], while A71V had local effects on the proximal 70's beta-sheets [**Figure 3.17E**]. Lastly, addition of I54L in 11Mut caused considerable changes to the dynamic ensemble compared to 10Mut, both proximal and distal. Most noticeably, side chain dihedrals of catalytic residues were impacted (Ψ and χ^2 of D25 and χ^2 of D25'), which is consistent with decreasing hydrogen bonding between the catalytic residues and the central hydroxyl of DRV [**Figure 3.14**]. In addition to changes at the active site, the dynamic ensemble of the flaps and 80's loop of chain B were affected. These results indicate that distal resistance mutations propagate their effects throughout the protein structure, which can be captured as alterations to the dynamic ensemble of protein dihedral angles.

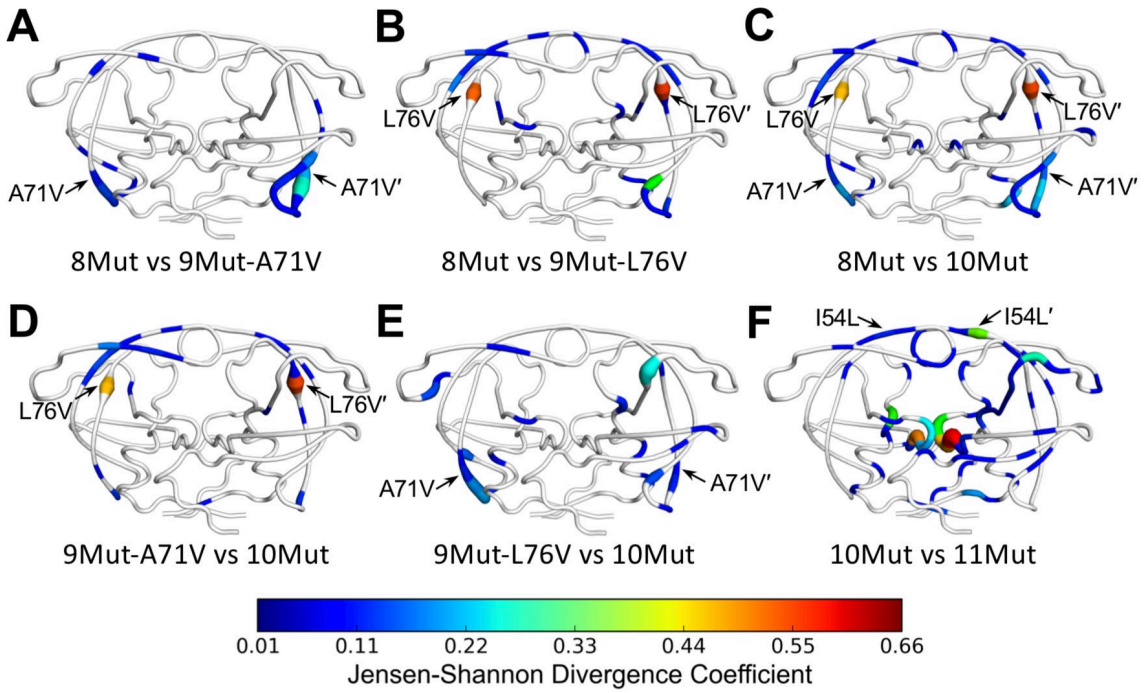


Figure 3.17 Comparison of protease dynamics via Jensen-Shannon divergence of dihedral angles between (A) 8Mut and 9Mut-A71V, (B) 8Mut and 9Mut-L76V, (C) 8Mut and 10Mut, (D) 9Mut-A71V and 10Mut, (E) 9Mut-L76V and 10Mut, and (F) 10Mut and 11Mut from MD simulations, mapped onto protease structure. Tube thickness and warmer colors indicate larger perturbation of the dynamic ensemble. Residues in white had no significant difference between the two variants. Arrows denote location of mutation(s) different between the variants compared.

3.4 Discussion

Drug resistance occurs when the target evolves and gains mutations to thwart inhibition while still maintaining biological function. We investigated a highly evolved and mutated variant of HIV-1 protease that accumulated 11 mutations to confer high-level resistance to DRV. We systematically dissected the mutations in this variant by generating 8 different subset variants, including separating active site mutations from distal mutations and leveraging a double mutant cycle analysis [**Figure 3.1A**]. Integrating activity and inhibition assays, high-resolution crystal structures, and molecular dynamics simulations, we determined how active site and distal mutations contribute to drug resistance. Our results show that resistance progressively increased as the mutations were added, while catalytic activity was maintained. Mutations were selected both inside and outside the active site where the inhibitor binds. We found that the distal mutations play a pivotal role in conferring drug resistance, as proximal mutations alone cannot explain the observed near μM resistance. Double mutant cycle analysis revealed that two such distal mutations are interdependent, with A71V restoring catalytic activity while L76V decreasing inhibition. Our findings challenge the convention that distal mutations are merely compensatory and have secondary or minimal contribution to resistance. On the contrary, we demonstrated that distal mutations act interdependently through complex molecular mechanisms to confer high levels of resistance.

In addition to avoiding inhibition, the mutated protease needs to continue processing viral substrates to allow viral replication. The highly mutated variants (8Mut, 10Mut, 11Mut) detected from viral passaging had somewhat compromised enzymatic activity relative to WT protease, although still allowed for robust viral replication. Unlike the detectable variants, both 9Mut-A71V and 9Mut-L76V variants had severely compromised catalytic activity. The protease must evolve through one of the 9Mut intermediates en route from 8Mut to 10Mut variant, but the virus cannot efficiently replicate to populate detectable levels. A71V is a well-known compensatory mutation that is often observed together with I50V and was reported to restore catalytic activity lost due to this primary mutation[76, 91]. The 8Mut variant does not contain I50V, and when added to the 8Mut variant A71V actually obliterated catalytic activity (9Mut-A71V). In the presence of the L76V mutation though the catalytic activity was restored (10Mut). Thus, the effect of A71V on catalytic activity is context dependent, and may be beneficial or detrimental depending on the background mutation(s).

While resistance due to primary active site mutations are rather straightforward to explain especially using the substrate envelope, how combinations of mutations including distal mutations act interdependently to confer resistance requires further investigations. The crystal structures showed consistent losses in protein-inhibitor vdW contacts, differences in internal distances, and subtle alterations in the binding conformation of DRV. However, given the high levels of resistance and orders of magnitude change in inhibition

constants, we expected further and amplified impact on the conformational dynamics of the protease–inhibitor complex. MD simulations highlighted the extent of unfavorable protease–DRV interactions with increasing mutation accumulation. Highly mutated variants resulted in increased protein and inhibitor fluctuations and, most importantly, reduced cross-correlated motions. We previously found that the loss in correlated inhibitor–target fluctuations correlated with reduced inhibitor potency for HCV NS3/4A protease[104] and analogous inhibitors binding to WT HIV-1 protease[55]. We find here that the same loss exists for highly mutated and resistant variants of HIV-1 protease. In addition, the hydrogen bonds and vdW interactions that contribute to DRV's pM potency in WT HIV protease were greatly diminished. Alterations in the dynamics of the protease–inhibitor complex are crucial in the molecular mechanisms of resistance, and thus the molecular dynamics need to be investigated in addition to static crystal structures structural to elucidate how the effect of distal mutations can be propagated to cause alterations in inhibitor binding.

Here we develop an alternative approach to previously explored methods to elucidate how distal mutations propagate dynamic changes throughout the protease. Expanding upon hydrophobic sliding[52, 138] and internal hydrogen bonding networks[54], quantitative comparison of protein dihedral angles can be used to examine dynamic structural changes. As dihedral angles from MD simulations comprise highly similar probability distributions, a rigorous statistical analysis involving Jensen-Shannon divergence implementation is needed to determine

distinct changes. This approach previously quantified differences in long-range dynamics of proteins with and without a ligand bound[140]. Dihedral angle comparison illustrated the propagation of dynamic effects due to distal mutations throughout the protein, including impact of A71V/L76V on the catalytic residues. In the most mutated and most resistant variant 11Mut, addition of I54L mutation conferred many additional changes to the active site and flaps while also propagating changes throughout the protease structure. This analysis demonstrated how mutations cause dynamic changes that propagate to distal regions of the protein, and thus how considering local changes to the static structure is not sufficient to elucidate molecular mechanisms of drug resistance.

Our work provides evidence that distal mutations do indeed contribute to drug resistance through complex dynamic processes, significantly altering protease dynamics and perturbing inhibitor binding. In other HIV-1 protease variants and, perhaps more significantly, other drug targets where distal mutations are observed in clinically relevant resistant variants, the possibility that distal mutations directly confer resistance cannot be ignored. This also has the implication that drug design needs to consider both structural and dynamic changes to the protein–inhibitor complex to be able to effectively target these variants. The increasing threat of drug resistance in many clinically relevant systems highlights the need to characterize resistance mutations, including distal mutations and combinations of mutations that interdependently confer resistance through complex dynamic resistance mechanisms.

3.5 Methods

3.5.1 Protease Gene Construction.

Protease gene construction was carried out as previously described[40, 104]. In short, the WT and 10Mut protease variant genes were purchased on a pET11a plasmid with codon optimization for protein expression in *E. coli* (Genewiz). The remaining variants were constructed using QuikChange site-directed mutagenesis (Genewiz). A Q7K mutation was included to prevent autoproteolysis[101].

3.5.2 Protein Expression and Purification

The expression, isolation, and purification of WT and mutant HIV-1 proteases used for all assays and crystallization were carried out as previously described[40, 99]. Briefly, the gene encoding the desired HIV protease was subcloned into the heat-inducible pXC35 expression vector (ATCC) and transformed into *E. coli* TAP-106 cells. Cells grown in 6 L of Terrific Broth were lysed with a cell disruptor twice, and the protein was purified from inclusion bodies[102]. The inclusion body centrifugation pellet was dissolved in 50% acetic acid followed by another round of centrifugation at 19K rpm for 30 minutes to remove impurities. Size exclusion chromatography was carried out on a 2.1-L Sephadex G-75 superfine (Sigma Chemical) column equilibrated with 50% acetic acid to separate high molecular weight proteins from the desired protease. Pure fractions of HIV protease were refolded into a 10-fold dilution of refolding buffer [0.05 M sodium acetate at pH 5.5, 5% ethylene glycol, 10% glycerol, and 5 mM DTT]. Folded protein was concentrated down to 0.5–3 mg/mL and stored. The stored protease was used in

K_M and K_i assays. For crystallography, an additional purification step was performed with a Pharmacia Superdex 75 FPLC column equilibrated with refolding buffer. Protease fractions purified from the size exclusion column was concentrated to 1–2 mg/mL using an Amicon Ultra-15 10-kDa device (Millipore) for crystallization.

3.5.3 K_M Assay

K_M values were determined as previously described[37, 89-90]. Briefly, a 10-amino acid substrate containing the natural MA/CA cut site with an EDANS/DABCYL FRET pair was dissolved in 8% DMSO at 40 nM and 6% DMSO at 30 nM. The 30 nM substrate was 4/5 serially diluted from 30 nM to 6 nM. HIV protease was diluted to 120 nM and, using a PerkinElmer Envision plate reader, and 5 µL were added to the 96-well plate to obtain a final concentration of 10 nM. K_M assays were conducted in triplicate. The fluorescence was observed with an excitation at 340 nm and emission at 492 nm and monitored for 200 counts, for approximately 23 minutes. FRET inner filter effect correction was applied as previously described[103]. Data corrected for the inner filter effect was analyzed with Prism7. Fluorescence over time was used to calculate initial velocities by fitting the data to the one-phase association equation. Initial velocities were then fit by non-linear regression to the Michaelis-Menten equation to determine k_{cat} and K_M and the standard error of the mean (SEM) was calucated. The error from these two measurements was propagated to calculate the overall error of k_{cat}/K_M.

3.5.4 K_i Assay

Enzyme inhibition constants (K_i values) were determined as previously described[37, 89-90]. Briefly, in a 96-well plate, DRV was 3/5 serially diluted from 2000 nM for 8Mut, 1000 nM for 9Mut-A71V, 5000 nM for 9Mut-L76V, 5000 nM for 10Mut, or 10,000 nM for 11Mut. All samples were incubated with 5 nM protein for 1 hour. A 10-amino acid substrate containing an optimized protease cut site[90], purchased from BAchem, with an EDANS/DABCYL FRET pair was dissolved in 4% DMSO at 120 mM. Using a PerkinElmer Envision plate reader, 5 µL of the 120 mM substrate were added to the 96-well plate to a final concentration of 10 mM. Fluorescence was observed with an excitation at 340 nm and emission at 492 nm and monitored for 200 counts. Data was analyzed with Prism7, as described in Chapter 2.

3.5.5 Gag Polyprotein Cleavage Assay

The pET28a plasmid containing full length $Pr55^{Gag}$-TEV-His construct was a kind gift from Maria Bewley and John Flanagan. Protein expression and purification was done as described by Bewley et al. (*Protein Expr Purif*, 2017; 130:137-145) and consists of the removal of DNA and ammonium sulfate precipitation. On the day of the assay, ammonium sulfate pellets were dissolved in resuspension buffer (10 mM HEPES pH 7.5, 500 mM NaCl, 0.1 mM TCEP and 0.1 mM EDTA) and diluted 1:5 with the same buffer without NaCl. The final solution was centrifuged at 20k x *g* for 20 min. Cleavage of $Pr55^{Gag}$ polyprotein by HIV-1 protease was monitored by SDS-PAGE of cleavage products visualized by Coomassie staining. Samples were taken from the reaction mixture at designated time points, and the

cleavage reaction was quenched by adding gel running buffer containing SDS and boiling for 2 min.

3.5.6 Crystallography

Discovery of the condition producing hexagonal cocrystals of DRV bound to a highly mutated variant (8Mut) without seeding was achieved using the Protein Complex Suite Screen (Qiagen), in well G5, consisting of 1 M ammonium sulfate and 1 M potassium chloride and 0.1 M HEPES buffer at pH 7 with a protease concentration of 1.2 mg/mL and 5-fold molar excess of DRV. With previously generated orthorhombic and hexagonal cocrystals in hand, reproducible cocrystals of DRV bound to all variants were achieved as previously described[37]. Briefly, all cocrystals were grown at room temperature by hanging drop vapor diffusion method in 24-well VDX hanging-drop trays (Hampton Research) with protease concentrations between 1.0 to 2.4 mg/mL with 3, 5, or 10-fold molar excess of DRV. Crystallization drops were set with the reservoir solution consisting of 18–26% (w/v) ammonium sulfate and 0.1 M bis-Tris-methane buffer at pH 5.5 set with 2 mL of well solution and 1 mL protein and microseeded with a cat whisker. Crystal morphology and space group was entirely dependent on the microseeds. Diffraction quality crystals were obtained within 1 week. As data was collected at 100 K, cryogenic conditions contained the precipitant solution supplemented with 25% glycerol. For direct structural analysis without possible crystal lattice contact bias, structure of WT protease bound to DRV was also generated and determined in the hexagonal space group.

3.5.7 Data Collection and Structure Solution

Diffraction data were collected and solved as previously described[37, 40, 104]. Diffraction quality crystals were flash frozen under a cryostream when mounting the crystal at our in-house Rigaku Saturn944 X-ray system. The data for 10Mut cocrystal structure was collected at the Advanced Photon Source at the Argonne National Laboratory, beamline 19-ID. All cocrystal diffraction intensities were indexed, integrated, and scaled using HKL3000[105]. Structures were solved using molecular replacement with PHASER[106]. Model building and refinement were performed using Coot[107] and Phenix[108]. During refinement, all crystals utilized optimized stereochemical weights and non-crystallographic symmetry operators. Hexagonal crystals grew as pseudo-merohedral twins and were solved with a twin law applied (h,-h-k,-l). Ligands were designed in Maestro and the output sdf file was used in the Phenix program eLBOW[109] to generate the cif file containing atomic positions and constraints necessary for ligand refinement. Iterative rounds of crystallographic refinement were carried out until convergence was achieved. To limit bias throughout the refinement process, 5% of the data were reserved for the free R-value calculation[110]. MolProbity[111] was applied to evaluate the final structures before deposition in the PDB. Structure analysis, superposition, and figure generation were performed using PyMOL[141]. X-ray data collection and crystallographic refinement statistics are presented in **Table 3.2**.

3.5.8 Internal Distance Analysis of Crystal Structures

Distance-difference matrices were generated as previously described[113] to reveal structural changes between inhibitor–protease pairs.

3.5.9 Intermolecular vdW Contact Analysis of Crystal Structures

To calculate the intermolecular van der Waals (vdW) interaction energies the crystal structures were prepared using the Schrodinger Protein Preparation Wizard[114]. Hydrogen atoms were added, protonation states determined and the structures were minimized. The protease active site was monoprotonated at D25. Subsequently, forcefield parameters were assigned using the OPLS2005 force field[115]. Interaction energies between the inhibitor and protease were estimated using a simplified Lennard-Jones potential, as previously described in detail[116]. Briefly, the vdW energy was calculated for pairwise interactions depending on the types of atoms interacting and the distance in-between. For each protease residue, the change in vdW interactions relative to a WT complex in the same space group was also calculated for each mutant structure.

3.5.10 System Preparation

High-resolution crystal structures were prepared using the Protein Preparation Wizard from Maestro within the Schrodinger Suite[114] as previously described[85]. Briefly, cocrystallized phosphates were removed, missing atoms were added using Prime[142], and PROPKA[143-144] was used to determine the protonation state of side chains at pH 7.0. Lastly, the structure was minimized to a convergence criterion of 0.3 Å using Impref[145].

3.5.11 Molecular Dynamics Simulations

The prepared systems were placed in a cubic TIP3P water box measuring 12 Å on each side. MD simulations were carried out as previously described[85] using Desmond software suite within Schrodinger[146]. Briefly, chloride ions were used to neutralize the system and 0.15 M salt were added using sodium and chloride ions. The OPLS3 force field was used to parameterize the ligand and protein. Prior to starting the 100 ns MD simulations, the solvated system was minimized using the stepwise procedure described previously[85]. Triplicates of 100 ns simulations for WT, I84V, 2Mut, 4Mut, 8Mut, 9Mut-A71V, 9Mut-L76V, 10Mut, and 11Mut each with a randomized velocity were started using the protocol previously developed[85]. The root-mean-square deviation (RMSD) and root-mean-square fluctuation (RMSF) of C_α atoms were calculated by utilizing the simulation interactions diagram from Maestro within the Schrodinger Suite.

3.5.12 Intermolecular vdW Contact Analysis of MD Trajectories

Intermolecular vdW interactions were calculated using a previously published protocol utilizing Lennard-Jones potential[40, 147]. The force field is not optimized for a sulfur violating the octet rule such as that found in the P2' moiety of DRV. As such, the packing around the P2' sulfur was calculated by averaging the vdW packing of the four adjacent atoms.

3.5.13 Cross-Correlations of Protease/DRV Fluctuations

The cross-correlation coefficients between the protease C_α atoms and DRV heavy atoms were calculated using a previously published protocol[55]. Briefly, the

atom fluctuations were determined for each MD simulation. The cross-correlation between atom pairs, such as atoms i and j, was determined according to

$$C_{i,j} = \frac{\langle \Delta R_i \cdot \Delta R_j \rangle}{\sqrt{\langle \Delta R_i^2 \rangle \langle \Delta R_j^2 \rangle}}$$

where ΔR_i and ΔR_j are the positional changes of atom i and j, respectively, and the angle brackets denote an ensemble average. The cross-correlation coefficient ranges from -1 to 1, where -1 represents complete anticorrelation between the atom pair, zero represents no correlation, and 1 represents complete correlation. The cross-correlation values were calculated using an in-house script and mapped onto the protease structure by replacing the B-factors in the PDB coordinate file, using PyMOL[141].

3.5.14 Distance-Difference Matrices

The distances between all C_α atom pairs for a given structure were calculated as a *198×198* matrix for all frames in the trajectory. The distance was calculated for each C_α atom pair, giving the mean distance over the trajectory. The distance difference matrix was produced by subtracting each mean distance in the reference structure from the corresponding mean distance in the given structure. The overall difference for each residue was then calculated by taking the average of the absolute values of all the 198 distance differences involving that residue. PyMOL[141] was used for visualization of distance differences where the protease backbone was represented as a cartoon-putty with increasing thickness and warmer color for increasing deviation.

3.5.15 Jensen-Shannon Divergence Analysis of Dihedral Angles

Over the MD simulations, all the ϕ, ψ, χ dihedral angles of protease residues were measured. These dihedral angle measurements were used as input for the MutInf software package[140]. Utilizing this package, the Jensen-Shannon divergence (JSD), which analyzes the difference between two probability distributions, was calculated. An α of 0.05 was used as a threshold to filter out non-statistically significant differences in distributions, where a p-value > α was considered non-statistically significant. In such cases, the JSD was set to zero. For visualization purposes, using PyMOL[141], if a residue had multiple dihedral angles with a calculated JSD, the greatest value was selected.

3.6 Acknowledgements

This research was supported by National Institute of General Medical Sciences P01 GM109767. Some of the crystallographic data was collected at Argonne National Laboratory, Structural Biology Center (SBC) at the Advanced Photon Source. SBC-CAT is operated by UChicago Argonne, LLC, for the U.S. Department of Energy, Office of Biological and Environmental Research under contract DE-AC02-06CH11357. We thank the beamline specialists at 19-ID for their help in data collection. This work was supported by the National Institute of General Medical Sciences, National Institutes of Health (P01 - GM109767).

4 CHAPTER 4:

HIV-1 Protease Inhibitors Incorporating Stereochemically Defined P2′ Ligands to Optimize Hydrogen Bonding in the Substrate Envelope

4.1 Abstract

A structure-guided design strategy was used to improve the resistance profile of HIV-1 protease inhibitors by optimizing hydrogen bonding and van der Waals interactions with the protease while staying within the substrate envelope. Stereoisomers of 4-(1-hydroxyethyl)benzene and 4-(1,2-dihydroxyethyl)benzene moieties were explored as P2′ ligands providing pairs of diastereoisomers epimeric at P2′, which exhibited distinct potency profiles depending on the configuration of the hydroxyl group and size of the P1′ group. While compounds with the 4-(1-hydroxyethyl)benzene P2′ moiety maintained excellent antiviral potency against a panel of multidrug-resistant HIV-1 strains, analogues with the polar 4-(1,2-dihydroxyethyl)benzene moiety were less potent, and only the (*R*)-epimer incorporating a larger 2-ethylbutyl P1′ group showed improved potency. Crystal structures of protease-inhibitor complexes revealed strong hydrogen bonding interactions of both (*R*)- and (*S*)-stereoisomers of the hydroxyethyl group with Asp30′. Notably, the (*R*)-dihydroxyethyl group was involved in a unique pattern of direct hydrogen bonding interactions with the backbone amides of Asp29′ and Asp30′. The SAR data and analysis of crystal structures provide insights for optimizing these promising HIV-1 protease inhibitors.

4.2 Introduction

The HIV-1 protease is a major target for developing antiviral therapies against HIV-1. Drug discovery efforts, aided by structure-based drug design, have led to the development of nine FDA-approved HIV-1 protease inhibitors (PIs).[148] All approved HIV-1 PIs are competitive inhibitors, and most contain different dipeptide isosteres as transition state mimetics.[149] In addition to extensive hydrophobic interactions, PIs mainly rely on a number of direct and water-mediated hydrogen bonds with the protease for potency. The clinical efficacy of HIV-1 PIs has significantly improved since this drug class was first introduced in the mid-1990s. The development of second-generation PIs with improved potency, tolerability, and pharmacokinetic profiles, and the introduction of low-dose ritonavir as a pharmacokinetic booster in PI-based antiretroviral therapy (ART) has led to improved clinical outcomes. Moreover, PIs allow exceedingly high level of viral inhibition at clinical concentrations due to cooperative dose-response curves with high slopes.[150-151] These transition-state mimetic inhibitors represent the most potent anti-HIV-1 drugs.

Despite much success, PI-based therapies are associated with drawbacks that limit their effectiveness, including major side effects, unfavorable pharmacokinetics, and the acquisition of many viable multidrug-resistant (MDR) protease variants.[89, 152] Drug resistance remains a major issue, as the prevalence of PI resistance mutations increases with duration of ART and is much higher among PI-experienced patients.[153-154] Even the most effective PI-based regimens have been reported to select multiple resistance mutations in the protease, resulting in reduced virologic response.[153-156] Moreover, the prevalence of transmitted drug resistance continues to increase, which is an added

challenge in the treatment of HIV infections.[157-158] The emergence of MDR protease variants presents a challenge to structure-based drug design (SBDD), as the target is not a single protein but an ensemble of closely related proteases. The new PIs must maintain activity not only against existing MDR protease variants, but also against future variants that may emerge. These challenges require developing new SBDD strategies to optimize PI potency while avoiding resistance.

The design strategy to maximize interactions in the HIV-1 protease active site, particularly with the protein backbone atoms, has been quite successful leading to the development of the FDA-approved drug darunavir.[159-160] Darunavir (DRV, **1**) is a highly potent inhibitor of HIV-1 protease with a low pM inhibition constant (*K*i) and a high genetic barrier to resistance. The key structural feature that distinguishes DRV from previous generation HIV-1 PIs is the *bis*-tetrahydrofuran (*bis*-THF) moiety at the P2 position [**Figure 4.1**] that makes strong hydrogen bonding interactions with the backbone NH of Asp29 and Asp30 in the protease active site.[148, 159, 161] The application of the design strategy to maximize interactions with the backbone atoms of HIV-1 protease has recently led to the discovery of DRV analogues that maintain exceptional potency against MDR HIV-1 variants including DRV-resistant strains.[162-165]

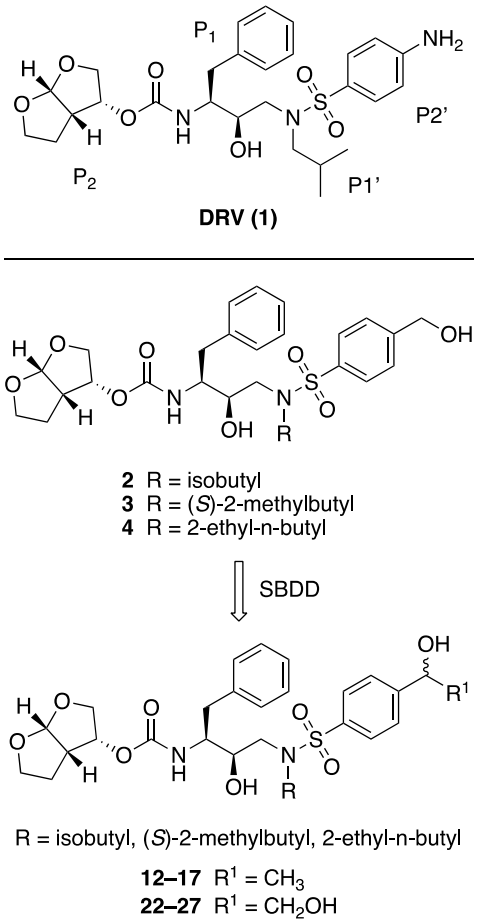


Figure 4.1 Structures of HIV protease inhibitors DRV (1), P2′ 4-(1-hydroxymethyl)benzene analogues (2–4) and designed compounds with stereochemically defined P2′ 4-(1-hydroxyethyl)benzene and (12–17) and 4-(1,2-dihydroxyethyl)benzene (22–27) moieties. The canonical nomenclature for inhibitor moiety position is indicated using DRV.

Among the strategies proposed to rationally design PIs against drug-resistant HIV-1 variants, the substrate envelope model provides a more comprehensive framework to incorporate drug resistance considerations into SBDD.[166-168] This structure-guided design strategy aims to optimize hydrogen bonding and van der Waals (vdW) interactions with the protease to improve potency against MDR protease variants, while constraining inhibitors within the substrate envelope to avoid resistance.[169] We previously used the substrate envelope-guided design strategy to develop a series of highly potent DRV analogues that maintained excellent antiviral potencies against a panel of clinically relevant MDR HIV-1 strains.[169] The X-ray cocrystal structures confirmed that the designed PIs optimally fill the substrate envelope and make enhanced vdW contacts and hydrogen bonding interactions with HIV-1 protease.[169-170]

Among these PIs, compounds **3** and **4** [**Figure 4.1**] with the 4-(hydroxymethyl)benzene P2′ moiety were identified as promising lead compounds due to distinct polar interactions of the benzylic hydroxyl group with the protease. Cocrystal structures of PIs **3** and **4** bound to wild-type HIV-1 protease revealed that the P2′ benzylic hydroxyl group is positioned close to the backbone NH of Asp29′ and Asp30′ in the S2′ subsite. The hydroxyl group makes a direct hydrogen bond only with the backbone NH of Asp30′, which is slightly closer than the NH of Asp29′ [**Figure 4.2**].[169-170] We reasoned that introduction of a small hydrophobic group such as a methyl at the benzylic position could shift the resulting secondary hydroxyl group closer to the backbone NH of Asp29′, allowing polar interactions with both residues, as well as additional vdW contacts in the S2′ subsite. Similarly, a hydroxymethyl group at the benzylic position could serve the

same purpose, with the primary hydroxyl group making additional polar interactions in the S2′ subsite of HIV-1 protease [**Figure 4.1**]. Optimal polar interactions of the P2′ moiety with the backbone NH of Asp29′ and Asp30′ could potentially mimic the strong hydrogen bonding interaction of the P2 *bis*-THF moiety, further improving potency against MDR HIV-1 variants. Previous efforts to identify an optimal P2′ moiety have only been partially successful, and none of the reported P2′ moieties make direct hydrogen bonding interactions with the backbone atoms of both Asp29′ and Asp30′.[162, 171-173]

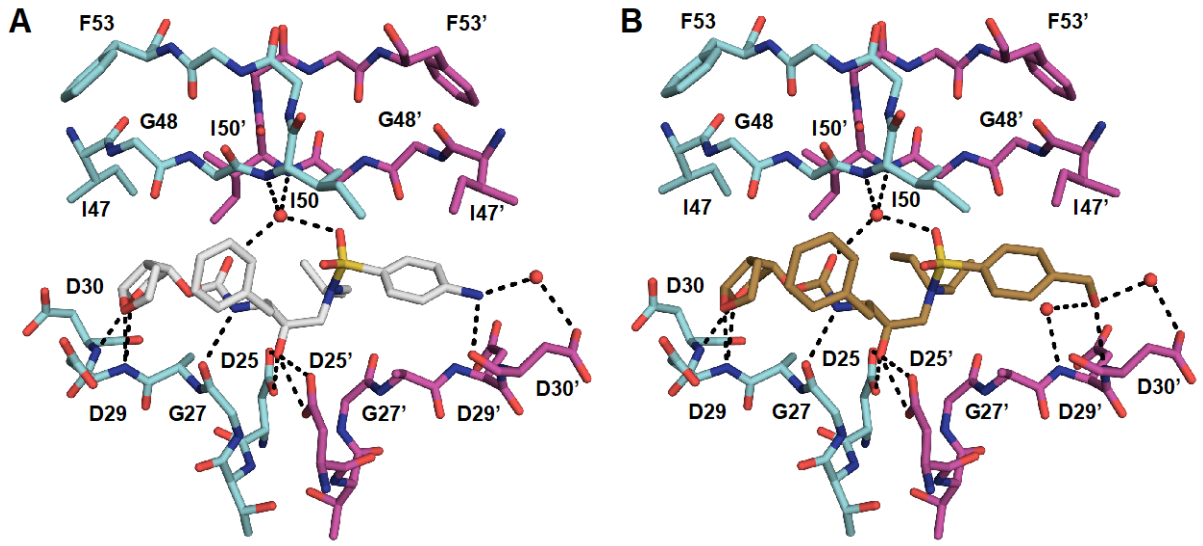


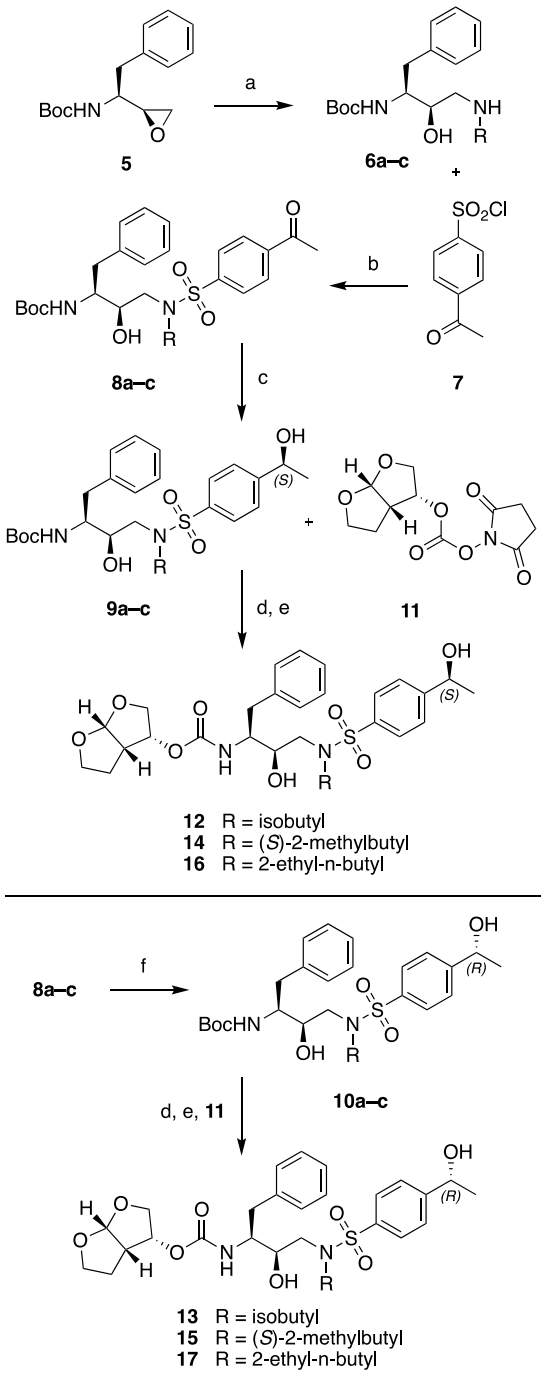
Figure 4.2 Comparison of hydrogen bonding network of (A) DRV (1) (PDB 6DGX) and (B) parent compound 4 (PDB 6OXQ). Hydrogen bonding pattern is similar throughout the active site except in the S2′ subsite. DRV makes a direct hydrogen bond with the backbone carbonyl of Asp30′ and water-mediated interactions with the side chain carboxylate group of the same residue. In contrast, inhibitor 4 makes a direct hydrogen bond with the backbone NH of Asp30′ and water-mediated hydrogen bonding with the backbone NH of Asp29′ and side chain carboxylate group of Asp30′.

On the basis of insights from the structural analysis and modeling, modifications of the P2′ moiety of PIs **3** and **4**, in combination with variations at the P1′ group, were explored to enhance inhibitor interactions in the S1′ and S2′ subsites. Here, we describe the substrate envelope-guided design, synthesis, evaluation of biochemical and antiviral potency, and crystal structure analysis of a series of novel HIV-1 PIs. The inhibitors were designed by incorporating stereoisomers of the 4-(1-hydroxyethyl)benzene and 4-(1,2-dihydroxyethyl)benzene moieties as novel P2′ ligands to enhance hydrogen bonding interactions in the S2′ subsite of HIV-1 protease. Structure-activity relationship (SAR) studies identified a number of compounds with improved potency profiles compared to DRV and the parent compounds against highly drug-resistant HIV-1 strains representing the spectrum of clinically relevant MDR viruses. We also report high-resolution crystal structures of all new compounds bound to wild-type HIV-1 protease revealing key interactions of the chiral 4-(1-hydroxyethyl)benzene and 4-(1,2-dihydroxyethyl)benzene moieties in the protease active site. A unique pattern of hydrogen bonding interactions was observed for compounds incorporating the (*R*)-4-(1,2-dihydroxyethyl)benzene as P2′ ligand, where both P2 and P2′ moieties make direct hydrogen bonds with the backbone NH of residues Asp29/29′ and Asp30/30′.

4.3 Chemistry

The synthesis of HIV-1 PIs with the (*S*)- and (*R*)-(1-hydroxyethyl)benzene moieties at the P2′ position is outlined in **Scheme 4.1**. We envisioned generating both stereoisomers of the hydroxyethyl group at the P2′ moiety from the corresponding 4-acetylbenzene derivatives by enantioselective reduction of the acetyl group using the Corey–Bakshi–Shibata (CBS) catalyst.[174] The 4-acetylbenzenesulfonamide intermediates **8a**–**c** were prepared from the commercially available epoxide **5** in two steps. Ring opening of the chiral epoxide **5** with selected amines provided the amino alcohols **6a**–**c**, which were reacted with 4-acetylbenzenesulfonyl chloride **7** using Na_2CO_3 as a base under biphasic conditions to afford the 4-acetylbenzenesulfonamide intermediates **8a**–**c** in excellent yield. Reduction of the acetyl group using BH_3-THF in the presence of the chiral catalyst (*R*)-CBS-Me provided the required intermediates with the (*S*)-(1-hydroxyethyl)benzene moiety **9a**–**c** with excellent enantioselectivity. After purification by flash chromatography, the products were recrystallized from a mixture of EtOAc and hexanes to ensure chiral purity. Removal of the Boc protecting group with TFA and reaction of the resulting amine salts with the *bis*-THF activated carbonate **11** provided the target compounds **12**, **14** and **16** with the (*S*)-4-(1-hydroxyethyl)benzene moiety at the P2′ position. Inhibitors with the corresponding (*R*)-4-(1-hydroxyethyl)benzene at the P2′ position were prepared similarly using the (*S*)-CBS-Me as the catalyst during the stereoselective reduction step and provided the desired inhibitors **13**, **15** and **17**.

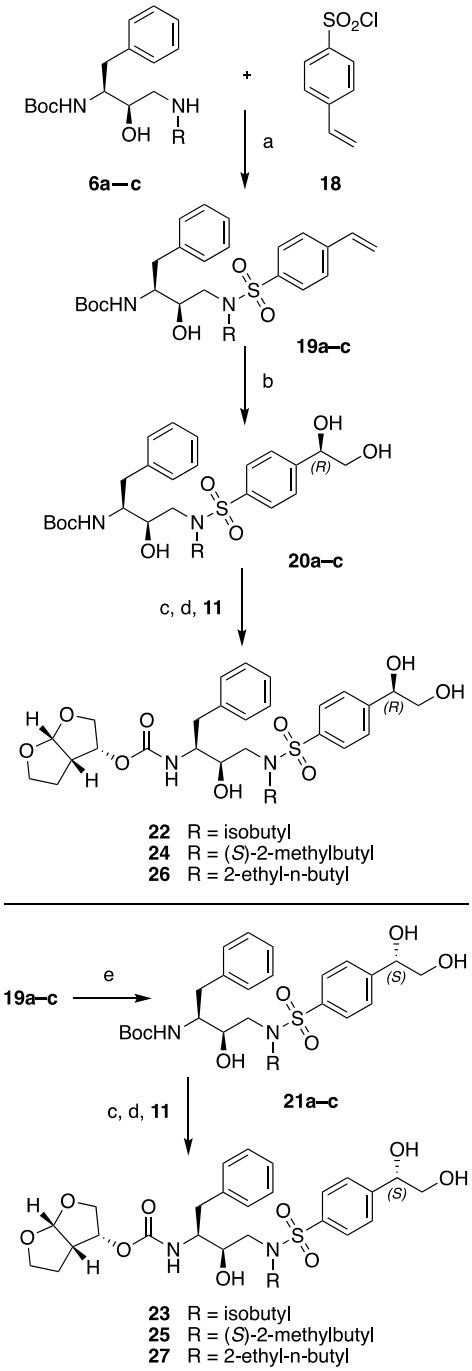
Scheme 4.1 Synthesis of protease inhibitors incorporating (*S*)- and (*R*)-4-(1-hydroxyethyl)benzene as P2′ ligands.



Reagents and conditions: (a) RNH_2, EtOH, 70 °C, 3 h, 79–91%; (b) Na_2CO_3, EtOAc, H_2O, RT, 18 h, 98–100%; (c) *R*-CBS-Me, BH_3-THF (1 M), THF, 0 °C to RT, 3 h, 57–85%; (d) TFA, CH_2Cl_2, RT, 1 h; (e) DIEA, CH_3CN, RT, 24 h, 73–91%; (f) *S*-CBS-Me, BH_3-THF (1 M), THF, 0 °C to RT, 3 h, 47–87%.

The protease inhibitors with the (*R*)- and (*S*)-4-(1,2-dihydroxyethyl)benzene moieties at the P2′ position were prepared from the corresponding styrene derivatives by Sharpless asymmetric dihydroxylation using the AD-mix catalyst[175-176] as outlined in **Scheme 4.2**. Briefly, reactions of amino alcohols **6a**–**c** with 4-vinylbenzenesulfonyl chloride **18** using Na_2CO_3 as a base under biphasic conditions provided the 4-vinylbenzenesulfonamide intermediates **19a**–**c**. The asymmetric dihydroxylation reaction using AD-mix-*β* as a chiral catalyst proceeded smoothly and provided the required intermediates with the (*R*)-4-(1,2-dihydroxyethyl)benzene moiety **20a**–**c**. In all cases the (*S*)-epimer was not detected by ^{1}H NMR; however, the products were recrystallized to provide enantiomerically pure intermediates **20a**–**c**. Analogous to the compound series with the 4-(1-hydroxyethyl)benzene P2′ moiety, Boc deprotection and the reaction of the resulting amine salts with the *bis*-THF carbonate **11** provided the target compounds **22**, **24** and **26** with the (*R*)-4-(1,2-dihydroxyethyl)benzene moiety at the P2′ position. Inhibitors with the corresponding (*S*)-4-(1,2-dihydroxyethyl)benzene moiety at the P2′ position were prepared in a similar fashion using the AD-mix-*α* as the chiral catalyst in the asymmetric dihydroxylation reaction and provided the target inhibitors **23**, **25** and **27**.

Scheme 4.2 Synthesis of protease inhibitors incorporating (S)- and (R)-4-(1,2-dihydroxyethyl)benzene as P2′ ligands.



Reagents and conditions: (a) Na_2CO_3, EtOAc, H_2O, RT, 18 h, 75–94%; (b) AD-mix-*β*, *t*-BuOH, H_2O, RT, 4 h, 68–87%; (c) TFA, CH_2Cl_2, RT, 1 h; (d) DIEA, CH_3CN, RT, 24 h, 33–79%; (e) AD-mix-*α*, *t*-BuOH, H_2O, RT, 4 h, 47–86%.

4.4 Results and Discussion

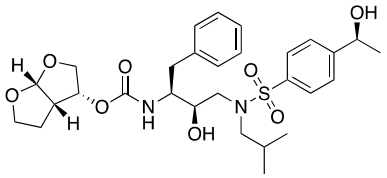
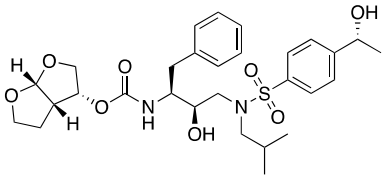
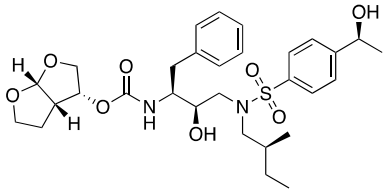
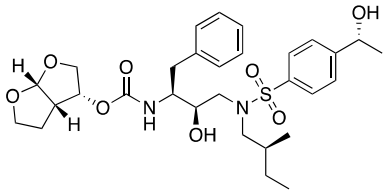
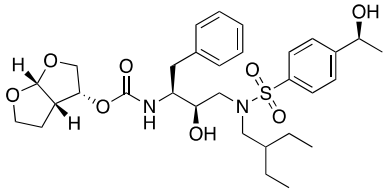
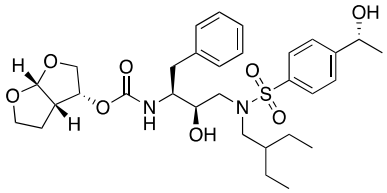
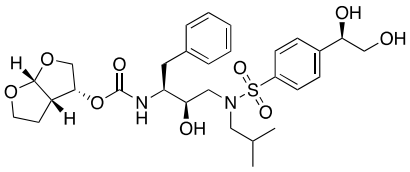
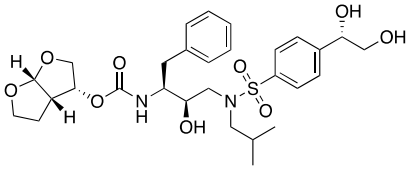
Our goal was to improve potency of PIs against MDR HIV-1 variants by optimizing hydrogen bonding interactions in the S2′ subsite of HIV-1 protease. Specifically, we explored the possibility of additional hydrogen bonding interactions between the hydroxyl group of the P2′ moiety and the backbone amide NH of Asp29′ and Asp30′ on the basis of an analysis of **3**- and **4**-bound HIV-1 protease structures and molecular modeling. Addition of a methyl group to the hydroxymethyl substituent was expected to shift the hydroxyl group closer to the backbone NH of Asp29′, while increasing vdW interactions with residues in the S2′ subsite. Similarly, replacement of the hydroxymethyl group with a more polar 1,2-dihydroxyethyl was intended to enhance polar interactions; particularly the second hydroxyl group was expected to replace the water-mediated interactions between the P2′ moiety and the Asp30′ side chain. However, due to the free rotation of the hydroxyethyl and dihydroxyethyl groups, modeling did not provide clear distinction which of the two stereoisomers would provide optimal polar and vdW interactions. Thus, both (*R*)- and (*S*)-stereoisomers of 4-(1-hydroxyethyl)benzene and 4-(1,2-dihydroxyethyl)benzene moieties were explored as P2′ ligands in combination with three P1′ groups of varying size and hydrophobicity.

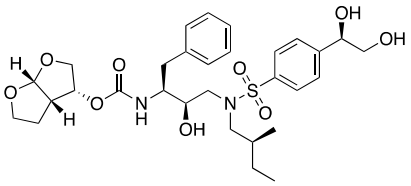
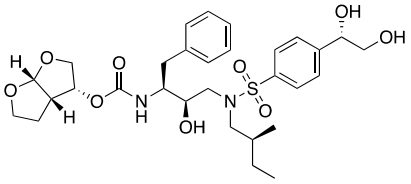
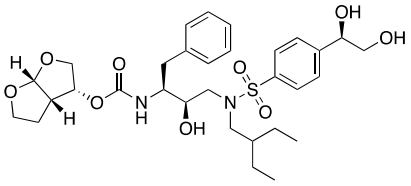
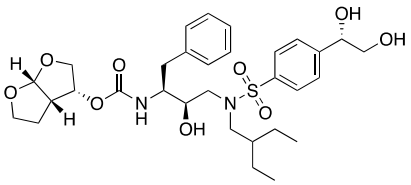
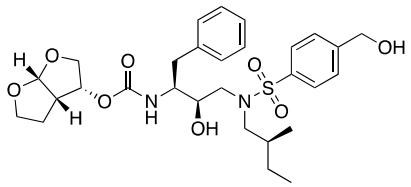
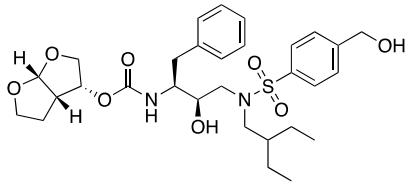
4.4.1 Enzyme Inhibition Assays

The binding affinity of HIV-1 PIs has reached a level where assessing the inhibition of wild-type protease using standard FRET-based assays is quite difficult. Accurate measurement of inhibition constants in the low pM range remains a challenge even with the recently reported highly sensitive fluorogenic assay.[177] Due to these limitations, two drug-resistant protease variants, I84V and I50V/A71V were selected to assess the

potency of new PIs. I84V is a common drug resistance mutation that reduces susceptibility to all FDA-approved PIs, while I50V/A71V mutations arise in response to therapy with APV and DRV-based regimens, resulting in reduced susceptibility to these PIs. The enzyme inhibition constants (*K*i) were determined against the I84V and I50V/A71V protease variants using the highly sensitive FRET assay[177]; DRV was used as a control in all assays [**Table 4.1**].

Table 4.1 Protease inhibitory activity of PIs 12–17 and 22–27 against drug-resistant variants.

Inhibitor	Structure	K_i (nM)	
		I84V	I50V/A71V
12		0.048 ± 0.004	0.057 ± 0.008
13		0.109 ± 0.007	0.092 ± 0.008
14		0.111 ± 0.008	0.093 ± 0.013
15		0.095 ± 0.005	0.132 ± 0.014
16		0.020 ± 0.002	0.102 ± 0.010
17		0.028 ± 0.003	0.107 ± 0.014
22		0.098 ± 0.008	0.090 ± 0.008
23		0.283 ± 0.024	NT

24		0.216 ± 0.011	NT
25		0.504 ± 0.039	NT
26		0.057 ± 0.004	0.074 ± 0.006
27		0.134 ± 0.007	0.220 ± 0.011
3		0.010 ± 0.003	0.080 ± 0.006
4		0.005 ± 0.002	0.055 ± 0.004
DRV		0.025 ± 0.006	0.075 ± 0.006

NT: not tested

DRV retained excellent potency against the I84V and I50V/A71V protease variants (*K*i = 25 pM and 75 pM, respectively) compared to that reported for wild-type protease (*K*i = 5–16 pM).[159, 169, 177] Compound **3**, which incorporates a 4-(hydroxymethyl)benzene moiety as P2′ ligand and (*S*)-2-methylbutyl group at the P1′ position, showed better potency against the I84V protease but was equipotent to DRV against the I50V/A71V variant. Analogue **4** with a larger 2-ethylbutyl group at the P1′ position was more active than DRV against both the I84V and I50V/A71V protease variants. The corresponding analogue **2** with the isobutyl P1′ group was previously reported which was equipotent to DRV against wild-type protease but exhibited relatively lower antiviral potency, presumably due to reduced hydrophobicity.[159, 178] Thus, as we have previously shown, PIs with 4-(hydroxymethyl)benzene as the P2′ moiety in combination with larger, more hydrophobic (*S*)-2-methylbutyl and 2-ethylbutyl P1′ groups maintained potency against drug-resistant protease variants.

Replacement of the 4-(hydroxymethyl)benzene P2′ moiety of PIs **2**–**4** with (*S*)- and (*R*)-4-(1-hydroxyethyl)benzene resulted in pairs of diastereoisomers epimeric at P2′. The P2′ hydroxyethyl analogues **12**–**17** showed excellent inhibitory potencies against the I84V and I50V/A71V protease variants with *K*i values ranging between 20 to 132 pM. In each case, both diastereoisomers were relatively less potent than the corresponding parent hydroxymethyl compound, particularly against the I84V protease variant. However, no major difference in inhibitory potency was observed between P2′ epimeric compounds. Only compounds with the isobutyl P1′ group showed minor differences in inhibitory potencies between the P2′ epimers, as diastereoisomer **12** with (*S*)-configuration of the hydroxyethyl group was slightly more potent than the corresponding (*R*)-epimer **13**. While

this work was underway, Ghosh *et al*. reported the synthesis of PIs **12** and **13** using a much longer reaction sequence; both compounds were tested against wild-type HIV-1 protease, and only compound **13** was evaluated for antiviral activity.[81]

Diastereoisomers **14** and **15** with the (*S*)-2-methylbutyl P1′ group showed similar potencies against the I84V and I50V/A71V protease variants, and both were less potent than the parent **3**. Similarly, no difference in inhibitory potency was observed between diastereoisomers **16** and **17** incorporating the 2-ethylbutyl P1′ group. However, analogues **16** and **17** were more potent against the I84V protease (*K*i = 20 and 28 pM, respectively) and exhibited overall potency profiles comparable to that of DRV. Together, these data demonstrate that minor modifications to the P1′ and P2′ moieties result in distinct potency profiles.

In contrast to the 4-(1-hydroxyethyl)benzene P2′ series, compounds with the 4-(1,2-dihydroxyethyl)benzene P2′ moiety showed clear differences in inhibitory potency between the P2′-epimers. Compounds incorporating the (*R*)-4-(1,2-dihydroxyethyl)benzene P2′ moiety were 2–3-fold more potent than the corresponding (*S*)-epimers. Diastereoisomer **22**, with the isobutyl group at the P1′ position and (*R*)-configuration of the dihydroxyethyl group, exhibited excellent inhibitory potency against the I84V and I50V/A71V proteases, while the corresponding (*R*)-epimer **23** was about 3-fold less active against the I84V protease. A similar trend was observed for diastereoisomers **24** and **25** incorporating the (*S*)-2-methylbutyl P1′ group, though both compounds were 2-fold less active against the I84V protease than the corresponding isobutyl P1′ analogues **23** and **24**. The analogue **26** incorporating the larger 2-ethylbutyl P1′ group and (*R*)-configuration of the dihydroxyethyl group was the most potent in this

series, with enzyme potency against the I50V/A71V protease variant comparable to that of DRV. Again, the corresponding P2′ (*S*)-epimer **27** was 2–3-fold less potent against the protease variants tested. Thus, shape of the hydrophobic group at P1′ and the configuration of the dihydroxyethyl group at P2′ affected inhibitor potency against drug-resistant protease variants.

4.4.2 Antiviral Assays

The potency and resistance profiles of PIs were evaluated using a cell-based antiviral assay against wild-type HIV-1 and three representative MDR variants selected from a panel of clinically relevant multi-PI resistant recombinant clones [**Table 4.3-4.4**].[179] Relative to the wild-type HIV-1 strain NL4-3 (with a near consensus protease sequence), the MDR variants contain 19, 20 and 24 amino acid substitutions in the protease and are named SLK19, VSL20, and KY24, respectively. All three variants show high-level resistance and cross-resistance to multiple PIs including DRV, as determined by the Monogram Biosciences (South San Francisco, CA) PhenoSense assay.[179] Accordingly, in antiviral assays, the MDR variants exhibited increasingly high-level resistance to DRV: compared to wild-type HIV-1, DRV was 6-fold less potent against SLK19, 58-fold less potent against VSL20, and 210-fold less potent against KY24. Thus, the three selected MDR HIV-1 strains represent clinically relevant MDR viruses with a wide spectrum of DRV resistance.

Table 4.2 Antiviral potency of PIs 12–17 and 22–27 against WT HIV-1 and drug-resistant variants

Inhibitor	cLogP[a]	Antiviral EC_{50} (nM) (fold change)[b]			
		WT	SLK19	VSL20	KY24
12	2.885	4.7	31 (7)	117 (25)	352 (75)
13	2.885	7.5	100 (13)	148 (20)	360 (48)
14	3.414	4.6	49 (11)	281 (61)	490 (107)
15	3.414	4.2	91 (22)	291 (69)	342 (81)
16	3.943	4.4	26 (6)	112 (25)	196 (45)
17	3.943	5.1	32 (6)	420 (82)	345 (68)
22	1.673	176	NT	NT	NT
23	1.673	189	NT	NT	NT
24	2.202	80	NT	NT	NT
25	2.202	92	NT	NT	NT
26	2.731	40	116 (3)	100 (3)	507 (13)
27	2.731	41	151 (4)	417 (10)	1154 (28)
3	3.105	5.7	52 (9)	310 (54)	1087 (191)
4	3.634	4.2	36 (9)	426 (101)	495 (118)
DRV	2.387	5.5	33 (6)	320 (58)	1157 (210)

[a]cLogP values were calculated using ChemDraw 18. [b]The multidrug-resistant HIV-1variants SLK19, VSL20 and KY24 contain the following amino acid substitutions in HIV-1 protease compared to the WT NL4-3 strain: SLK19 (L10I, V11I, I13V, K14R, A22V, E35D, M36I, N37D, G48M, F53L, I54V, I62V, L63P, A71V, T74S, V82A, I84V, T91S, Q92K); VSL20 (L10I, K20R, L33F, E34A, E35D, M36I, N37D, K43T, M46I, G48V, I50V, I54T, I62V, L63P, A71V, I72V, V77I, V82A, I85V, I93L); KY24 (L10V, T12V, I13V, I15V, K20M, V32I, L33F, K43T, M46I, I47V, I54M, D60E, Q61N, I62V, L63P, C67Y, H69K, A71I, I72L, G73S, V77I, V82A, L89V, L90M). NT: not tested.

Table 4.3 Antiviral potency of PIs 12–17 and 22–27 against WT HIV-1 and drug-resistant variants

Inhibitor	cLogP	Antiviral EC_{50} (nM) (95% CI: lower, upper)			
		WT	SLK19	VSL20	KY24
12	2.885	4.7 (4.2, 5.2)	31 (23, 43)	117 (84, 162)	352 (311, 398)
13	2.885	7.5 (6.4, 8.7)	100	148 (110, 200)	360 (278, 466)
14	3.414	4.6 (4.0, 5.3)	49 (39, 62)	281 (215, 366)	490 (402, 599)
15	3.414	4.2 (3.7, 4.8)	91 (48, 180)	291 (227, 372)	342 (246, 477)
16	3.943	4.4 (3.8, 5.1)	26 (13, 50)	112 (78, 160)	196 (162, 238)
17	3.943	5.1 (4.4, 6.3)	32 (26, 39)	420 (311, 566)	345 (294, 405)
22	1.673	176 (154, 201)	NT	NT	NT
23	1.673	189 (181, 196)	NT	NT	NT
24	2.202	80 (69, 92)	NT	NT	NT
25	2.202	92 (88, 96)	NT	NT	NT
26	2.731	40 (7.2, 72)	116 (92, 145)	100 (74, 134)	507 (425, 605)
27	2.731	41 (25, 57)	151 (100, 229)	417 (268, 649)	1154 (1017, 1311)
3	3.105	5.7 (1.6, 9.9)	52 (37, 66)	310 (252, 382)	1087 (866, 1364)
4	3.634	4.2 (3.4, 5.0)	36 (17, 54)	426 (330, 550)	495 (399, 615)
DRV	2.387	5.5 (4.6, 6.5)	33 (30, 35)	320 (302, 340)	1157 (994, 1347)

NT: not tested

The 4-(hydroxymethyl)benzene P2′ analogues **3** and **4** were equipotent to DRV against wild-type HIV-1 with EC_{50} values of 5.7 and 4.2 nM, respectively. The two analogues retained potent activity against the SLK19 variant but were significantly less potent against the VSL20 variant, with EC_{50} values similar to that of DRV. Compared to DRV and compound **3**, both of which showed similar potencies against the three MDR variants, analogue **4** retained 2-fold better potency against the KY24 variant. The slightly improved resistance profile of compound **4** is in agreement with previous results and indicates that optimizing vdW interactions in the S1′ pocket can improve potency against MDR viruses.[169]

All compounds incorporating the 4-(1-hydroxyethyl)benzene P2′ moieties (**12**–**17**) potently inhibited wild-type HIV-1 with EC_{50} values similar to that of DRV and parent compounds **3** and **4**. However, the three MDR HIV-1 variants showed distinct susceptibilities to each of these compounds, depending on the P1′ group and the configuration of the hydroxyethyl group at the P2′ moiety. Compound **12**, with the isobutyl P1′ group and the (*S*)-4-(1-hydroxyethyl)benzene P2′ moiety, was equipotent to DRV against SLK19 but exhibited 2–3-fold better potency against the highly resistant MDR variants VSL20 and KY24, resulting in an overall improved resistance profile. The corresponding epimer **13**, incorporating the (*R*)-4-(1-hydroxyethyl)benzene P2′ moiety, while showing a 3-fold lower potency against SLK19, retained similar potency as **12** against VSL20 and KY24. Diastereoisomers **14** and **15** with the (*S*)-2-methylbutyl P1′ group exhibited similar potency profiles as the corresponding analogues **12** and **13**, except for a 2-fold loss in potency against VSL20. Further improvement in overall potency profile was realized with analogue **16** incorporating a larger 2-ethylbutyl P1′ group and

the (*S*)-4-(1-hydroxyethyl)benzene P2′ moiety, which maintained similar potency against SLK19 as DRV but demonstrated about 3- and 6-fold increase in potency against VSL20 and KY24, respectively. Moreover, compared to the parent compound **4**, analogue **16** demonstrated 4- and 2-fold improved potency against VSL20 and KY24. Compared to **16**, the (*R*)-epimer **17** maintained similar potency against SLK19 but exhibited lower potency against VSL20 and KY24. While all compounds with the 4-(1-hydroxyethyl)benzene as P2′ ligands maintained excellent potency against MDR HIV-1 variants, each exhibited distinct resistance profile depending on the configuration of the hydroxyethyl group and size of the P1′ group.

The introduction of a more polar 4-(1,2-dihydroxyethyl)benzene moiety as the P2′ ligand resulted in significantly reduced potency against wild-type HIV-1, despite compounds **22**–**27** showing pM inhibitory activities in biochemical assays. The observed loss of potency in cellular assays is likely due to reduced hydrophobicity, as indicated by lower calculated partition coefficient (LogP) values compared to DRV [**Table 4.1**]. The cellular potency of HIV-1 protease inhibitors has been shown to strongly correlate with the hydrophobicity descriptor LogP, suggesting that membrane transport is a key factor affecting antiviral potency.[180]

The observed antiviral potencies of the 4-(1,2-dihydroxyethyl)benzene containing compounds indeed correlate with their cLogP values [**Figure 4.3**]. In all cases both diastereoisomers exhibited similar antiviral potencies against wild-type HIV-1. Compounds **22** and **23** incorporating the isobutyl P1′ group showed significantly lower potency than DRV. Replacement of the isobutyl P1′ group with a slightly more hydrophobic (*S*)-2-methylbutyl group, providing analogues **24** and **25**, resulted in further

improvement in antiviral potency. Compounds **26** and **27** incorporating a larger 2-ethylbutyl P1′ group further increased hydrophobicity and resulted in a 2-fold improvement in potency compared to the (*S*)-2-methylbuty P1′ analogues **24** and **25**. However, both diastereoisomers **26** and **27** were 8-fold less active than DRV against wild-type HIV-1 despite **26** exhibiting similar potency in enzyme inhibition assays. Compared to DRV the P2′ (*R*)-epimer **26** maintained better potency against MDR variants VSL20 and KY24 and exhibited much lower fold potency losses. The corresponding (*R*)-epimer **27**, though less active against SLK19, showed similar potency as DRV against VSL20 and KY24. Thus, the combination of a more hydrophobic 2-ethylbutyl P1′ group and the (*R*)-4-(1,2-dihydroxyethyl)benzene P2′ moiety provided compounds with improved potency profiles.

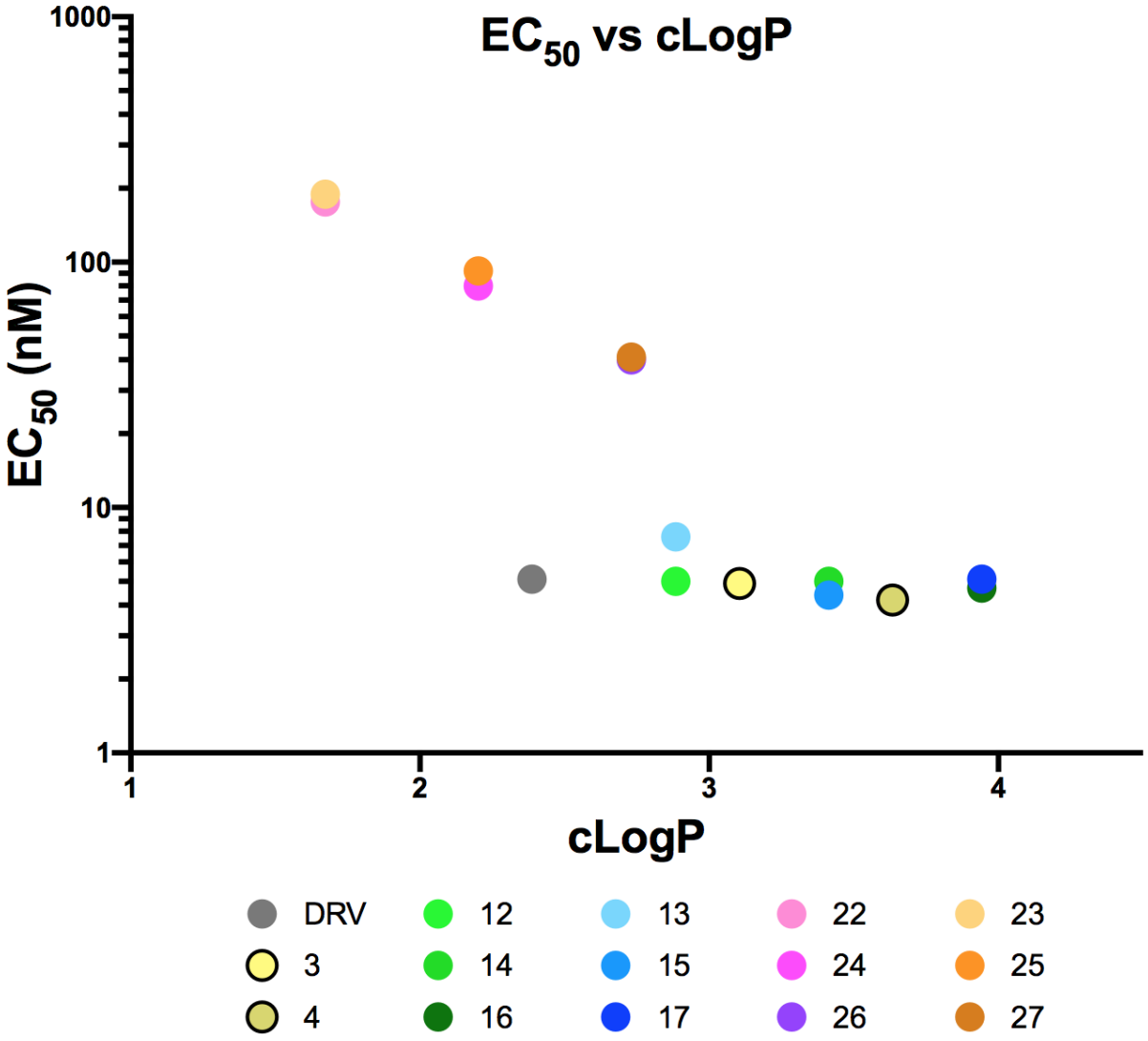


Figure 4.3 Correlation of observed antiviral potencies of PIs against wild-type HIV-1 with hydrophobicity descriptor cLogP.

4.4.3 Analysis of Protease-Inhibitor Complexes

To explore molecular interactions of the chiral 4-(1-hydroxyethyl)benzene and 4-(1,2-dihydroxyethyl)benzene P2′ moieties, we determined crystal structures of PIs **12**–**17** and **22**–**27** bound to wild-type HIV-1 protease of the NL4-3 strain. The cocrystal structures of PIs **2**–**4** were also determined with the same protease enzyme, as the previously reported cocrystal structures of PIs **3** and **4** were with the wild-type protease of SF-2 sequence, which is slightly different from NL4-3. The two variants differ by four amino acids which caused minor structural differences at the distal loops, but the active sites were nearly identical, including inhibitor binding and crystallographic waters. The crystallographic data collection and refinement statistics are summarized in **Table 4.4**. All 15 high-resolution (1.86–2.03 Å) cocrystal structures were solved in the same $P2_12_12_1$ space group with one protease homodimer in the asymmetric unit, and only one orientation of the bound inhibitor in the protease active site, which was crucial for direct comparison and analysis.

Table 4.4 X-ray data collection and crystallographic refinement statistics.

Inhibitor	***1***	***2***	***3***	***4***	***12***	***13***	***14***	***15***	***16***	***17***	***22***	***23***	***24***	***25***	***26***	***27***
PDB Code	6DGX	6OXO	6OXP	6OXQ	6OXR	6OXS	6OXT	6OXU	6OXV	6OXW	6OXX	6OXY	6OXZ	6OY0	6OY1	6OY2
Resolution (Å)	2.00	2.00	1.97	1.89	2.03	1.99	1.86	1.86	1.99	1.98	1.96	1.96	1.96	2.00	2.00	1.99
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Cell dimensions:																
a (Å)	51.0	51.1	51.0	50.9	50.8	50.9	51.1	51.0	50.9	50.9	51.0	51.0	50.9	51.1	50.9	51.1
b (Å)	58.3	58.4	58.0	58.1	58.6	59.0	58.1	58.0	58.8	58.9	57.9	58.2	57.7	58.1	57.8	58.2
c (Å)	61.9	62.2	62.1	62.3	61.8	61.6	62.2	62.0	61.7	61.7	62.3	62.2	62.4	62.3	62.2	62.1
β (°)	90	90	90	90	90	90	90	90	90	90	90	90	90	90	90	90
Completeness (%)	98.1	99.7	94.5	96.8	95.8	97.8	95.4	95.1	99.6	95.3	96.8	97.1	95.4	98.1	92.7	95.0
Total reflections	84332	86563	55435	104830	88407	84420	90485	84249	47857	86496	79161	62349	62708	82784	58994	61152
Unique reflections	12758	13010	12890	14875	11816	12983	15390	15237	13165	12786	13311	13435	13109	12855	12065	12586
Average I/σ	31.1	15.7	24.9	33.6	26.2	24.9	23.5	26.7	22.2	25.6	17.6	25.3	27.0	25.8	32.3	24.5
Redundancy	6.6	6.7	4.3	7.0	7.5	6.5	5.9	5.5	3.6	6.8	5.9	4.6	4.8	6.4	4.9	4.9
R_{sym} *(%)*[a]	5.7	10.4	5.3	5.5	7.5	7.1	8.3	6.9	5.6	6.8	8.8	5.9	5.9	7.2	5.1	6.4
RMSD[b] *in:*																
Bond lengths (Å)	0.003	0.002	0.004	0.009	0.002	0.003	0.003	0.008	0.007	0.003	0.005	0.005	0.003	0.003	0.008	0.002
Bond angles (°)	0.590	0.547	0.689	0.971	0.553	0.657	0.703	0.904	0.956	0.608	0.784	0.767	0.640	0.766	0.913	0.584
R_{factor} *(%)*[c]	19.2	20.1	17.3	19.7	18.9	18.7	19.8	18.5	17.7	18.8	19.2	16.5	17.1	17.9	17.8	19.3
R_{free} *(%)*[d]	22.8	25.0	22.2	23.7	23.4	22.8	23.1	23.3	21.9	23.8	23.3	19.6	21.9	21.8	22.5	24.0

[a] $R_{sym} = \Sigma \mid I - \langle I \rangle \mid / \Sigma I$, where I = observed intensity, $\langle I \rangle$ = average intensity over symmetry equivalent; values in parentheses are for the highest resolution shell.
[b] RMSD, root mean square deviation.
[c] $R_{factor} = \Sigma \mid\mid F_o\mid - \mid F_c\mid\mid / \Sigma \mid F_o\mid$.
[d] R_{free} was calculated from 5% of reflections, chosen randomly, which were omitted from the refinement process.

The overall binding conformations of all PIs incorporating the 4-(hydroxymethyl)benzene (**2**–**4**), 4-(1-hydroxyethyl)benzene (**12**–**17**), and 4-(1,2-dihydroxyethyl)benzene (**22**–**27**) P2′ moieties are similar to that of DRV [**Figure 4.4A**]. However, clear differences in contacts were observed in the binding of the P2′ moieties depending on the substituent at the 4-position of the benzene ring and the configuration of the hydroxyl group [**Figure 4.4B, 4.4D**]. Compared to DRV, all PIs with modified P2′ ligands maintain similar hydrogen bonding interactions with the protease, except in the S2′ subsite [**Figures 4.5 and 4.6**]. The cocrystal structures of PIs **2**–**4** in complex with protease showed minor changes in the conformation of the 4-(hydroxymethyl)benzene moiety compared to the 4-aminobenzene of DRV in the S2′ subsite [**Figure 4.7**]. In the DRV-bound protease structure, the amino group of the 4-aminobenzene P2′ moiety is involved in hydrogen bonding interactions with the main chain carbonyl of Asp30′ and water-mediated interactions with the side-chain carboxylate of Asp30′ [**Figure 4.2**]. The primary hydroxyl group of the 4-(hydroxymethyl)benzene P2′ moiety in compounds **2**–**4** is oriented toward Asp29′ and Asp30′ backbone and makes a direct hydrogen bond with the backbone NH of Asp30′. Moreover, the hydroxyl group interacts with the backbone NH of Asp29′ and the side chain of Asp30′ through water-mediated hydrogen bonds. This network of hydrogen bonding interactions in the S2′ subsite likely underlie the improved potency of compounds **2**–**4** compared to DRV [**Figure 4.2**].

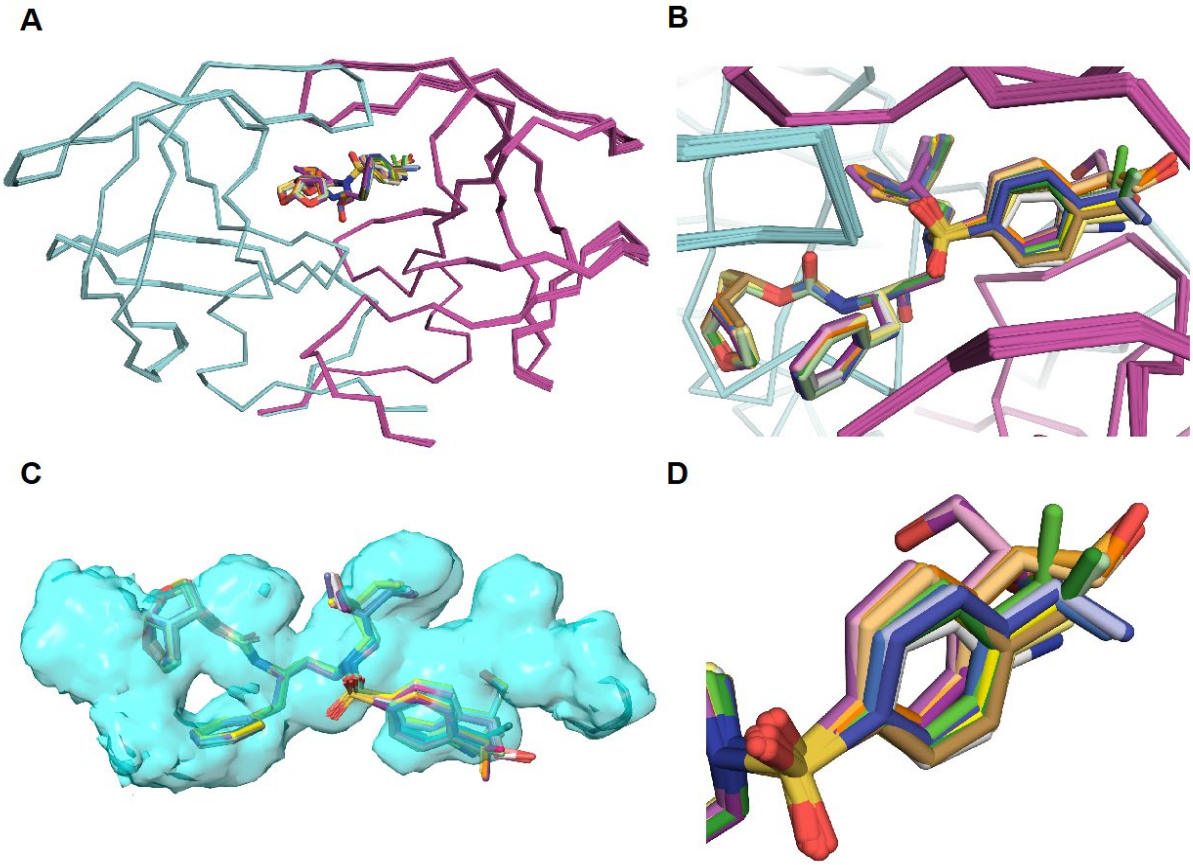


Figure 4.4 Comparison of binding modes of protease inhibitors with 4-(1-hydroxyethyl)benzene and 4-(1,2-dihydroxyethyl)benzene P2′ moieties in the active site of wild-type HIV-1 protease. The two protease monomers are in cyan (denoted as non-prime) and magenta (denoted as prime). Superposition of protease complexes with DRV (1), parent compounds (2–4), and new analogues (12–17 and 22–27). The inhibitors are shown as sticks and HIV protease dimers are shown as ribbons. (B) Zoomed-in active site of superimposed complexes. The inhibitors bind to wild-type HIV protease in similar conformations except minor variations in the S2′ subsite. (C) Fit of inhibitors within the substrate envelope. The substrate envelope is in blue space filling representation, and the superimposed inhibitors are displayed as sticks. There is minimal protrusion of inhibitors outside the active site. (D) Variations in the binding mode of the inhibitors' P2′ moieties.

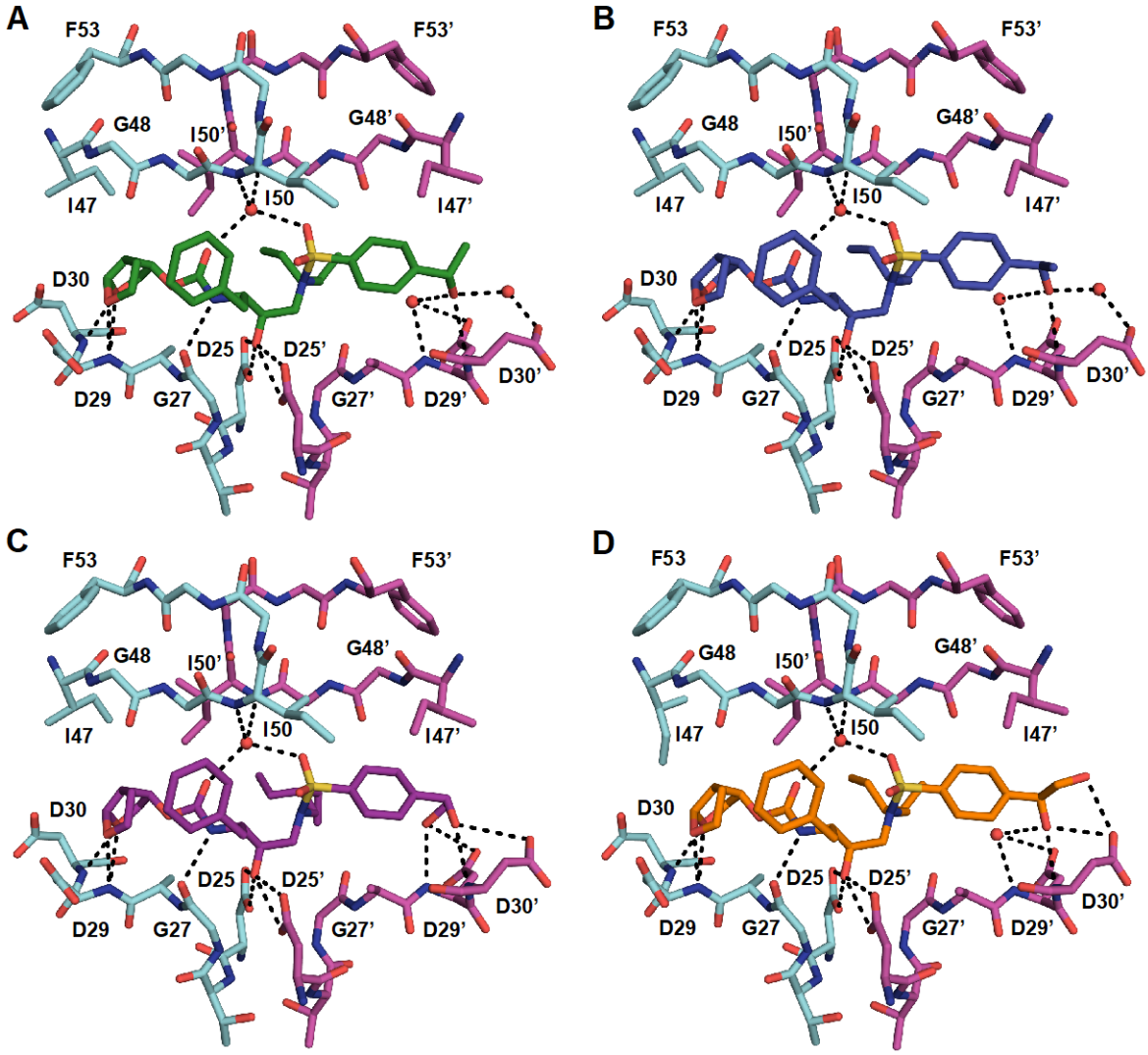


Figure 4.5 Crystal structures of wild-type HIV-1 protease in complex with inhibitors (A) 16, (B) 17, (C) 26, and (D) 27. Both (*R*)- and (*S*)-stereoisomers of the P2′ 4-(1-hydroxyethyl)benzene moiety make direct hydrogen bonding interactions with the backbone NH of Asp30′ in the S2′ subsite (A and B). The (*R*)-stereoisomer of the P2′ 4-(1,2-dihydroxyethyl)benzene moiety makes hydrogen bonding interactions with backbone NH of Asp29′ and Asp30′ (C). The (*S*)-stereoisomer of the P2′ 4-(1,2-dihydroxyethyl)benzene moiety makes hydrogen bonding interactions with the backbone NH and side chain carboxylate group of Asp30′ (D).

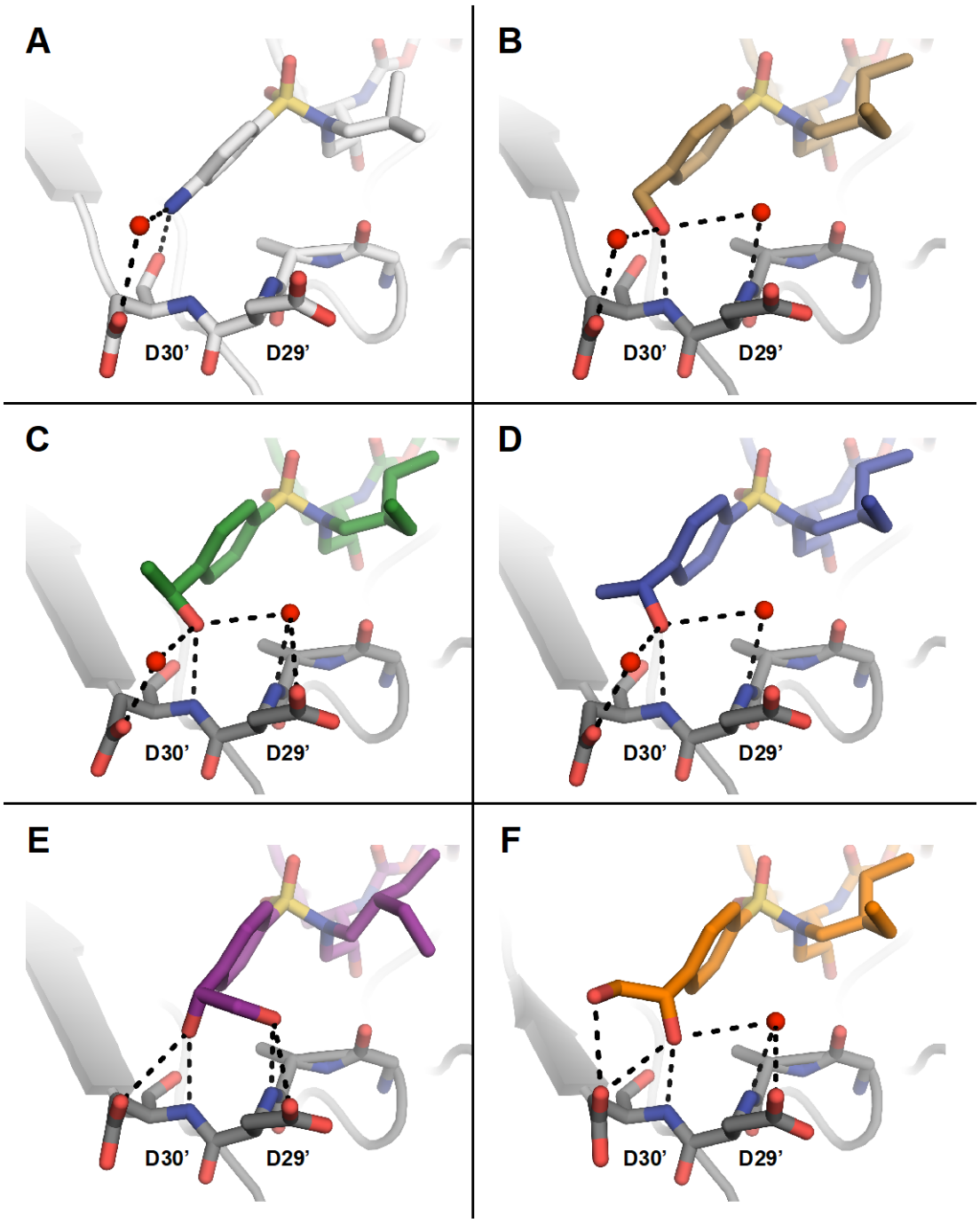


Figure 4.6 Comparison of binding interactions of representative PIs with modified P2′ moieties in the S2′ subsite of HIV-1 protease. Binding interactions of (A) DRV (PDB 6DGX), (B) parent compound 4 (PDB 6OXQ) with the 4-(hydroxymethyl)benzene P2′ moiety, (C) inhibitor 16 with (*S*)-4-(1-hydroxyethyl)benzene P2′ moiety, (D) inhibitor 17 with the (*R*)-configuration of the P2′ moiety, (E) inhibitor 26 with (*R*)-4-(1,2-dihydroxyethyl)benzene P2′ moiety, and (F) inhibitor 27 with the (*S*)-configuration of the P2′ moiety.

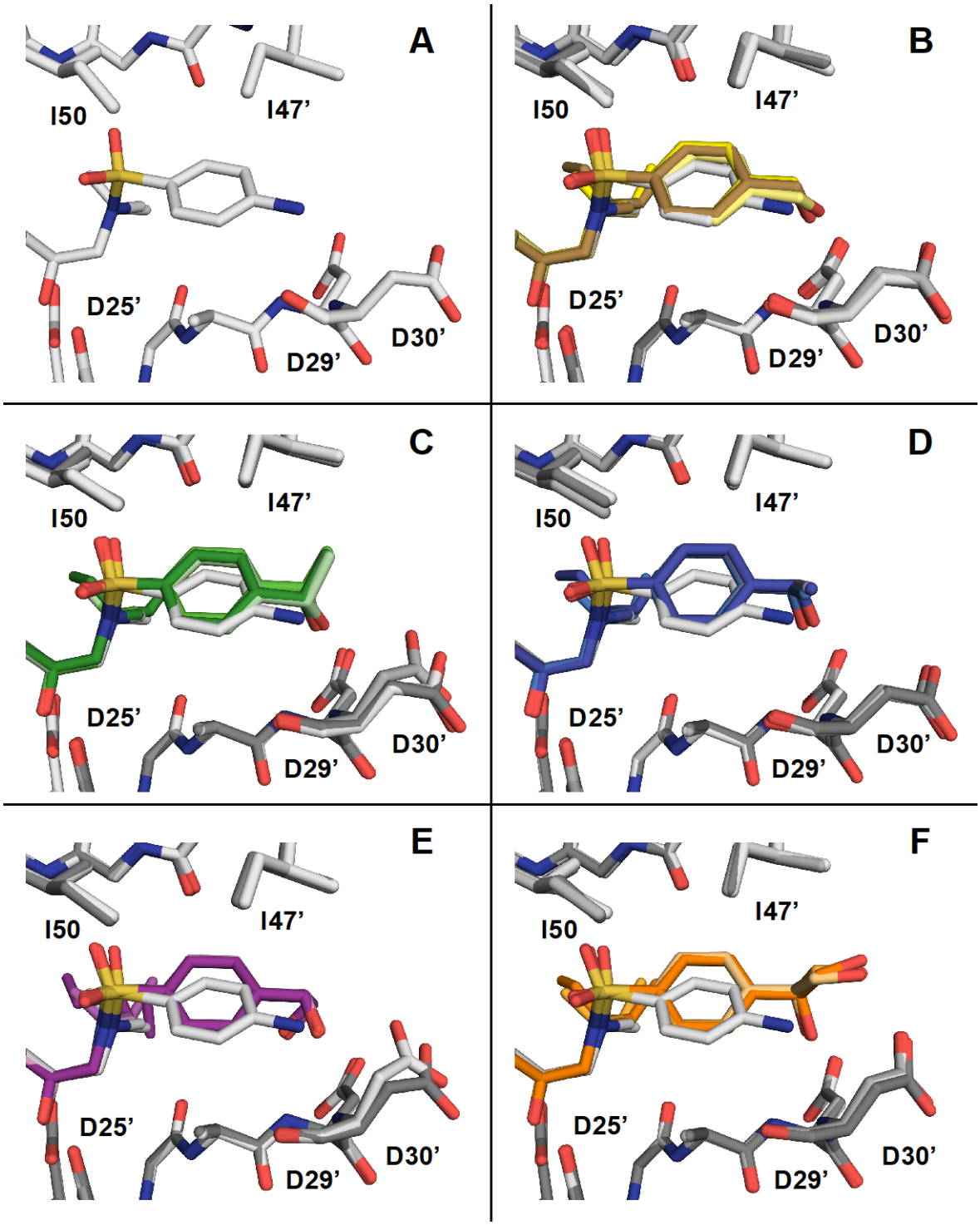


Figure 4.7 X-ray crystal structures of inhibitors superimposed with DRV showing shift of the benzene ring of the P2′ moiety. Superimposed structures of (B) parent compounds 2–4; (C) analogues 12, 14 and 16; (D) 13, 15 and 17; (E) 22, 24 and 26; and (F) 23, 25 and 27.

The PIs **12**–**17** with the 4-(1-hydroxyethyl)benzene moiety at the P2′ position were designed to better position the hydroxyl group between the backbone NH of Asp29′ and Asp30′. In addition, the methyl group was expected to make vdW interactions with the hydrophobic residues around the S2′ subsite. Similar to the parent compounds, PIs **12**–**17** fit well within the substrate envelope [**Figure 4.4C**]. For both the (*S*)- and (*R*)-4-(1-hydroxyethyl)benzene P2′ moieties only a portion of the hydroxyethyl substituent protrudes outside the envelope. The structures of PIs **12**, **14** and **16** bound to protease superimpose very well with the corresponding parent compound structures with only subtle changes in the position and orientation of the (*S*)-4-(1-hydroxyethyl)benzene P2′ moiety. The differences in the conformation of the P2′ moiety are more evident for PIs **13**, **15** and **17** incorporating the (*R*)-4-(1-hydroxyethyl)benzene, which is shifted upward toward the Gly48 backbone with a slight change in the orientation of the benzene ring compared to the position in the corresponding epimers, likely to maintain polar interactions between the hydroxyl group and the backbone NH of Asp30′ [**Figure 4.7**].

Despite clear differences in the overall binding conformations of (*S*)- and (*R*)-4-(1-hydroxyethyl)benzene P2′ moieties, the secondary hydroxyl group is oriented in the same direction of Asp29′ and Asp30′ backbone [**Figure 4.6**]. Instead, the orientation of the methyl group is altered. As a result, regardless of the configuration of the hydroxyl group, the PIs incorporating the 4-(1-hydroxyethyl)benzene moiety (**12**–**17**) make identical direct and water-mediated hydrogen bonding interactions in the protease active site as the parent compounds **2**–**4** [**Figure 4.6**]. In both configurations, the hydroxyl group is positioned between the backbone NH of Asp29′ and Asp30′. However, in all structures the hydroxyl group makes a direct hydrogen bond only with the backbone NH of Asp30′,

which is closer (3.0–3.2 Å) than the NH of Asp29′ (3.6–3.8 Å). The distance between the hydroxyl group and the backbone NH of Asp30′ is slightly shorter (3.0–3.1 Å) for compounds with the (*S*)-4-(1-hydroxyethyl)benzene P2′ moiety than the corresponding (*R*)-epimers (3.2 Å) [**Table 4.5**]. The hydroxyl group also interacts with the side chain of Asp30′, mostly through water-mediated hydrogen bonds but directly in case of compound **14**. In the **14**-protease complex structure the side chain of Asp30′ is shifted toward the P2′ moiety, resulting in a strong, direct hydrogen bond between the hydroxyl and carboxylate groups. Moreover, the hydroxyl group makes water-mediated interactions with the backbone NH of Asp29′, which are not observed in DRV-protease complex. Thus, the hydroxyl group at the P2′ moiety is involved in a network of direct and water-mediated interactions in the protease active site [**Figure 4.6**].

Table 4.5 Intermolecular hydrogen bonds between the protease and inhibitors in the cocrystal structures (distance in Å). Shorter distances are more red and longer distances are more blue (cutoff 3.5 Å).

Residue	D30	D29	D29	G27	D25	D25	D25'	D25'	I50	I50	I50'	I50'	D30'	D30'	D30'	D30'	D30'	D29'	D29'	D29'	D29'	D29'
Atom	MC-N	MC-N	MC-N	MC-O	SC-O	SC-O	SC-O	SC-O	MC-N	MC-N	MC-N	MC-N	SC-O	SC-O	SC-O	SC-O	MC-N	MC-N	SC-O	H2O	MC-N	SC-O
Type	Direct	Direct	Direct	Direct	Direct	Direct	Direct	Direct	Water	Water	Water	Water	Water	Water	Direct	Direct	Direct	Direct	Direct	Water	Water	Water
DRV	2.9	3.2	2.8	3.0	3.2	2.5	2.9	2.5	3.0	3.1	2.8	2.9	2.9	2.9								
2	3.0	3.2	2.7	3.1	3.1	2.5	2.8	2.4	2.9	3.1	2.5	3.1					3.1					
3	3.0	3.2	2.9	3.0	3.2	2.6	2.9	2.5	2.8	3.0	2.7	3.1	2.5	2.1			3.0			3.0	2.9	
4	3.0	3.1	2.8	3.0	3.3	2.6	2.9	2.4	2.7	3.0	2.8	3.0	2.7	2.8			2.9			3.2	2.9	
12	2.9	3.1	2.8	2.9	3.1	2.4	2.8	2.4	3.0	3.0	2.6	3.0	2.6	2.4			3.1			3.3	3.1	
14	2.9	3.1	2.8	3.0	3.3	2.6	2.9	2.5	2.8	2.9	2.6	3.0			3.0		3.0			3.1	2.8	
16	3.0	3.2	2.9	3.0	3.3	2.5	2.9	2.4	2.9	2.9	2.7	2.9	2.6	2.3			3.1			3.0	3.0	3.5
13	2.9	3.1	2.8	3.0	3.2	2.4	2.9	2.5	2.9	3.0	2.6	3.0	2.4	2.6			3.2			3.3	3.0	
15	3.1	3.1	2.9	3.1	3.3	2.6	2.9	2.5	2.7	2.9	2.7	3.0	2.7	2.4			3.2			2.9	2.8	
17	2.9	3.2	2.8	3.0	3.2	2.5	2.8	2.3	2.9	2.9	2.6	3.0	2.5	2.5			3.2			3.3	3.1	
22	2.9	3.0	2.8	3.0	3.4	2.6	2.9	2.3	2.8	3.0	2.6	3.0			2.5		2.9	3.1	3.4			
24	2.9	3.1	2.8	3.0	3.4	2.6	2.8	2.4	2.9	3.0	2.6	3.0			2.4		2.9	3.1	3.3			
26	3.0	3.1	2.8	3.1	3.3	2.6	2.9	2.3	2.7	3.1	2.7	2.9			3.4		2.9	3.1	3.5			
23	3.0	3.2	2.9	3.0	3.3	2.6	2.9	2.4	2.8	3.1	2.6	3.1			3.2	2.5	3.2			3.3	3.1	
25	2.9	3.0	2.8	3.0	3.2	2.5	2.8	2.4	2.8	3.1	2.5	3.0			2.9	2.4	3.0					
27	3.0	3.1	2.8	2.9	3.2	2.5	2.8	2.4	3.0	3.0	2.5	2.9			3.1	3.0	3.0			3.3	3.0	3.4

MC: Main chain; SC: side chain

Cocrystal structures were analyzed for vdW contacts, and compounds with the modified P2′ moieties showed enhanced vdW interactions with protease residues in the S2′ subsite compared to the corresponding parent PIs **2**–**4** [**Figure 4.8**]. To evaluate the changes in vdW interactions, total vdW energies per residue were calculated for each protease-inhibitor complex and the corresponding values for parent compound subtracted [**Figure 4.9**]. The additional methyl group is oriented toward Ile47 in complexes with PIs **12**, **14** and **16** incorporating the (*S*)-4-(1-hydroxyethyl)benzene P2′ moiety and toward Val32 and Leu76 in the corresponding (*R*)-epimers **13**, **15** and **17**. Compounds with the same P1′ group show minor differences in vdW interactions with residues in the S2′ subsite depending on the configuration of the P2′ hydroxyl group. The methyl group of analogue **12** with the (*S*)-configuration of the hydroxyl group makes additional vdW interactions with the methylene portion of Asp30′ side chain compared to the parent compound **2**. Surprisingly, there is no increase in vdW contacts with Ile47 even though the methyl group of compound **12** is oriented toward this residue. The methyl group of the P2′ (*R*)-epimer **13** forms vdW contacts with Ile47 and Leu76. The corresponding analogues with the (*S*)-2-methylbutyl and 2-ethylbutyl P1′ groups showed largely similar interactions as observed for **12** and **13**, respectively, but a slight shift in the benzene ring caused minor differences in the vdW contacts. As previously shown for parent compounds **3** and **4**, larger (*S*)-2-methylbutyl and 2-ethylbutyl P1′ groups result in increased vdW contacts in the S1′ subsite,

compared to the corresponding isobutyl group. Compounds with the same P1′ group show similar vdW contacts in the S1′ subsite. Thus, enhanced vdW interaction of PIs **12**–**17** in the S2′ subsite result from the additional methyl group and altered position of the modified P2′ moiety.

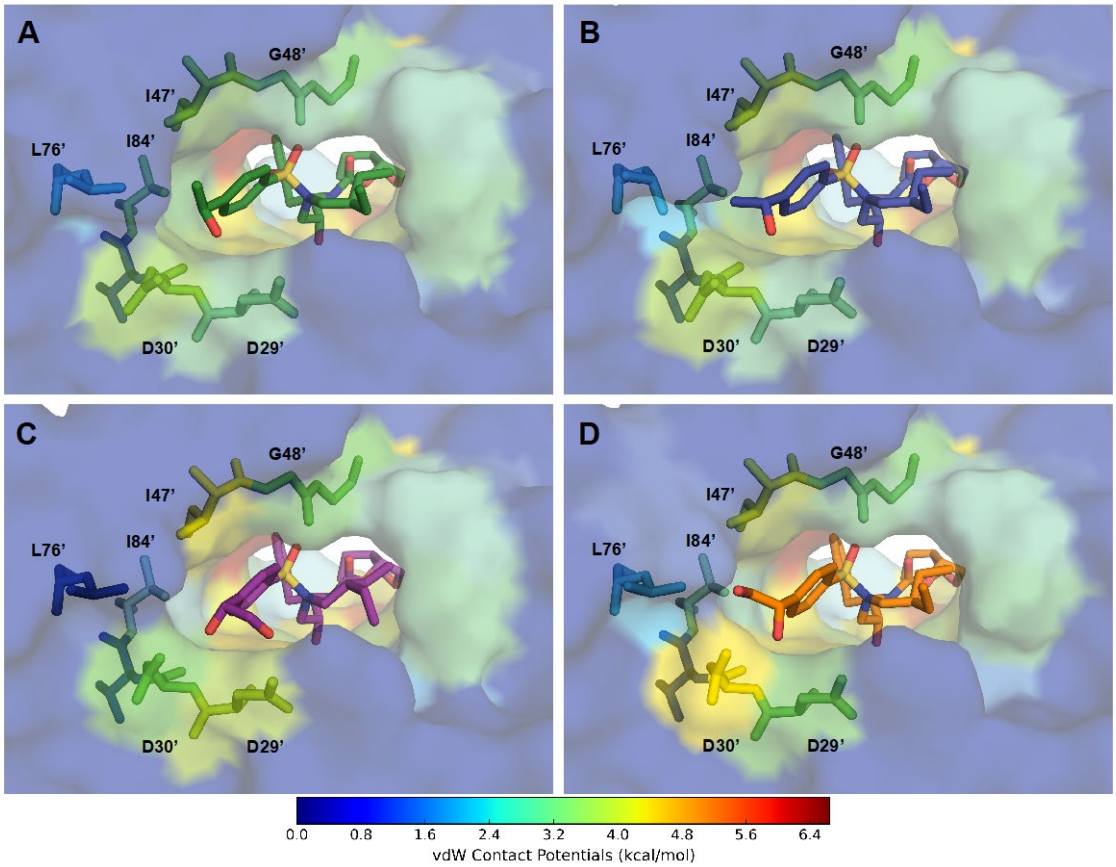


Figure 4.8 Packing of inhibitors (A) 16, (B) 17, (C) 26, and (D) 27 in the S2′ subsite of HIV-1 protease. The protease residues are colored blue to red for increasing van der Waals (vdW) contact potentials with the inhibitor mapped onto the surface of cocrystal structures. Inhibitors and key residues are shown as sticks.

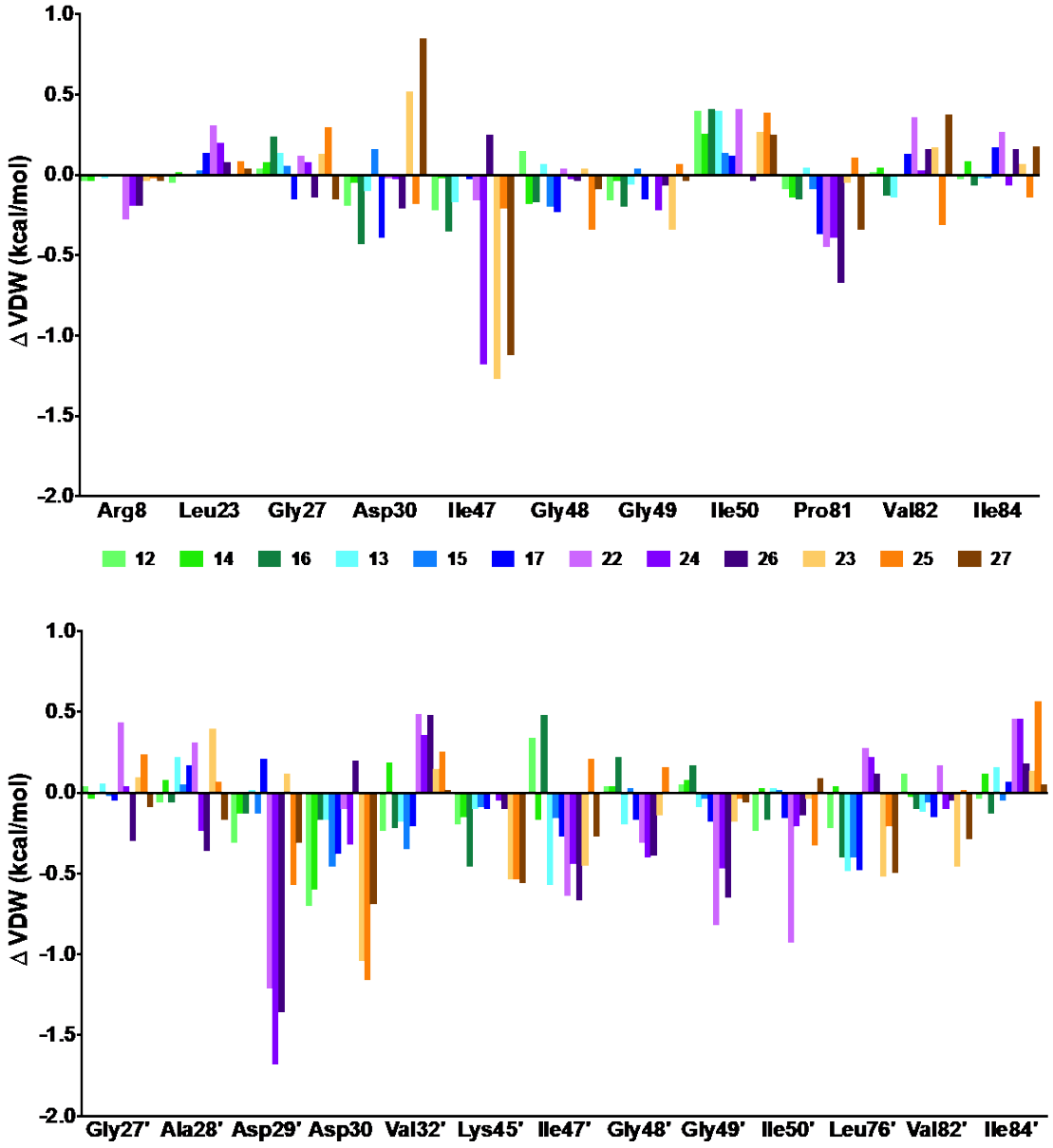


Figure 4.9 Differences in vdW interactions for PIs incorporating the modified P2′ moieties compared to the corresponding parent compound. Differences in per residue vdW interactions were calculated by subtracting the corresponding parent compound from each analogue.

While compounds with the (*S*)- and (*R*)-4-(1-hydroxyethyl)benzene P2′ moieties show minor changes in vdW contacts in the S2′ subsite, both diastereoisomers make identical hydrogen bonding interactions with the protease because the orientation of the hydroxyl group is similar in both configurations. The largely similar overall binding interactions of P2′ epimeric compounds correlates with the similar enzyme inhibitory potencies observed between diastereoisomers. PIs with the (*S*)-4-(1-hydroxyethyl)benzene P2′ moiety showed improved antiviral potency against MDR variants likely due to stronger polar interactions with the backbone atoms without relying on contacts with side chains that could mutate to cause resistance. Overall, the cocrystal structures provided insights into the binding of these new PIs with modified P2′ moieties, revealing the structural basis for the observed inhibitory potencies.

The cocrystal structures of PIs **22**–**27** incorporating the 4-(1,2-dihydroxyethyl)benzene in complex with protease revealed a larger shift in the position of the P2′ moiety compared to DRV [**Figure 4.7**]. The benzene ring is moved toward the flaps, likely to accommodate the larger substituent at the 4-position. With the (*R*)-configuration of the dihydroxyethyl group, the primary hydroxyl group in PIs **22**, **24** and **26** is oriented toward the Asp29′ backbone and the S3′ subsite. As a result, these PIs fit very well within the substrate envelope [**Figure 4.4C**]. In contrast, the primary hydroxyl group in the corresponding (*S*)-epimer compounds **23**, **25** and **27** is oriented in the opposite direction, toward the Asp30′ and Leu76 side chains, and lie largely outside the substrate envelope. The

cocrystal structures of PIs **22**–**27** also showed subtle variations in the binding of the secondary hydroxyl group of the (*R*)- and (*S*)-4-(1,2-dihydroxyethyl)benzene P2′ moieties. Despite differences in the binding of the dihydroxyethyl stereoisomers, similar conformations of the benzene ring were observed for both (*R*)- and (*S*)-4-(1,2-dihydroxyethyl)benzene P2′ moieties.

The hydrogen bonding patterns of compounds incorporating the 4-(1,2-dihydroxyethyl)benzene as P2′ ligands varied considerably depending on the configuration of the secondary hydroxyl group [**Figure 4.6**]. In the cocrystal structures of PIs **22**, **24**, and **26** with the (*R*)-configuration of the dihydroxyethyl group, the secondary hydroxyl group makes a strong, direct hydrogen bond with the backbone NH of Asp30′. Another weaker but direct hydrogen bond connects the secondary hydroxyl group with the side chain carboxylate of Asp30′, replacing the water-mediated interactions observed for compounds with the 4-(hydroxymethyl)- and 4-(1-hydroxyethyl)-benzene P2′ moieties. The primary hydroxyl group of the (*R*)-dihydroxyethyl moiety is oriented toward the protease backbone and forms a direct hydrogen bond with the backbone NH of Asp29′, displacing a water molecule observed in the parent compound structures [**Figure 4.10**]. The primary hydroxyl group is also involved in weaker (3.3–3.5 Å), direct hydrogen bonding interactions with the side chain carboxylate of Asp29′. The unique pattern of polar interactions of the (*R*)-dihydroxyethyl group in the S2′ subsite, particularly involving the backbone NH of Asp29′ and Asp30′, mimics the interactions of the *bis*-THF moiety in the S2 subsite.

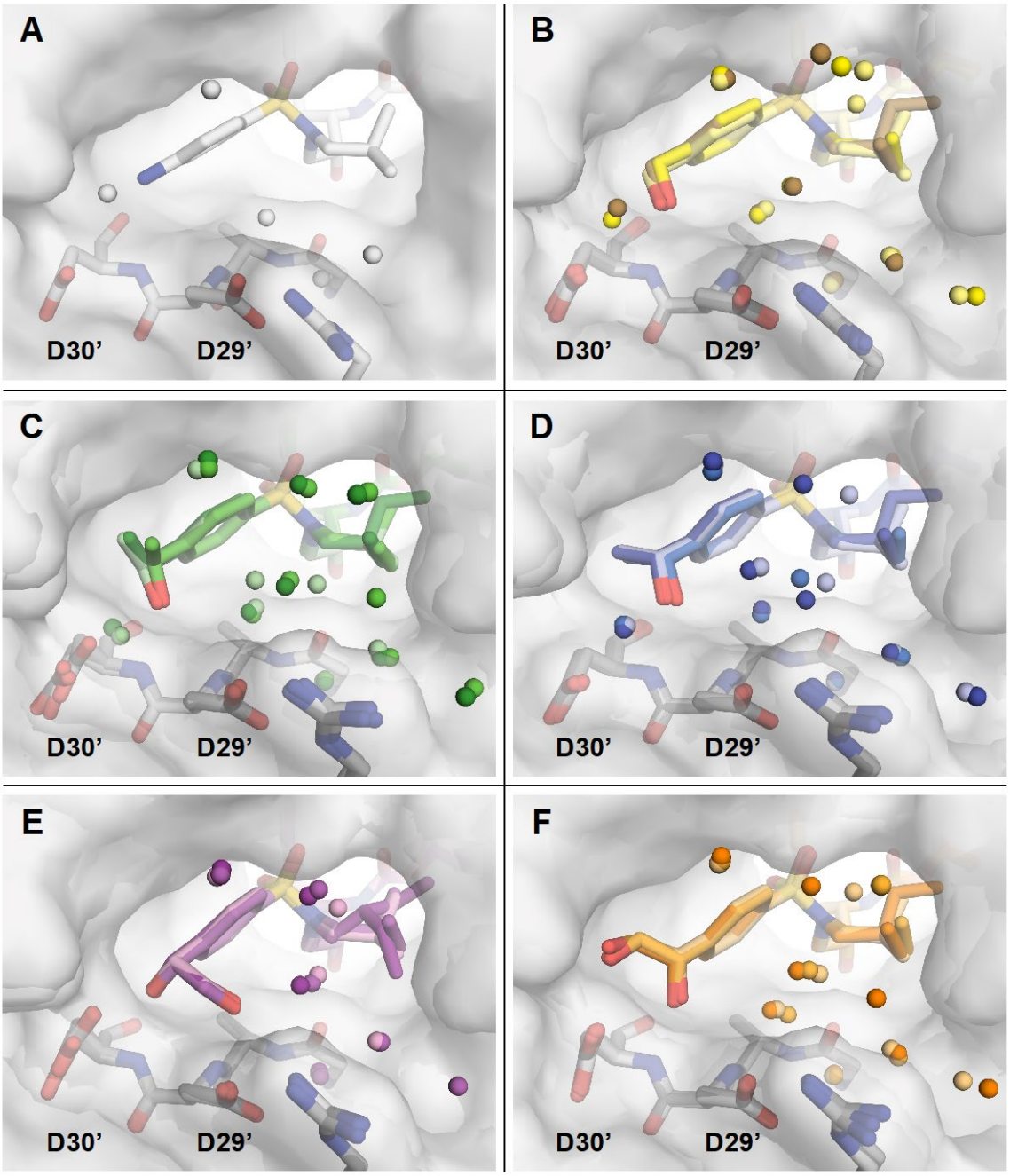


Figure 4.10 Comparison of water networks in the S2′ subsite. Most crystallographic waters are conserved in protease cocrystal structures of inhibitors with similar P2′ moieties. (A) Water structure in DRV-protease complex. Superimposed structures of (B) parent compounds **2**–**4**; (C) analogues **12**, **14** and **16**; (D) **13**, **15** and **17**; (E) **22**, **24** and **26**; and (F) **23**, **25** and **27**.

The cocrystal structures of PIs **23**, **25**, and **27** with the (*S*)-configuration of the P2′ dihydroxyethyl group also revealed a unique network of hydrogen bonds between the hydroxyl groups and the protease. The secondary hydroxyl group forms two direct hydrogen bonds, one with the backbone NH of Asp30′ and another with the side chain carboxylate of the same residue. The secondary hydroxyl group is also positioned close to the backbone NH of Asp29′ (3.5–3.6 Å) and makes water-mediated interactions with both the backbone NH and side chain carboxylate of this residue. The primary hydroxyl group is oriented toward the side chain of Asp30′ and is involved in a direct hydrogen bonding interaction with the carboxyl group, replacing the water molecule observed in complexes with compounds incorporating the 4-(hydroxymethyl)- and 4-(1-hydroxyethyl)-benzene P2′ moieties. In this orientation, the dihydroxyethyl group forms an additional direct hydrogen bond with the protease replacing a water-mediated interaction.

In addition, PIs **22**–**27** incorporating the 4-(1,2-dihydroxyethyl)benzene moieties as P2′ ligands form enhanced vdW contacts with the protease compared to the corresponding parent compounds **2**–**4**. The altered position of the P2′ moiety places the benzene ring closer to Gly48′ in the flaps, resulting in an overall increase in vdW contacts for all PIs in this series. Minor changes in vdW interactions were observed between P2′ epimeric compounds, with additional subtle variations resulting from changes at the P1′ position [**Figure 4.9**]. In general, compounds with the (*R*)-configuration of the modified P2′ moiety **(22**, **24**, and **26**)

make enhanced vdW interactions with Asp29′, Pro81, and residues 47′–50′ in the flap region, but slightly reduced interactions with Val32 and Ile84 [**Figure 4.8**]. The corresponding (*S*)-epimers (**23**, **25**, and **27**) formed more vdW contacts with a different set of residues around the active site, including Asp30′, Lys45′, Ile47, and Leu76′, and weaker contacts with Asp30 and Ile50. However, few exceptions to these general trends were observed for compounds **24** and **25** that incorporate the (*S*)-2-methylbutyl group at the P1′ position, likely due to the asymmetric shape of the P1′ group [**Figure 4.9**]. Overall, for PIs incorporating the 4-(1,2-dihydroxyethyl)benzene P2′ moieties, the increase in vdW interactions with specific protease residues mainly depend on the configuration of the dihydroxyethyl group.

The cocrystal structures of PIs with the 4-(1,2-dihydroxyethyl)benzene P2′ moieties revealed unique hydrogen bonding patterns and vdW contacts that vary significantly, depending on the configuration of the modified P2′ moiety. The improved potency profiles of PIs with the (*R*)-4-(1,2-dihydroxyethyl)benzene P2′ moiety compared to the corresponding (*S*)-epimers likely result from enhanced backbone interactions with residues Asp29′ and Asp30′, as well as increased vdW contacts with the side chains of Asp29′ and residues in the flap region. Notably, these PIs also fit very well in the substrate envelope. Compound **26**, with the 2-ethylbutyl P1′ group, is a promising lead for exploring further modifications,

particularly in combination with more hydrophobic P1 and P2 groups, to further improve potency against MDR variants.

4.5 Conclusions

New HIV-1 protease inhibitors were designed by incorporating stereoisomers of the 4-(1-hydroxyethyl)benzene and 4-(1,2-dihydroxyethyl)benzene moieties as P2′ ligands to enhance hydrogen bonding interactions in the S2′ subsite. Compounds with the polar dihydroxyethyl group at the P2′ moiety were generally less potent than DRV while inhibitors with the hydroxyethyl group exhibited improved antiviral potency against MDR variants. Crystal structures of protease-inhibitor complexes show that all PIs make enhanced hydrogen bonding and vdW interactions with the protease. The reduced cellular potency of PIs with the dihydroxyethyl group is likely due to their relatively low hydrophobicity, which can be counterbalanced with increased hydrophobicity at other positions such as P1. Overall, the PI potency profile was determined by the polar substituent at the 4-position of the P2′ benzene ring, configuration of the secondary hydroxyl group, and the size of the P1′ group. Exploration of both stereoisomers of P2′ moiety revealed unique polar and vdW interactions, including a network of direct and water-mediated hydrogen bonding with the backbone and side chain atoms of Asp29′ and Asp30′. Notably, PIs with the (*R*)-4-(1,2-dihydroxyethyl)benzene moiety make hydrogen bonding interactions with the backbone NH of Asp29′ and Asp30′ in the S2′ subsite, mimicking the polar interactions of the P2 *bis*-THF moiety, and show improved potency and resistance profiles compared to the

corresponding (*S*)-epimers. In general, compounds that make increased hydrogen bonding interactions with the backbone and vdW contacts with invariant residues, and fit better within the substrate envelope maintain better potency against highly resistant MDR HIV-1 strains. These SAR data and structural insights may allow further optimization of these promising inhibitors.

4.6 Experimental Section

4.6.1 General

All reactions were performed in oven-dried round- bottomed fitted with rubber septa under argon atmosphere unless otherwise noted. All reagents and solvents, including anhydrous solvents, were purchased from commercial sources and used as received. Flash column chromatography was performed on an automated Teledyne ISCO CombiFlash Rf+ system equipped with a UV-Vis detector using disposable Redisep Gold high performance silica gel columns, or manually using silica gel (230−400 mesh, EMD Millipore). Thin- layer chromatography (TLC) was performed using silica gel (60 F- 254) coated aluminum plates (EMD Millipore), and spots were visualized by exposure to ultraviolet light (UV), exposure to iodine adsorbed on silica gel, and/or staining with alcohol solutions of phosphomolybdic acid (PMA) and ninhydrin followed by brief heating. ^{1}H NMR and ^{13}C NMR spectra were acquired on Varian Mercury 400 MHz and Bruker Avance III HD 500 MHz NMR instruments. Chemical shifts are reported in ppm (δ scale) with the residual solvent signal used as reference, and coupling constant (*J*) values are reported in hertz (Hz). Data are presented as follows: chemical shift, multiplicity (s = singlet, d

= doublet, dd = doublet of doublet, dd = doublet of triplet, t = triplet, m = multiplet, br s = broad singlet), coupling constant in Hz, and integration. High-resolution mass spectra (HRMS) were recorded on a Thermo Scientific Orbitrap Velos Pro mass spectrometer coupled with a Thermo Scientific Accela 1250 UPLC and an autosampler using electrospray ionization (ESI) in the positive mode. The purity of final compounds was determined by analytical HPLC and was found to be ≥95% pure. HPLC was performed on an Agilent 1200 system equipped with a multiple wavelength detector and a manual injector under the following conditions: column, Phenomenex Hypersil-BDS-5u-C18 (5 μm, 4.6 mm × 250 mm, 130 Å); solvent A, H_2O containing 0.1% trifluoroacetic acid (TFA); solvent B, CH_3CN containing 0.1% TFA; gradient, 20% B to 100% B over 15 min followed by 100% B over 5 min; injection volume, 20 μL; flow rate, 1 mL/min. The wavelengths of detection were 254nm and 280 nm. Retention times and purity data for each target compound and intermediate can be found online.

4.6.2 Antiviral Assays

293T and TZM-BL[181] cells (NIH AIDS Research and Reference Reagent Program) were maintained in Dulbecco's modified Eagle's medium supplemented with 10% fetal calf serum in the presence of penicillin and streptomycin at 37 °C with 5% CO_2. To determine the concentration of drugs achieving 50% inhibition of infection compared with the drug-free control, 4.5×10^6 293T cells were seeded onto a 10-cm plate 24 h before transfection. Cells were transfected with 8 μg of either the wild-type plasmid, infectious molecular clone pNL-CH derived from the

pNL4-3 clone of HIV-1, or each PI-resistant HIV-1 variant[179] using FuGENE 6 transfection reagent (Roche). The culture supernatant of 293T cells transfected with wild-type or PI-resistant HIV-1 variant was removed 18 h after transfection and the cells were washed with 1 × PBS. The 293T cells were collected and transferred to wells of a 24-well plate. Briefly, each drug was serially diluted in the culture medium and the dilutions were added to the wells of a 24-well plate. The 293T cells (0.5×10^6 per well) collected from the transfection were added to wells containing various concentrations of drug. The culture supernatant containing virus particles was harvested 18 h after the 293T cells were reseeded in the presence of drug. This supernatant was filtered through a 0.45-μm-pore-size membrane (Millipore) to remove cell debris then used to infect 2×10^4 TZM-BL cells in a 96-well plate following a procedure previously described.[182] The culture supernatant was removed from each well 48 h post-infection, and the cells were washed with 1 × PBS. For the luciferase assay, infected TZM-BL cells were lysed in 1× reporter lysis buffer (Promega) and the cells were kept at −80 °C. After one freeze-thaw cycle, the cell lysates were transferred into a 96-well assay plate (Costar), and luciferase activity was measured using a luminometer (Promega). The culture supernatant harvested from 293T cells reseeded in the absence of drugs was used as a drug-free control. EC_{50} was determined based on a dose-response curve generated using GraphPad Prism (version 7.0).

4.6.3 Protease Gene Construction

Protease gene construction was carried out as previously described.[183-184] The NL4-3 strain has four naturally occurring polymorphisms in the protease relative to the SF2 strain.[100] In short, the protease variant genes (I84V, I50V/A71V) were constructed using QuikChange site-directed mutagenesis (Genewiz) onto NL4-3 wild-type protease on a pET11a plasmid with codon optimization for protein expression in Escherichia coli. A Q7K mutation was included to prevent autoproteolysis.[185]

4.6.4 Protein Expression and Purification

The expression, isolation, and purification of WT and mutant HIV-1 proteases used for the kinetic assays and crystallization were carried out as previously described.[183-184] Briefly, the gene encoding the HIV protease was subcloned into the heat-inducible pXC35 expression vector (ATCC) and transformed into E. coli TAP-106 cells. Cells grown in 6 L of Terrific Broth were lysed with a cell disruptor and the protein was purified from inclusion bodies.[186] The inclusion body centrifugation pellet was dissolved in 50% acetic acid followed by another round of centrifugation to remove impurities. Size exclusion chromatography was used to separate high molecular weight proteins from the desired protease. This was carried out on a 2.1 L Sephadex G-75 superfine column (Millipore Sigma) equilibrated with 50% acetic acid. The cleanest fractions of HIV protease were refolded into a 10-fold dilution of 0.05 M sodium acetate at pH 5.5, 5% ethylene glycol, 10% glycerol, and 5 mM DTT. Folded protein was concentrated down to

0.7–2 mg/mL and stored. This stored protease was used in K_M and K_i assays. For crystallography, a final purification was performed with a Pharmacia Superdex 75 FPLC column equilibrated with 0.05 M sodium acetate at pH 5.5, 5% ethylene glycol, 10% glycerol, and 5 mM DTT. Protease fractions purified from the size exclusion column were concentrated to 1–2 mg/mL using an Amicon Ultra-15 10-kDa device (Millipore) for crystallization.

4.6.5 Enzyme Inhibition Assays

The K_M and K_i Assays were carried out as previously described.[89, 177] In the K_M assay, a 10-amino acid substrate containing the natural MA/CA cleavage site with an EDANS/DABCYL FRET pair was dissolved in 8% DMSO at 40 nM and 6% DMSO at 30 nM. The 30 nM of substrate was 4/5th serially diluted from 30 nM to 6 nM, including a 0 nM control. HIV protease was diluted to 120 nM and, using a PerkinElmer Envision plate reader, 5 µL was added to the 96-well plate to obtain a final concentration of 10 nM. The fluorescence was observed with an excitation at 340 nm and emission at 492 nm and monitored for 200 counts, for approximately 23 minutes. FRET inner filter effect correction was applied as previously described.[187] Corrected data were analyzed with Prism7, as described in Chapter 3.

To determine the enzyme inhibition constant (K_i), in a 96-well plate, each inhibitor was 2/3 serially diluted from 3 nM to 52 pM, including a 0 pM control, and incubated with 0.35 nM protein for 1 hour. A 10-amino acid substrate containing an optimized protease cleavage site with an EDANS/DABCYL FRET pair was

dissolved in 4% DMSO at 120 μM. Using the Envision plate reader, 5 μL of the 120 μM substrate was added to the 96-well plate to a final concentration of 10 μM. The fluorescence was observed with an excitation at 340 nm and emission at 492 nm and monitored for 200 counts, for approximately 60 min. Data was analyzed with Prism7, As described in Chapter 2.

4.6.6 Protein Crystallization

The condition reliably producing cocrystals of NL4-3 WT protease bound to PIs was discovered and optimized as previously described.[188] Briefly, all cocrystals were grown at room temperature by hanging drop vapor diffusion method in a 24-well VDX hanging-drop trays (Hampton Research) with a protease concentration of 1.0–1.7 mg/mL with 3-fold molar excess of inhibitors. Crystallization drops 1 μL protein-inhibitor solution and 2 μL reservoir solution consisting of 22–26% (w/v) ammonium sulfate with 0.1 M bis-Tris-methane buffer at pH 5.5. Drops were micro-seeded with a cat whisker. Diffraction quality crystals were obtained within 1 week. As data were collected at 100 K, cryogenic conditions contained the precipitant solution supplemented with 25% glycerol.

4.6.7 X-Ray Data Collection and Structure Solution

X-ray diffraction data were collected and solved as previously described.[184, 188-189] Diffraction quality crystals were flash frozen under a cryostream when mounting the crystal at our in-house Rigaku_Saturn944 X-ray system. The cocrystal diffraction intensities were indexed, integrated, and scaled using HKL3000.[190] Structures were solved using molecular replacement with

PHASER.[191] Model building and refinement were performed using Coot[192] and Phenix.[193] Ligands were designed in Maestro and the output sdf files were used in the Phenix program eLBOW[109] to generate cif files containing atomic positions and constraints necessary for ligand refinement. Iterative rounds of crystallographic refinement were carried out until convergence was achieved. To limit bias throughout the refinement process, five percent of the data were reserved for the free R-value calculation.[110] MolProbity[194] was applied to evaluate the final structures before deposition in the PDB. Structure analysis, superposition and figure generation was done using PyMOL.[195] X-ray data collection and crystallographic refinement statistics are presented in the Supporting Information [**Table 4.4**].

4.6.8 Molecular Modeling

Molecular modeling was carried out using MacroModel.[196] Briefly, inhibitors were modeled into the active site of wild-type HIV-1 protease using the protease-3 and protease-4 co-complex structures (PDB code: 3O9B and 3O9G).[169] Structures were prepared using the Protein Preparation tool in Maestro 11.[196] 2D chemical structures were modified with the appropriate changes using the Build tool in Maestro. Once modeled, molecular energy minimizations were performed for each inhibitor–protease complex using the PRCG method with 2500 maximum iterations and 0.05 gradient convergence threshold. PDB files of modeled complexes were generated in Maestro for structural analysis.

4.7 Acknowledgments

This work was supported by a grant from the National Institute of General Medical Sciences of the NIH (P01 GM109767). We thank Kaneka USA for generous gifts of chiral epoxides, Ellen A. Nalivaika for providing the HIV-1 protease variants, the AIDS Research and Reference Reagent Program (NIAD, NIH) for reference protease inhibitors, and Drs. William Royer, Brian Kelch and members of the Schiffer and Miller laboratories for helpful discussions.

5 CHAPTER 5:

Structural Analysis of Potent Hybrid HIV-1 Protease Inhibitors Containing Bis-Tetrahydrofuran in a Pseudo-symmetric Dipeptide Isostere

5.1 Abstract

Drug discovery efforts against HIV-1 protease have focused mainly on exploring darunavir (DRV) analogues with similar sulfonamide-based dipeptide isosteres and P2/P2′ moieties. The conformationally constrained P2 *bis*-tetrahydrofuran (*bis*-THF) of DRV and related ligands have not been extensively explored in other dipeptide isosteres. In this study, we investigated HIV-1 protease inhibitors (PIs) containing *bis*-THF in a pseudo-C_2-symmetric dipeptide isostere and characterized their binding using experimental and computational methods. Specifically, series of hybrid PIs were designed by incorporating *bis*-THF on either side of the Phe-Phe isostere of lopinavir in combination with hydrophobic amino acids on the opposite P2/P2′ position. Enzyme inhibition and antiviral data indicated that the *bis*-THF moiety can be attached at either the P2 or P2′ position without significantly affecting potency. In contrast, the moiety on the opposite P2 or P2′ position had a dramatic effect on potency depending on the size and shape of the hydrophobic amino acid. Cocrystal structures of inhibitors with wild-type HIV-1 protease revealed the *bis*-THF moiety retained similar interactions as observed in the DRV-protease complex regardless of position on the Phe-Phe isostere. Analyses of cocrystal structures and molecular dynamics simulations provide insights for optimizing HIV-1 PIs containing *bis*-THF in non-sulfonamide dipeptide isosteres.

5.2 Introduction

Combinations of drugs targeting HIV-1 enzymes essential for viral replication have been highly effective in reducing viral load in infected individuals and significantly increasing their life expectancy.[152] HIV-1 protease inhibitors (PIs) remain essential components of combination antiretroviral therapy (cART) for AIDS patients. There are nine HIV-1 PIs currently approved by the US FDA; however, clinical use of most of these has diminished over the years due to side effects, unfavorable pharmacokinetics, and drug resistance. Currently only three PIs, darunavir (DRV), lopinavir (LPV), and atazanavir (ATV) [**Figure 5.1**] – boosted with small doses of ritonavir (RTV) – are being used in cART as they have comparatively higher potency and better resistance profiles.[126, 158]

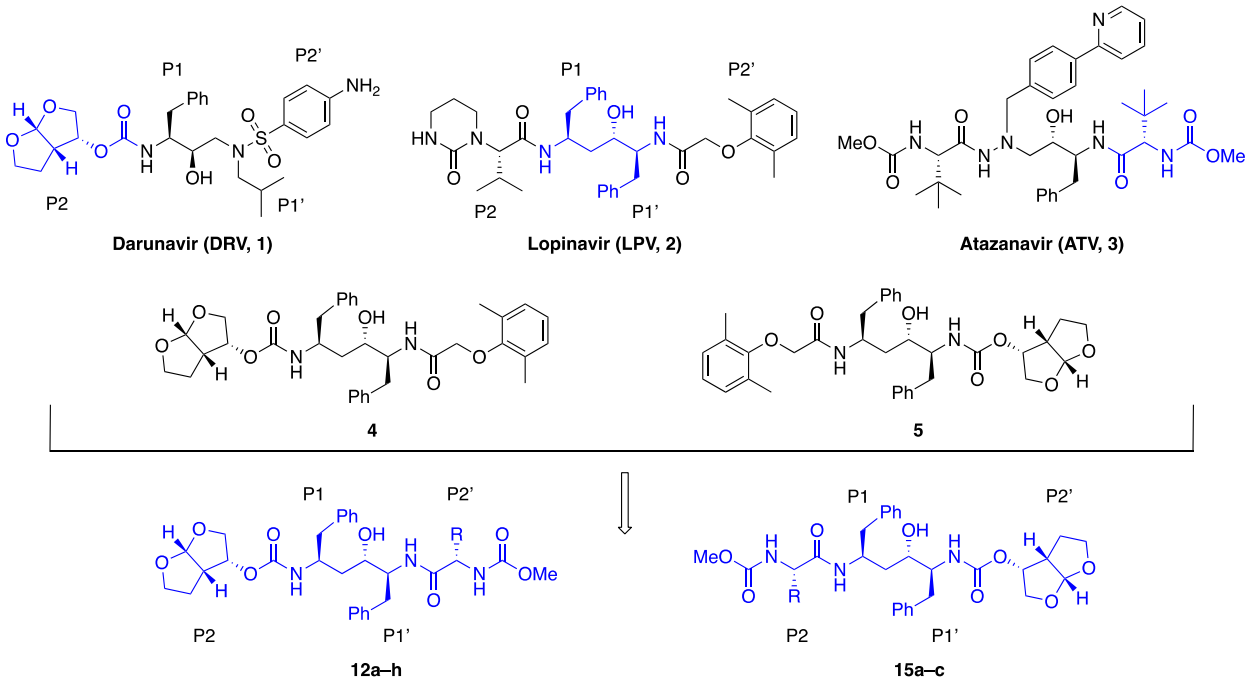


Figure 5.1 Structures of HIV-1 protease inhibitors darunavir (DRV) (1), lopinavir (LPV) (2), atazanavir (ATV) (3), previously reported hybrid compounds 4 and 5, and new hybrid compounds designed by incorporating the *bis*-THF moiety of DRV into the pseudo-symmetric Phe-Phe dipeptide isostere of LPV at the P2 (12a–h) or P2′ position (15a–c), in combination with an ATV-like hydrophobic amino acid moiety at the P2′ or P2 position, respectively. The canonical nomenclature is used to indicate the position of each inhibitor moiety.

HIV-1 protease is a C_2-symmetric homodimer with a pair of catalytic aspartic acid residues and a series of subsites forming the active site along the dimer interface. The protease recognizes and cleaves a variety of substrates, which adopt a conserved shape when bound in the active site defined as the substrate envelope.[43] A series of diverse amino acid sequences can be recognized as specific substrates due to dynamic subsite interdependence within the active site.[197] Most HIV-1 PIs are substrate-based peptidomimetics containing a hydroxyl group as a transition state mimetic and diverse moieties that target the S1/S1′ and S2/S2′ subsites.[27, 148] Of these PIs, DRV is the most potent with the highest barrier to resistance and efficacy against multidrug-resistant HIV-1 strains.[78] The excellent potency and resistance profile of DRV is attributed to the *bis*-tetrahydrofuran (*bis*-THF) moiety, which forms strong hydrogen bonding interactions with the main-chain NH of Asp29 and Asp30 in the S2 subsite of the enzyme.[198] Moreover, the extensive van der Waals (vdW) interactions of the *bis*-THF moiety, along with the sulfonamide-based Phe-Leu dipeptide isostere and the 4-aminobenzene group, also significantly contribute to the potency of DRV [**Figure 5.1**].[161]

Efforts to improve potency, pharmacokinetics, and resistance profiles of HIV-1 PIs have led to the discovery of exceptionally potent compounds exhibiting a wide range of properties.[148, 199-201] The classical design approaches of elaborating existing drugs and combining pharmacophores to form new hybrid scaffolds while employing structural insights from crystal structures have been the primary tools used to design these new inhibitors.[148, 202] In recent years, drug discovery efforts against HIV-1 protease have focused mainly on sulfonamide-based compounds similar to DRV. As such, many analogues of DRV containing sulfonamide-based dipeptide isosteres with modifications

at P1/P1′ and P2/P2′ positions have been explored extensively, including novel *bis*-THF analogues with improved hydrogen bonding and vdW interactions with the protease.[80, 148, 203-204] Moreover, detailed structural analyses informing structure-based design of inhibitors with the DRV scaffold have been reported numerous times.[148, 161, 166] However, very few inhibitors containing the *bis*-THF moiety combined with other isosteres, in particular the pseudo-C_2-symmetric dipeptide isosteres of LPV and ATV, have been explored.[205]

The variety of dipeptide isosteres and P2/P2′ moieties in HIV-1 PIs hitherto developed suggest that protease subsites can be targeted with diverse ligands but may require careful consideration of subsite interdependence. The *bis*-THF moiety of DRV has emerged as a privileged ligand for targeting the S2 subsite especially when incorporated into the (*R*)-(hydroxyethylamino)sulfonamide isostere.[206] Despite the unique binding characteristics of *bis*-THF, only a few attempts have been made to use this ligand in combination with other dipeptide isosteres. Chen and colleagues at Abbott Laboratories explored PIs with *bis*-THF on either side of the pseudo-C_2-symmetric Phe-Phe dipeptide isostere of LPV,[205] while Cannizzaro and coworkers at Gilead Sciences designed similar analogues using the aza-dipeptide isostere of ATV.[207] Although some of the Phe-Phe isostere-based compounds, such as **4** and **5** [**Figure 5.1**], exhibited significant antiviral activity against wild-type HIV-1, in the absence of direct comparison, it is unclear if their potency is similar to that of DRV.[205] The stereochemical preference for the *bis*-THF moiety has been clearly established, with the same (3*R*)-stereoisomer preferred in the Phe-Phe isostere as in the DRV scaffold.[208] However, due to limited structure-activity relationship studies, the S2/S2′ subsite preference of the *bis*-THF moiety in the context of Phe-Phe

isostere remains unclear,[205, 208] as is the identity of an optimal moiety for targeting the opposite S2/S2′ subsite.[205] Furthermore, no cocrystal structures of such hybrid PIs have been reported, limiting opportunities for structure-guided design and optimization.

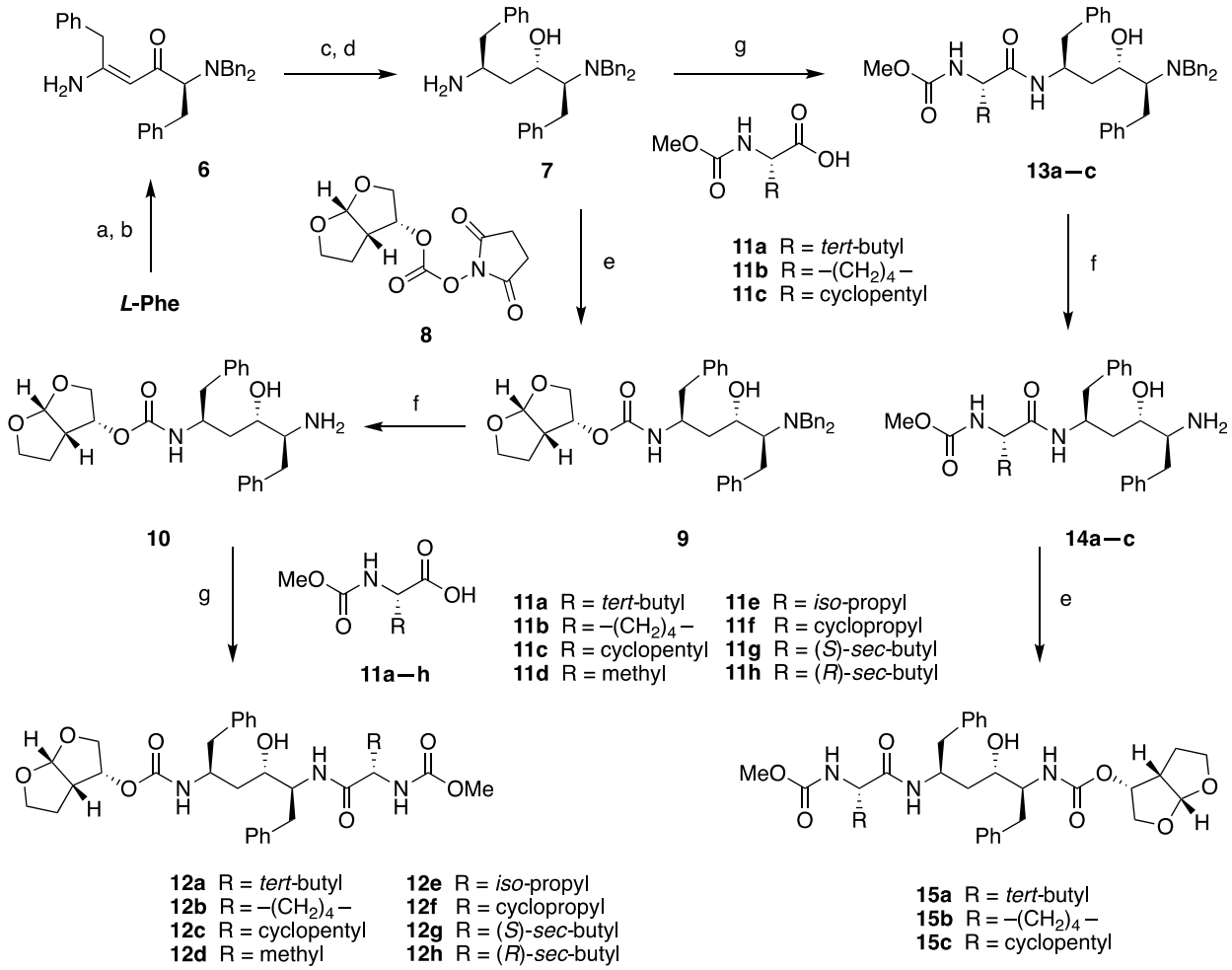
The active site of HIV-1 protease exhibits subsite interdependence in substrate recognition and inhibitor binding,[55, 197] but this feature of the protease has rarely been considered in inhibitor design. While identifying optimal moieties for targeting each of the subsites in the protease active site is important, understanding the subsite interdependence and selection of the best combination of moieties are also crucial to achieving higher potency especially against drug-resistant variants.[203] The *bis*-THF moiety potentially can be incorporated in diverse dipeptide isosteres in combination with suitable ligands targeting the other subsites. Here, we explored a series of hybrid HIV-1 PIs incorporating the *bis*-THF moiety of DRV on either side of the pseudo-C_2-symmetric Phe-Phe dipeptide isostere of LPV in combination with ATV-like hydrophobic P2/P2′ ligands of varying size and shape [**Figure 5.1**]. The hybrid PIs were tested for inhibitory activity against wild-type and two drug-resistant variants and for antiviral activity in cellular assays. We also determined high-resolution X-ray crystal structures of all designed hybrid compounds (and **4**) bound to wild-type HIV-1 protease, and characterized the binding interactions of *bis*-THF in the context of the Phe-Phe isostere. In addition, molecular dynamics simulations were utilized to assess the stability of protease-inhibitor interactions in solution. The comprehensive structural and dynamic analyses of hybrid PIs have provided insights into inhibitor binding and interdependence of protease subsites.

5.3 Chemistry

The synthesis of hybrid PIs incorporating the *bis*-THF moiety at either the P2 or P2′ position is outlined in **Scheme 5.1**. The required Phe-Phe dipeptide isostere core intermediate **7** was synthesized from phenylalanine following previously reported methods with minor modifications.[209] Reaction of the dipeptide isostere **7** with *bis*-THF activated carbonate **8** in the presence of DIEA provided intermediate **9**. Debenzylation of **9** using ammonium formate and 10% palladium on activated carbon in methanol gave the corresponding deprotected amine **10**, which was subsequently coupled with the amino acid derivatives **11a**–**h** using HATU and DIEA to provide the target compounds **12a**–**h** with *bis*-THF at the P2 position. While *N*-(methoxycarbonyl)-protected amino acid derivatives **11a** and **11e** were commercially available, intermediates **11b**–**d** and **11f**–**h** were prepared from the corresponding amino acids by reaction with methyl chloroformate under basic conditions following a reported procedure.[210]

The PIs **15a**–**c** with the *bis*-THF at the P2′ position were synthesized using a similar reaction sequence. Coupling of the dipeptide isostere **7** with acids **11a**–**c** under HATU peptide coupling conditions gave the corresponding intermediates **13a**–**c**. Catalytic debenzylation of compounds **13a**–**c** followed by coupling with *bis*-THF activated carbonate **8** in the presence of DIEA provided the target compounds **15a**–**c**. A pair of previously reported PIs **4** and **5** containing the dimethylphenoxy acetate and *bis*-THF moieties as the P2 and P2′ ligands in the Phe-Phe isostere were also synthesized using the same reaction sequences.

Scheme 5.1 Synthesis of Protease Inhibitors Incorporating *Bis*-Tetrahydrofuran in a Pseudo-Symmetric Phe-Phe Dipeptide Isostere



Reagents and conditions: (a) K_2CO_3, KOH, BnCl, H_2O, 100 °C, 5 h; (b) $NaNH_2$, CH_3CN, MTBE, 0 °C, 2 h, then BnMgCl (2 M in THF), RT, 14 h; (c) $NaBH_4$, MsOH, *i*-PrOH, EGDME, 0 °C, 12 h; (d) $NaBH_4$, TEA, DMA, 0–15 °C, 2 h; (e) DIEA, CH_3CN, RT, 24 h; (f) Pd/C, HCO_2NH_4, MeOH, 50 °C, 12 h; (g) *N*-(methoxycarbonyl)-capped amino acid, HATU, DIEA, DMF, RT, 4 h.

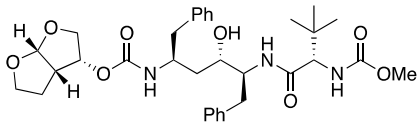
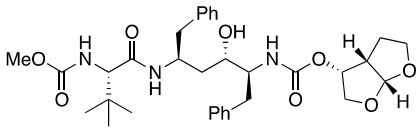
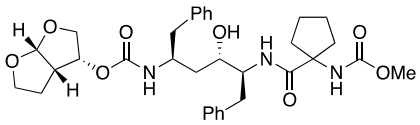
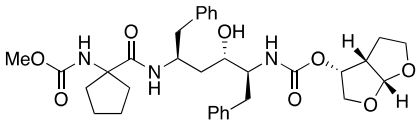
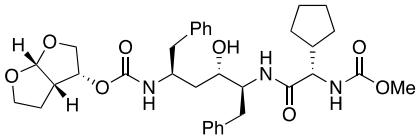
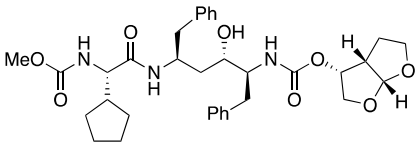
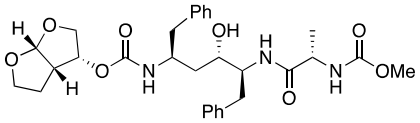
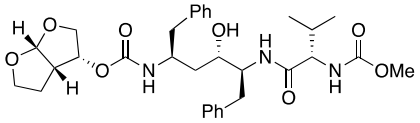
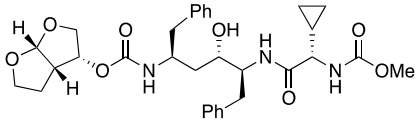
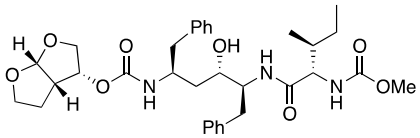
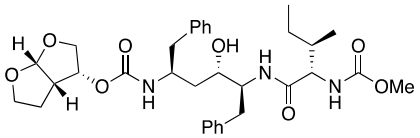
5.4 Results and Discussion

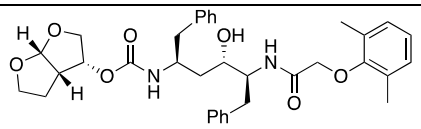
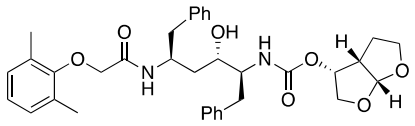
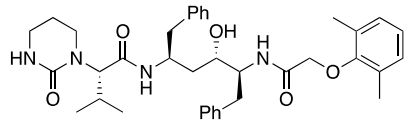
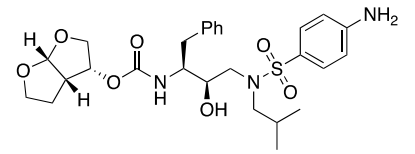
We explored hybrid HIV-1 PIs containing the *bis*-THF moiety of DRV in the Phe-Phe isostere of LPV, with the goal to characterize the binding interactions of this P2 ligand in the context of a pseudo-C_2-symmetric dipeptide isostere. The *bis*-THF moiety was used in combination with diverse *N*-(methoxycarbonyl)-capped amino acid derivatives similar to the P2/P2′ moiety of ATV to optimize hydrogen bonding and vdW interactions in both the S2 and S2′ subsites of HIV-1 protease. To explore subsite preference for the *bis*-THF moiety, pairs of compounds were synthesized with the *bis*-THF moiety attached at either the *distal* (P2) (**12a**–**c**) or *proximal* (P2′) (**15a**–**c**) position with respect to the central hydroxyl group of the Phe-Phe isostere. Considering potential subsite interdependence, incorporating a relatively small and conformationally flexible moiety at the other P2/P2′ position was expected to allow the P2/P2′ *bis*-THF moiety to optimally interact with the backbone atoms in the S2/S2′ subsite, likely improving overall inhibitor binding with the protease. Structure-activity relationships were explored by incorporating hydrophobic amino acids of varying size and shape to identify an optimal ligand for targeting the other S2/S2′ subsite.

5.4.1 Enzyme Inhibition and Antiviral Assays

The potency of PIs was assessed using biochemical and antiviral assays. The enzyme inhibition constants (K_i) were determined against wild-type protease and two drug-resistant variants (I84V and I50V/A71V) using a highly sensitive fluorogenic assay [**Table 5.1**].[90] For a subset of compounds, antiviral potencies (EC_{50}) were determined against wild-type HIV-1 (NL4-3 strain) using a cell-based antiviral assay. DRV was used as a control in all assays.

Table 5.1 Inhibitory Activity of PIs against WT HIV-1 Protease and Drug-Resistant Variants

Inhibitor	Structure	*K*i (nM)			EC_{50} (nM)
		WT	I84V	I50V/A71V	WT
12a		0.018 ± 0.003	0.208 ± 0.014	0.273 ± 0.043	11
15a		0.013 ± 0.001	0.372 ± 0.021	0.072 ± 0.005	39
12b		46 ± 3	236 ± 39	484 ± 70	NT
15b		7.3 ± 0.5	191 ± 30	40 ± 3	NT
12c		0.235 ± 0.016	0.71 ± 0.10	2.35 ± 0.39	NT
15c		0.161 ± 0.011	1.18 ± 0.26	1.39 ± 0.12	NT
12d		0.019 ± 0.004	0.069 ± 0.013	0.073 ± 0.003	66
12e		0.074 ± 0.006	0.302 ± 0.022	0.295 ± 0.082	55
12f		0.571 ± 0.076	1.82 ± 1.50	1.66 ± 0.48	NT
12g		0.055 ± 0.004	0.198 ± 0.014	0.560 ± 0.094	70
12h		0.027 ± 0.004	0.091 ± 0.013	0.029 ± 0.003	240

4		0.017 ± 0.002	0.069 ± 0.005	0.072 ± 0.002	38
5		0.037 ± 0.005	0.259 ± 0.055	NT	16
LPV		< 0.005	0.051 ± 0.004	0.061 ± 0.011	12
DRV		< 0.005	0.025 ± 0.006	0.075 ± 0.006	5

NT = Not Tested

Compounds **12a** and **15a**, with the *bis*-THF moiety at the P2 (distal) and P2′ (proximal) position, respectively, and the *N*-(methoxycarbonyl)-*tert*-leucine at the opposite P2/P2′ position, showed excellent inhibitory potencies against wild-type protease and the drug resistant variants. Compared to DRV and LPV, both compounds were less active, particularly against the I84V variant. Interestingly, compounds **12a** and **15a** exhibited similar K_i values, though the latter was 4-fold more active against the I50V/A71V variant, indicating no clear S2/S2′ subsite preference for the *bis*-THF moiety. However, in antiviral assays compound **12a** was 3-fold more potent than **15a** against wild-type HIV-1. The distal compound **12a**, with the *bis*-THF moiety at the P2 position, exhibited excellent antiviral potency (EC_{50} = 11 nM) comparable to that of LPV (EC_{50} = 12 nM) and DRV (EC_{50} = 5 nM). A similar trend was observed for previously reported compounds **4** and **5** containing the *bis*-THF and the dimethylphenoxy acetate moieties as the P2 and P2′ ligands. Compound **5** exhibited 2-fold better antiviral potency than **4**, indicating a reverse subsite preference for the *bis*-THF moiety relative to the **12a**/**15a** pair.

The small variation in the antiviral potencies of compounds **12a**/**15a** and **4**/**5** indicate that the position of the *bis*-THF moiety with respect to the central hydroxyl group of the Phe-Phe isostere, though not critical for binding, may affect potency in cellular assays.

Although the potency of the hybrid compounds was not particularly sensitive to the position of the *bis*-THF moiety, compounds with different hydrophobic amino acid derivatives at the opposite P2/P2′ position showed markedly different inhibitory potencies. The compounds with the cycloleucine moiety, **12b** and **15b**, were significantly less active (>2500-fold and 560-fold, respectively) than the corresponding *tert*-leucine analogues **12a** and **15a**. The dramatic loss of potency is likely due to the altered shape and flexibility of the hydrophobic amino acid moiety. This was further supported by the fact that the corresponding cyclopentylglycine analogues **12c** and **15c**, though also an order of magnitude less active than **12a** and **15a**, still exhibited sub-nanomolar to low nM potency against wild-type protease and drug-resistant variants. Together these data confirm that the hydrophobic amino acid moiety at the opposite P2/P2′ position significantly influence potency in both biochemical and antiviral assays.

Since potency of the compounds with *bis*-THF moiety attached to the Phe-Phe isostere largely depended on the hydrophobic amino acid moiety at the opposite side of the core scaffold, we investigated size and shape requirements for optimally targeting the S2′ subsite. Analogues of compound **12a** were synthesized with diverse hydrophobic amino acid derivatives as P2′ ligands while keeping the *bis*-THF moiety at the P2 position. The alanine analogue **12d** exhibited excellent inhibitory potency with *K*i values against resistant variants comparable to that of LPV and DRV. The corresponding valine analogue **12e** was 4-fold less active against wild-type protease but maintained similar

inhibitory potency as **12a** against the resistant variants. However, compared to **12a** compounds **12d** and **12e** were 6- and 5-fold less active in antiviral assays, respectively, indicating that factors other than binding affinity are responsible for lower antiviral potency. The cyclopropyl analogue **12f** was considerably less active than **12a**, with K_i values similar to the cyclopentyl analogue **12c**, strongly suggesting that cyclic moieties at the P2/P2′ position are unfavorable. Replacement of *tert*-leucine with an isoleucine (compound **12g**) or alloisoleucine (compound **12h**) moiety at the P2′ position also resulted in slightly lower activity against wild-type protease but similar or better potency against resistant variants. This further confirmed earlier findings that the P2/P2′ group opposite to the *bis*-THF moiety on the Phe-Phe core has a significant impact on potency. These results are in contrast to the sulfonamide-based dipeptide isostere of DRV, where diverse P2′ modifications result in relatively minor changes in potency.[78] The identity of the P2/P2′ moiety opposite to the *bis*-THF on the Phe-Phe core considerably impacts potency in both biochemical and antiviral assays.

5.4.2 Analysis of Protease-Inhibitor Complexes

To characterize protease-inhibitor molecular interactions, twelve high-resolution (1.8–2.0 Å) crystal structures of hybrid PIs and compound **4** bound to wild-type HIV-1 protease of the NL4-3 strain were determined. A cocrystal structure of LPV was also determined bound to the same protease enzyme for direct comparison. The crystallographic data collection and refinement statistics are summarized in **Table 5.2**. Following convention, protease chains were assigned chain A (non-prime) or chain B (prime) depending on the interactions between the central hydroxyl group of the Phe-Phe isostere and the catalytic aspartates Asp25/Asp25′.

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

Inhibitor	***DRV***	***LPV***	***4***	***12a***	***12b***	***12c***	***12d***	***12e***	***12f***	***12g***	***12h***	***15a***	***15b***	***15c***
PDB Code	6DGX	6PJB	6PJC	6PJD	6PJE	6PJF	6PJG	6PJH	6PJI	6PJK	6PJL	6PJM	6PJN	6PJO
Resolution (Å)	2.00	1.98	1.95	1.89	1.92	1.94	1.80	1.85	1.90	1.99	1.99	1.93	1.98	1.95
Space group	$P2_12_12_1$	$P2_12_12_1$	$P12_11$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Cell dimensions:														
a (Å)	51.0	51.4	51.7	51.3	51.4	51.3	51.3	51.5	51.1	51.3	51.5	51.3	51.3	51.4
b (Å)	58.3	57.8	62.5	58.3	58.5	58.1	58.6	58.7	58.6	58.8	58.2	58.8	58.9	58.6
c (Å)	61.9	62.3	59.8	62.1	61.9	61.5	62.0	62.0	61.7	61.9	61.7	61.8	61.8	61.7
β (°)	90	90	98.175	90	90	90	90	90	90	90	90	90	90	90
Completeness (%)	98.1 (94.5)	99.7 (98.8)	97.5 (94.2)	97.6 (88.2)	99.7 (94.8)	99.8 (98.2)	98.7 (97.6)	99.5 (99.9)	99.0 (98.6)	92.0 (93.4)	98.9 (99.2)	99.5 (92.2)	99.3 (90.9)	93.9 (88.1)
Total reflections	84332	91123	113979	60549	88652	93472	90337	80091	88965	47555	64344	86288	79533	74054
Unique reflections	12758	13342	26144	15065	14780	14151	17695	16774	15198	12050	12967	14564	13515	13293
Average I/σ	31.1 (3.2)	20.6 (4.2)	12.2 (2.5)	21.0 (3.9)	21.5 (3.6)	33.3 (3.9)	24.9 (2.9)	25.4 (2.3)	24.1 (2.1)	15.1 (2.2)	17.5 (2.5)	20.2 (3.9)	17.7 (3.8)	27.2 (7.4)
Redundancy	6.6 (4.4)	6.8 (5.9)	4.4 (3.5)	4.0 (2.0)	6.0 (3.1)	6.6 (5.0)	5.1 (5.0)	4.8 (4.8)	5.9 (5.5)	3.9 (3.7)	5.0 (4.7)	5.9 (3.5)	5.9 (3.9)	5.6 (4.0)
R_{sym} *(%)*[a]	5.7 (40.2)	8.1 (40.9)	10.1 (96.7)	6.6 (19.4)	8.3 (30.2)	5.2 (36.6)	8.4 (46.0)	8.0 (49.0)	9.4 (43.7)	9.8 (32.0)	12.4 (52.1)	9.0 (28.3)	9.1 (32.9)	6.5 (16.5)
RMSD[b] *in:*														
Bond lengths (Å)	0.003	0.003	0.009	0.008	0.005	0.004	0.011	0.005	0.008	0.011	0.007	0.003	0.003	0.003
Bond angles (°)	0.590	0.537	1.282	0.928	0.796	0.699	1.230	0.853	1.044	1.270	1.029	0.673	0.647	0.719
R_{factor} *(%)*[c]	19.2	19.7	23.9	17.0	18.2	20.7	18.1	18.5	18.8	17.9	20.0	17.4	17.6	18.9
R_{free} *(%)*[d]	22.8	22.5	27.4	21.9	22.8	24.6	21.5	21.9	22.5	23.7	23.1	21.5	22.2	23.9

[a]$R_{sym} = \Sigma \mid I - \langle I \rangle \mid / \Sigma I$, where I = observed intensity, $\langle I \rangle$ = average intensity over symmetry equivalent; values in parentheses are for the highest resolution shell.
[b]RMSD, root mean square deviation.
[c]$R_{factor} = \Sigma \mid\mid F_o\mid - \mid F_c\mid\mid / \Sigma \mid F_o\mid$.
[d]R_{free} was calculated from 5% of reflections, chosen randomly, which were omitted from the refinement process.

LPV and DRV achieve low picomolar affinity to wild-type HIV-1 protease by making a number of hydrogen bonds and vdW interactions with the active site residues [**Figure 5.2**]. The interactions of the transition state mimetic secondary hydroxyl group with the catalytic aspartates, together with the vdW interactions of the adjacent P1/P1′ moieties, are critical for overall inhibitor binding. Together these interactions determine the positions of the P2/P2′ moieties in the active site, which in turn determine the patterns of hydrogen bonding and vdW interactions of these moieties with the protease. DRV contains a sulfonamide-based Phe-Leu dipeptide isostere core with the (*R*)-configuration of the secondary hydroxyl group, whereas LPV has a pseudo-C_2-symmetric Phe-Phe dipeptide isostere with the (*S*)-configuration of the hydroxyl group. Though chemically quite distinct, the core scaffolds of LPV and DRV make largely similar hydrogen bonding and vdW interactions with the protease. These inhibitors, however, differ in their interactions with the protease in the S2 and S2′ subsites.

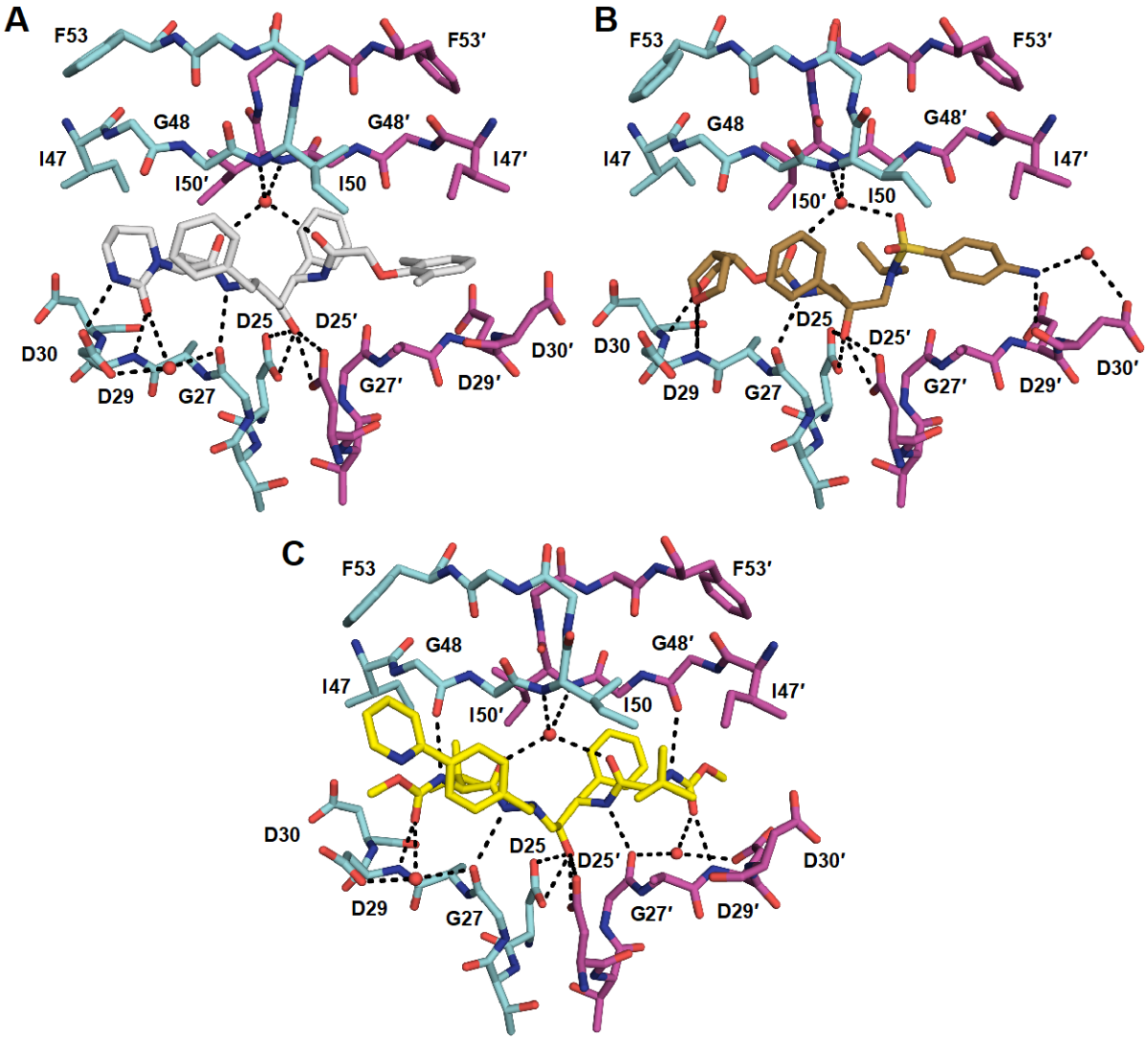


Figure 5.2 Binding mode of parent compounds (A) LPV and (B) DRV, and (C) ATV. Hydrogen bonds are shown as black dashes and waters are shown as red spheres.

The *bis*-THF moiety of DRV forms direct hydrogen bonds with the main-chain NH of Asp29 and Asp30 in the S2 subsite, while the 4-aminobenzene makes a direct hydrogen bond with the main-chain carbonyl of Asp30′ as well as a water-mediated hydrogen bond with the side-chain carboxylate of the same residue in the S2′ subsite [**Figure 5.2**]. The cyclic urea moiety of LPV makes direct hydrogen bonds with the main-chain NH and side-chain carboxylate of Asp29. The same moiety also makes water-mediated hydrogen bonds with the main-chain carbonyl of Gly27 and the side-chain carboxylate of Asp29 in the S2 subsite. However, in contrast to the P2′ moiety of DRV, the dimethylphenoxy acetate moiety of LPV only makes hydrophobic interactions in the S2′ subsite [**Figure 5.2**]. The hybrid compounds combine the pseudo-C_2-symmetric Phe-Phe isostere of LPV with the *bis*-THF moiety of DRV and an ATV-like amino acid moiety to enhance direct hydrogen bonding and vdW interactions in the S2 and S2′ subsites.

The cocrystal structures of hybrid compounds revealed similar binding conformations of the Phe-Phe isostere and the *bis*-THF moiety as observed in the protease complexes with LPV and DRV, respectively [**Figure 5.3**]. Importantly, the *bis*-THF moiety is positioned to interact with the main-chain NH of residues Asp29 and Asp30, mimicking the binding interactions of DRV [**Figure 5.4**]. This binding conformation is maintained in all hybrid compounds including **4** regardless of the position of the *bis*-THF moiety. As a result, within the distal and proximal compound series (with the *bis*-THF moiety at the P2 and P2′ position respectively), the structures aligned well with variations occurring only at the hydrophobic amino acid moiety [**Figure 5.5**].

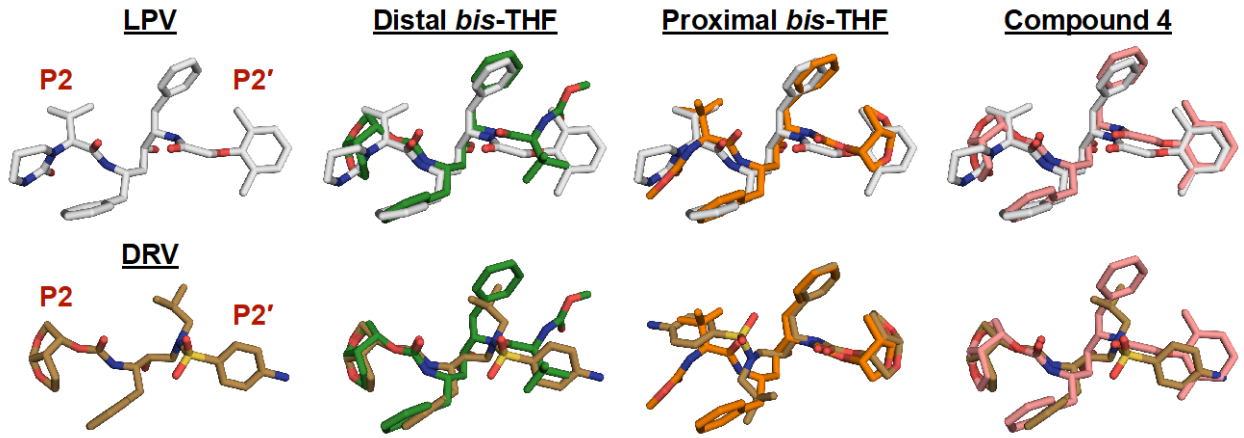


Figure 5.3 Overlays of protease cocrystal structures bound to LPV (grey, top) and DRV (brown, bottom) with compounds **12a** (green), **15a** (orange), and **4** (salmon) revealed similar binding conformations of the Phe-Phe isostere and the *bis*-THF moiety as observed in the protease complexes with LPV and DRV, respectively. Note that DRV is flipped 180º to compare with *bis*-THF of **15a** at the proximal position.

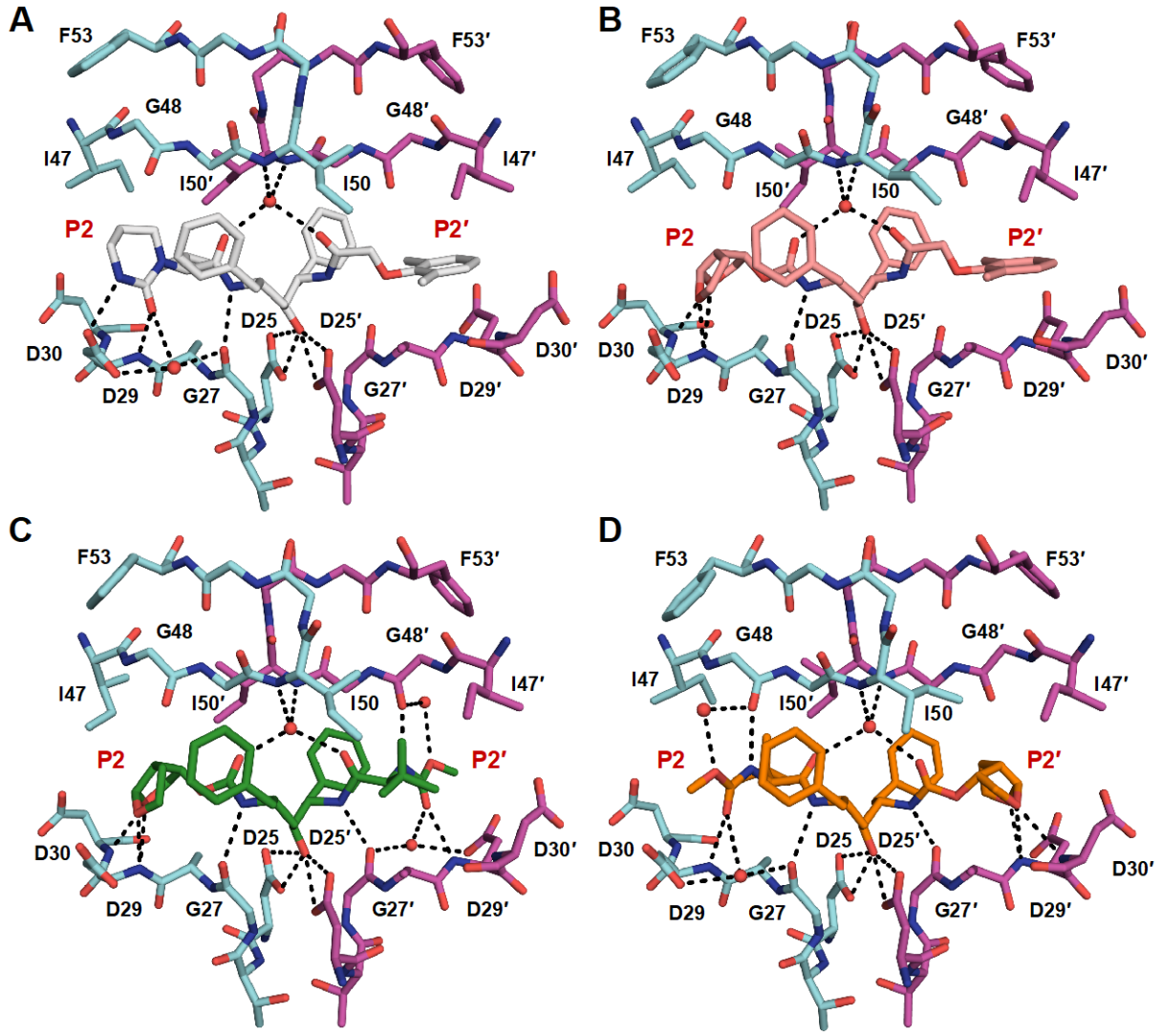


Figure 5.4 Crystal structures of wild-type HIV-1 protease in complex with (A) LPV and hybrid compounds (B) **4**, (C) **12a**, and (D) **15a**. The *bis*-THF moiety maintains the hydrogen bonding interactions with the main-chain NH of Asp29/Asp29′ and Asp30/Asp30′ in the S2/S2′ subsites regardless of the its position on the Phe-Phe core. Hydrogen bonds are shown as black dashes and waters are shown as red spheres.

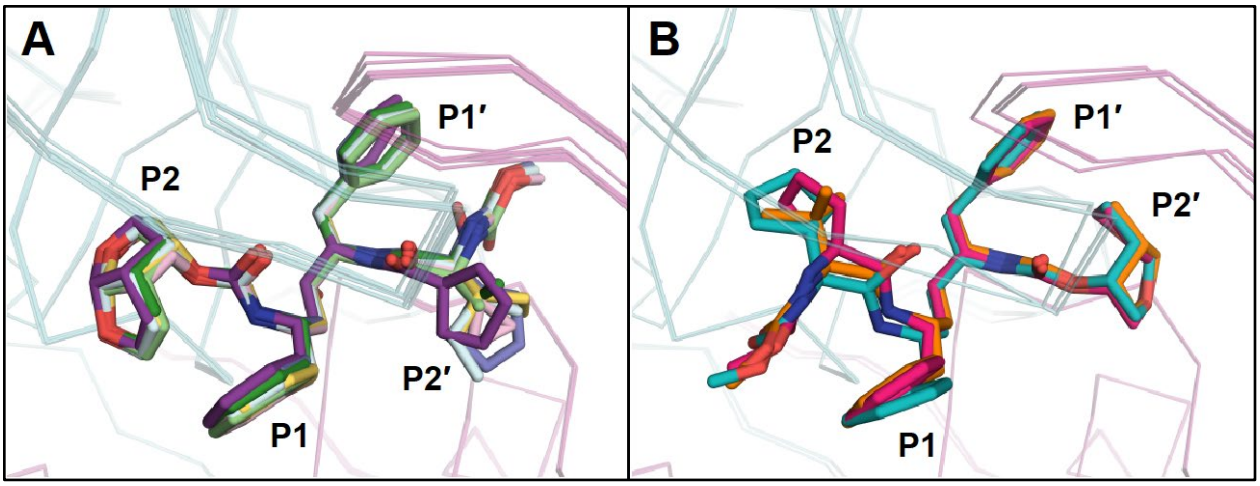


Figure 5.5 Superposition of HIV-1 protease complexes with (A) distal compounds **12a**–**12h** and (B) proximal compounds **15a**–**15c**, showing similar binding conformations of the *bis*-THF moiety regardless of the position, minor differences at the Phe-Phe isostere, but major differences at the P2/P2′ amino acid moiety. The protease active site is shown in a ribbon representation, with bound inhibitors depicted as sticks.

In all complexes, the central hydroxyl group of the Phe-Phe core maintains the key hydrogen bonds with the side-chain carboxylates of the catalytic residues Asp25/Asp25′. Nevertheless, subtle differences in hydrogen bonding and vdW interactions were observed between the Phe-Phe cores in LPV and the hybrid compounds. Notably, one of the four water-mediated interactions with the flap residue Ile50 is consistently longer (2.9–3.4 Å) than what is observed in the LPV structure (2.7 Å) [**Table 5.3**]. The two amide NH groups on each side of the Phe-Phe core could potentially interact with the backbone carbonyl of Gly27 and Gly27′, but in the LPV structure only the Gly27 is close enough to form a hydrogen bond (2.9 Å versus 3.9 Å) [**Figure 5.4**]. This phenomenon is also observed in the protease complex with compound **4** (2.9 Å versus 3.7 Å) indicating that the large hydrophobic P2′ moiety prevents the additional hydrogen bond from occurring with Gly27′. In contrast, in all hybrid compound structures the core NH groups are within 3.5 Å from the backbone carbonyls of Gly27 and Gly27′ [**Table 5.3**], although simultaneous interactions with both Gly27/Gly27′ may not be stable. Nevertheless, the combination of *bis*-THF and a flexible, hydrophobic amino acid moiety allow both core NH groups to interact with the main-chain carbonyls of residues Gly27/Gly27′.

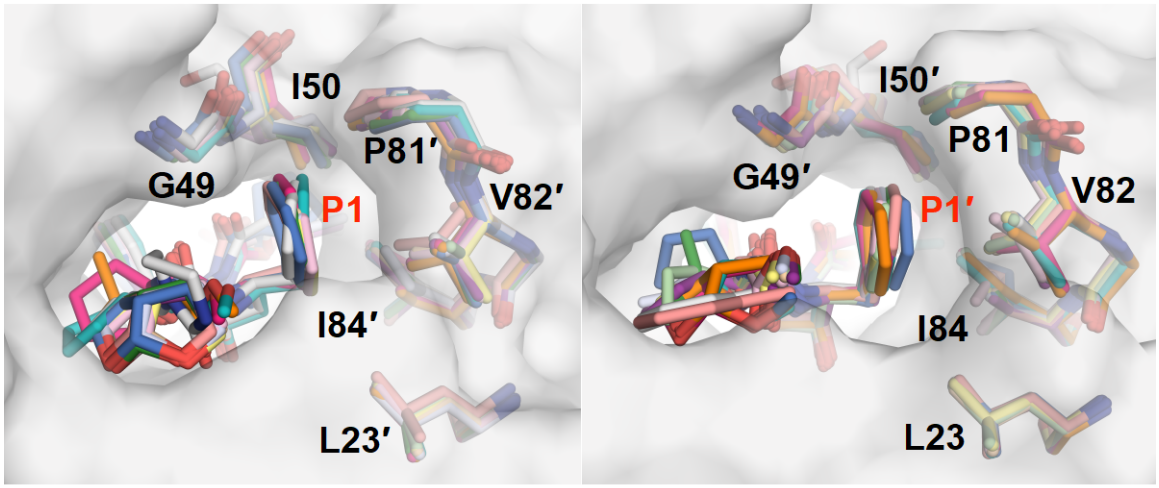
Table 5.3 Intermolecular hydrogen bonds between the protease and inhibitors in the cocrystal structures (distance in Å). Shorter distances are more red and longer distances are more blue (cutoff 3.5 Å).

Residue	30	29	29	29	29	27	8'	48	27	25	25	25'	25'	50	50	50'	50'	27'	48'	48'	48'	29'	29'	29'	27'	8	30'	30'	30'	30'	30'
Atom	MC-N	MC-N	MC-N	H2O	SC-O	MC-O	SC-N	MC-O	MC-O	SC-O	SC-O	SC-O	SC-O	MC-N	MC-N	MC-N	MC-N	SC-O	MC-O	H2O	MC-O	MC-N	H2O	SC-O	MC-O	SC-N	MC-O	H2O	SC-O	MC-N	MC-O
Type	Direct	Direct	Direct	Water	Water	Water	Water	Direct	Direct	Direct	Direct	Direct	Direct	Water	Water	Water	Water	Direct	Direct	Water	Water	Direct	Water	Water	Water	Water	Direct	Water	Water	Water	Water
1	2.9	3.2	2.8						3.0	3.2	2.5	2.9	2.5	3.0	3.1	2.8	2.9										3.1	2.9	2.9		
2			2.7	3.3	2.7	2.9	3.4		2.9	2.7	2.7	2.8	2.7	2.7	2.7	2.8	3.0	3.9													
3			3.0	3.1	2.8	3.0		3.1	3.0	3.6	2.9	2.8	2.5	2.7	3.0	2.8	3.0	3.2	3.0			3.0	3.2	2.8	2.8	2.9					
4	3.2	3.4	2.7						2.9	2.4	2.4	2.9	2.9	2.8	3.2	2.5	3.0	3.7													
12a	3.1	3.2	2.8						3.0	3.0	2.4	2.9	2.7	2.7	3.1	2.7	3.0	3.2	2.9	3.2	3.0	2.9	3.3	2.8	2.9	3.7					
12b	3.1	3.4	2.9						2.9	3.1	2.5	2.9	2.7	2.7	3.1	2.7	3.0	3.5				3.1						2.9		2.9	2.8
12c	3.2	3.4	2.8						2.9	3.0	2.6	2.7	2.6	2.8	3.0	2.7	3.1	3.1	2.9			3.0	3.4	2.7	3.0	3.5					
12d	3.4	3.5	2.8						2.7	2.8	2.6	2.7	2.7	2.9	3.1	2.8	2.8	3.2	2.9			3.0	3.5	2.6	3.0	3.5					
12e	3.2	3.3	2.8						2.8	2.7	2.6	2.7	2.6	2.8	3.1	2.8	2.9	3.2	2.8			3.0	3.5	2.5	3.0	3.3					
12f	3.3	3.5	2.8						2.7	2.8	2.5	2.6	2.6	2.8	3.2	2.8	2.8	3.1	2.9			3.0	3.4	2.6	3.0	3.5					
12g	3.4	3.5	2.8						2.7	2.8	2.6	2.7	2.6	2.8	3.4	2.7	2.9	3.1	2.9			3.0	3.5	2.5	3.1	3.6					
12h	3.1	3.2	2.8						3.0	2.9	2.6	2.8	2.5	2.8	2.9	2.8	3.0	3.4	2.9			2.9						2.9		2.9	3.2

Residue	30'	29'	29'	29'	29'	27'	8'	48'	27'	25'	25'	25	25	50'	50'	50	50	27	48	48	48	29	29	29	27	8	30	30	30	30	30
Atom	MC-N	MC-N	MC-N	H2O	SC-O	MC-O	SC-N	MC-O	MC-O	SC-O	SC-O	SC-O	SC-O	MC-N	MC-N	MC-N	MC-N	SC-O	MC-O	H2O	MC-O	MC-N	H2O	SC-O	MC-O	SC-N	MC-O	H2O	SC-O	MC-N	MC-O
Type	Direct	Direct	Direct	Water	Water	Water	Water	Direct	Direct	Direct	Direct	Direct	Direct	Water	Water	Water	Water	Direct	Direct	Water	Water	Direct	Water	Water	Water	Water	Direct	Water	Water	Water	Water
15a	3.2	3.4	2.9						2.9	3.1	2.7	2.7	2.6	2.8	3.1	2.8	3.0	3.0	2.9	3.0	2.9	3.0	3.4	2.8	2.9	3.5					
15b	3.1	3.3	2.9						3.1	3.0	2.8	2.7	2.5	2.9	3.0	2.6	3.2	3.4	3.0	3.0	3.0	3.3									
15c	3.4	3.6	2.7						3.2	2.7	2.7	2.6	2.7	3.0	2.9	2.7	3.1	2.6	3.1			3.0	3.5	2.8	3.0	3.6					

MC-N = main-chain nitrogen; MC-O = main-chain oxygen.

The Phe-Phe isostere in hybrid compounds, in addition to maintaining key hydrogen bonding interactions with the protease, makes a number of favorable vdW interactions with the hydrophobic residues that make up the S1/S1′ subsites. Despite the same phenylalanine side chains at the P1 and P1′ positions, there are a few notable differences in vdW contacts with the protease [**Figure 5.6**]. Compared to LPV, the P1/P1′ phenyl group adjacent to the amino acid moiety is shifted towards Val82 resulting in increased vdW contacts with that residue, particularly for compounds with larger P2/P2′ amino acid moieties. These localized gains in vdW contacts are slightly lower for proximal compounds (**15a**–**c**) than their distal counterparts (**12a**–**c**). The shift of the P1/P1′ phenyl group towards Val82, creating more vdW contacts, resulted in a nearly equal and opposite loss of contacts with Ile50 in the same subsite. The packing of the phenyl group adjacent to the *bis*-THF moiety resulted in a small increase in vdW contacts with the 80's loop (residues 81, 82 and 84) but decrease in contacts with flap residue Gly49. In the proximal compounds, gains in vdW contacts are smaller and the losses are greater. Overall, despite localized losses and gains in vdW interactions for both distal and proximal compounds, the packing of the P1 and P1′ groups remain generally similar to LPV.



		4	12a	12b	12c	12d	12e	12f	12g	12h			15a	15b	15c	
P1	B23	0.01	-0.11	0.18	-0.06	-0.21	-0.01	-0.28	-0.06	0.00		A23	-0.14	-0.03	-0.14	P1′
	A49	0.61	0.55	0.94	0.57	0.55	0.42	0.55	0.56	0.32		B49	1.01	1.22	0.81	
	A50	-0.16	0.19	-1.48	-0.85	0.16	0.08	-0.02	0.15	-0.41		B50	-0.57	-0.90	0.53	
	B81	-0.28	-0.14	-0.09	-0.31	-0.27	-0.17	-0.47	-0.25	-0.17		A81	-0.03	0.02	0.00	
	B82	-0.44	0.02	0.45	-0.65	-0.45	-0.33	-0.31	-0.51	-0.19		A82	0.02	-0.05	-0.32	
	B84	0.14	-0.05	-0.27	-0.27	0.29	0.12	-0.17	-0.02	-0.71		A84	-0.23	-0.18	-0.49	
P1′	A23	-0.17	-0.05	-0.27	-0.41	-0.11	-0.07	-0.06	-0.40	-0.18		B23	-0.08	0.04	-0.42	P1
	B49	0.23	0.10	-0.24	0.40	-0.33	0.33	-0.15	0.07	0.04		A49	-0.42	-0.57	-0.14	
	B50	1.25	1.54	1.27	1.16	1.01	1.05	1.10	1.06	1.20		A50	-0.16	-0.07	0.89	
	A81	0.09	0.07	0.07	-0.22	-0.01	0.00	0.06	-0.34	-0.02		B81	-0.38	-0.41	-0.06	
	A82	-0.80	-1.18	-1.26	-1.58	-1.33	-0.98	-1.16	-1.10	-1.45		B82	-0.98	-1.26	-1.26	
	A84	0.13	0.18	0.09	-0.08	0.52	0.52	0.11	0.75	0.12		B84	0.35	0.14	0.75	

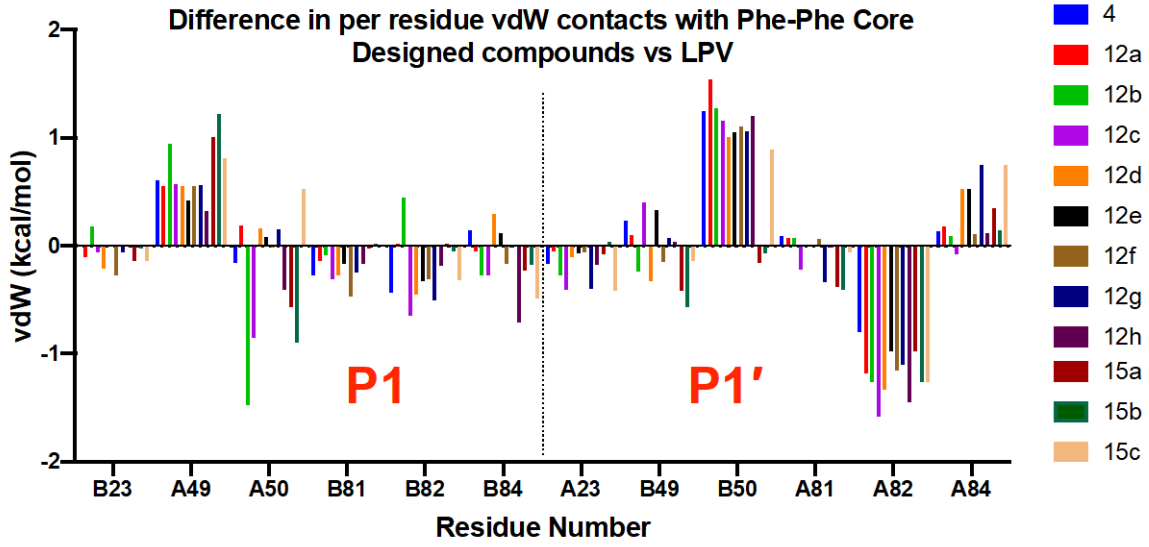


Figure 5.6 Comparison of per residue vdW contacts of the Phe-Phe core compared to LPV. Negative numbers indicate that the designed compounds are making better contacts with those residues.

The cocrystal structures of hybrid compounds (and **4**) also revealed similar binding interactions of the *bis*-THF moiety as in DRV-protease complex regardless of the position, with the two ring oxygen atoms positioned to form hydrogen bonds with the main-chain NH of Asp29 and Asp30. However, there is a subtle differences in the position and puckering of the *bis*-THF rings, causing the carbamate-linked THF ring oxygen atom to slightly shift away from the backbone NH of Asp30 compared to DRV (3.1–3.4 Å versus 2.9 Å) [**Figure 5.7**]. Whereas the oxygen atom of the other THF ring maintains a similar distance to the backbone NH of Asp29 (2.7–2.9 Å versus 2.8 Å) [**Table 5.3**]. For distal compounds, these subtle changes in the *bis*-THF moiety arise from the differences in the binding of core isosteres, where the P1 benzyl group and the carbamate are shifted towards the B-chain compared to their position in DRV. This is likely due to the additional carbon in the core or to accommodate the larger P1′ group. In the proximal compounds, the core scaffold from the central hydroxyl to the *bis*-THF moiety is identical. However, the opposite stereochemistry of the central hydroxyl causes a subtle shift preventing the *bis*-THF moiety from optimally interacting with Asp30′ [**Figure 5.7**]. The minor differences in the position and puckering of the *bis*-THF moiety in hybrid compounds result in weaker hydrogen bonding interactions with the protease in the S2/S2′ subsite compared to DRV.

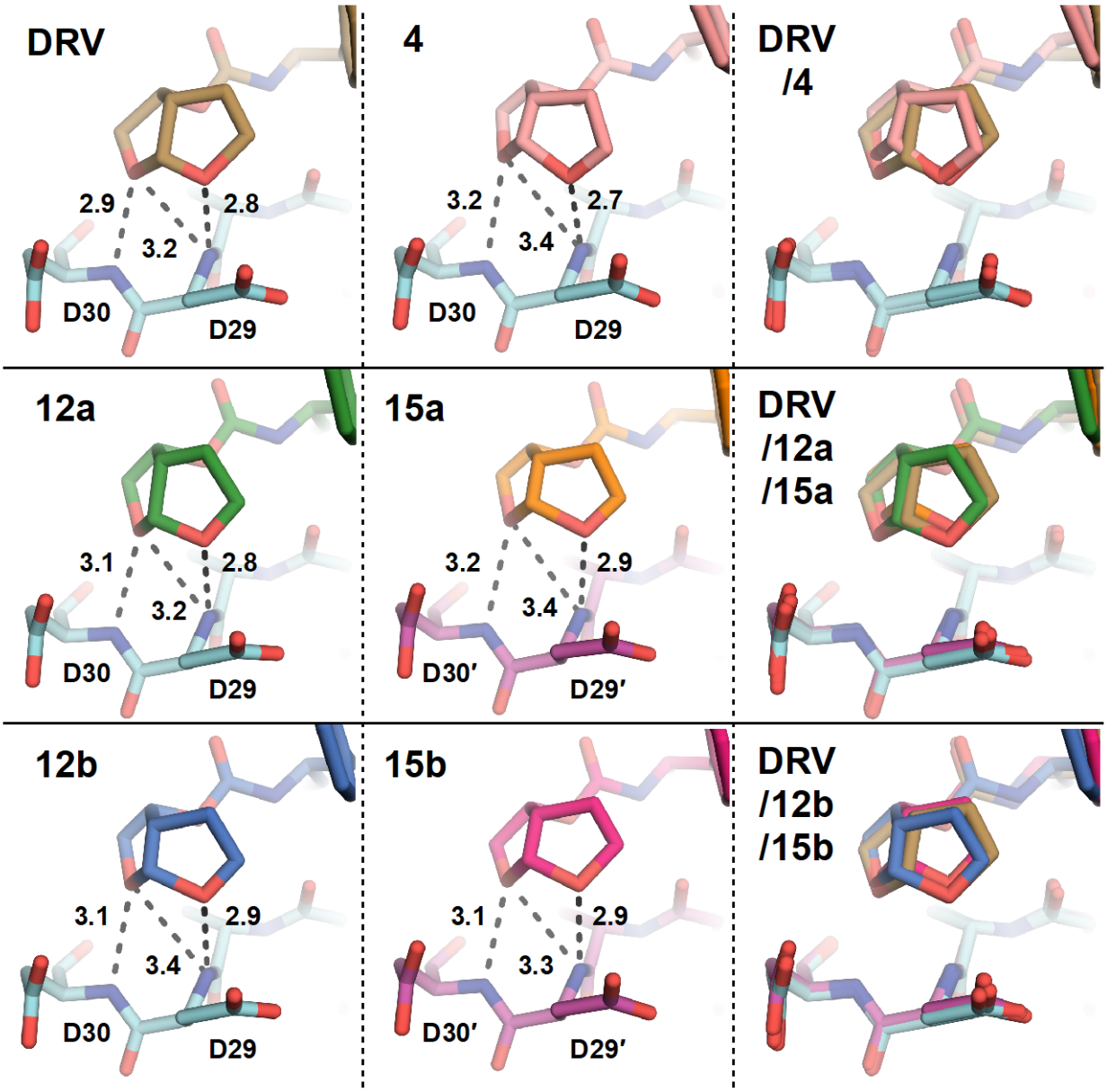
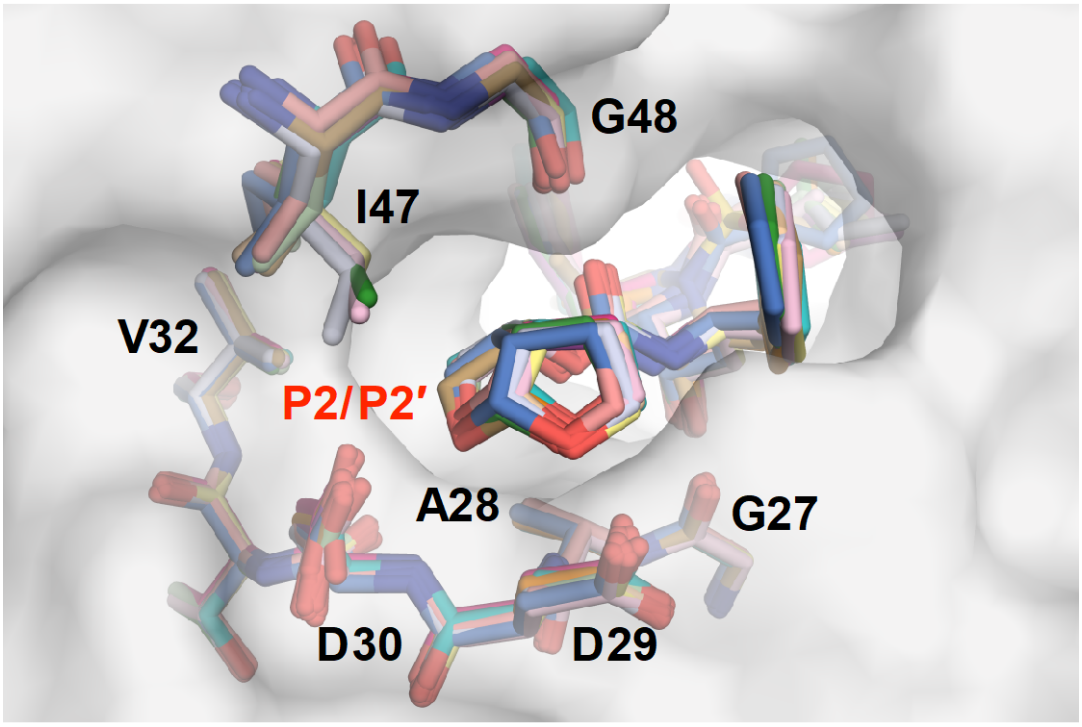


Figure 5.7 Comparison of the binding conformations of the *bis*-THF moiety and protease-inhibitor hydrogen bond distances with the main-chain NH of Asp29 and Asp30 in the S2/S2′ subsite of HIV-1 protease. As observed the DRV-protease complex, the *bis*-THF moiety retained similar interactions with the protease in complexes with compound **4** (top), the most potent hybrid compounds **12a** and **15a** (middle), and the least potent hybrid compounds **12b** and **15b** (bottom), irrespective of position on the Phe-Phe core.

The *bis*-THF moiety of DRV is critical to its potency, not only for the hydrogen bonding interactions with Asp29 and Asp30, but also vdW packing with residues in the S2 subsite, particularly Val32 and Ile47. In hybrid compounds, the minor change in the position and puckering of the *bis*-THF moiety causes reduced vdW interactions with Ala28 and Asp30, but increased contacts with Ile47 and Gly48 [**Figure 5.8**]. The main-chain carbonyl of Gly48 is oriented towards the hydrophobic region of *bis*-THF, so increased contacts with this residue may be unfavorable. In contrast, compounds **4**, **15a** and **15b** do not make these unfavorable contacts with Gly48. With the exception of enhanced interactions with Ile47, the *bis*-THF moiety in most hybrid compounds make less optimal vdW interactions with the protease compared to those of DRV.



	4	12a	12b	12c	12d	12e	12f	12g	12h		15a	15b	15c
A27	0.15	0.07	0.27	-0.13	-0.13	0.06	0.04	0.17	0.09	B27	-0.10	0.00	0.23
A28	0.18	0.25	0.16	0.14	0.15	0.20	0.17	0.21	0.18	B28	0.14	0.29	-0.06
A29	0.07	0.00	0.12	-0.05	0.04	0.01	0.06	-0.04	0.08	B29	0.23	0.38	0.19
A30	0.14	0.33	-0.27	-0.03	0.69	0.23	0.62	-0.09	0.81	B30	0.35	0.42	-0.18
A32	0.02	0.09	-0.02	-0.08	0.11	0.13	0.10	-0.08	0.09	B32	0.11	0.05	0.00
A47	-0.14	-1.19	-0.20	-0.28	-1.11	-0.06	-1.12	-0.34	-1.12	B47	0.09	0.20	-0.06
A48	-0.58	-0.71	-0.57	-0.84	-0.72	-0.63	-0.31	-0.58	-0.66	B48	0.20	0.42	-0.52

P2 P2′

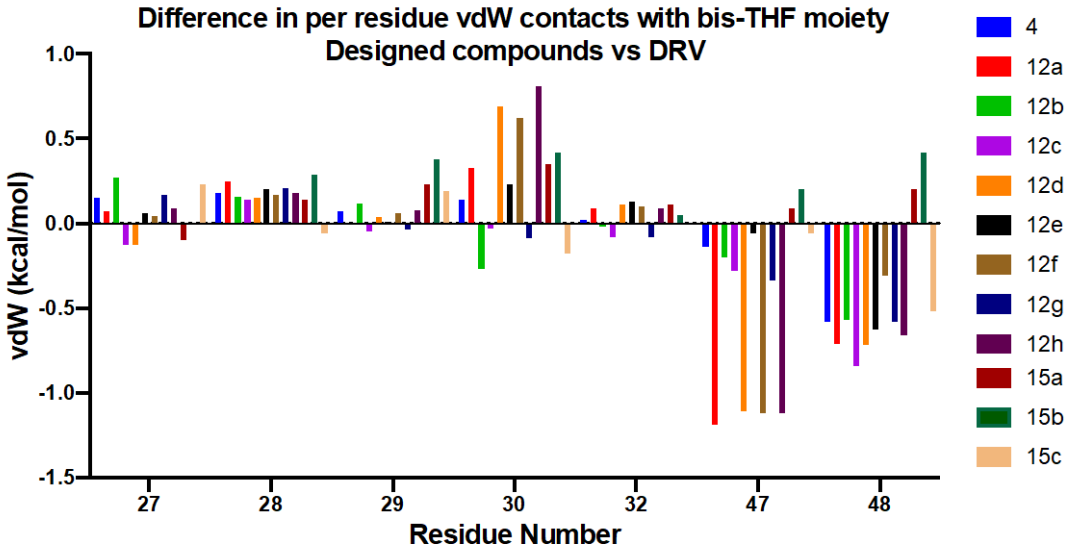


Figure 5.8 Comparison of per residue vdW contacts of the *bis*-THF moiety compared to DRV. Negative numbers indicate that the designed compounds are making better contacts with those residues.

The major difference between the hybrid compounds and the previously reported Phe-Phe core-based *bis*-THF containing PIs **4** and **5** is the introduction of an ATV-like amino acid moiety to target the opposite S2/S2′ subsite. The amino acid side chain interacts with hydrophobic protease residues, while the *N*-(methoxycarbonyl)-capping group allows for additional hydrogen bonding interactions with the protease. As already noted, LPV's P2′ moiety does not form any hydrogen bonds, and DRV's P2′ moiety makes one direct and one water-mediated hydrogen bond in the S2′ subsite [**Figure 5.2**]. In most hybrid compounds, the carbonyl oxygen of the carbamate makes one direct hydrogen bond with the backbone NH of Asp29′ (Asp29), and the carbamate NH makes another direct hydrogen bond with the backbone carbonyl of Gly48′ (Gly48) in the flaps [**Figure 5.9**]. There are also a number of coordinated water molecules that allow for additional water-mediated interactions. The carbonyl oxygen of the carbamate coordinates the same water-mediated hydrogen bonds as the cyclic urea moiety of LPV, with the backbone carbonyl of Gly27′ (Gly27) and the side-chain carboxylate of Asp29′ (Asp29) [**Figure 5.9**]. This network of polar interactions is maintained by all amino acid moieties in both the S2′ and S2 subsites except the cycloleucine in compound **12b**.

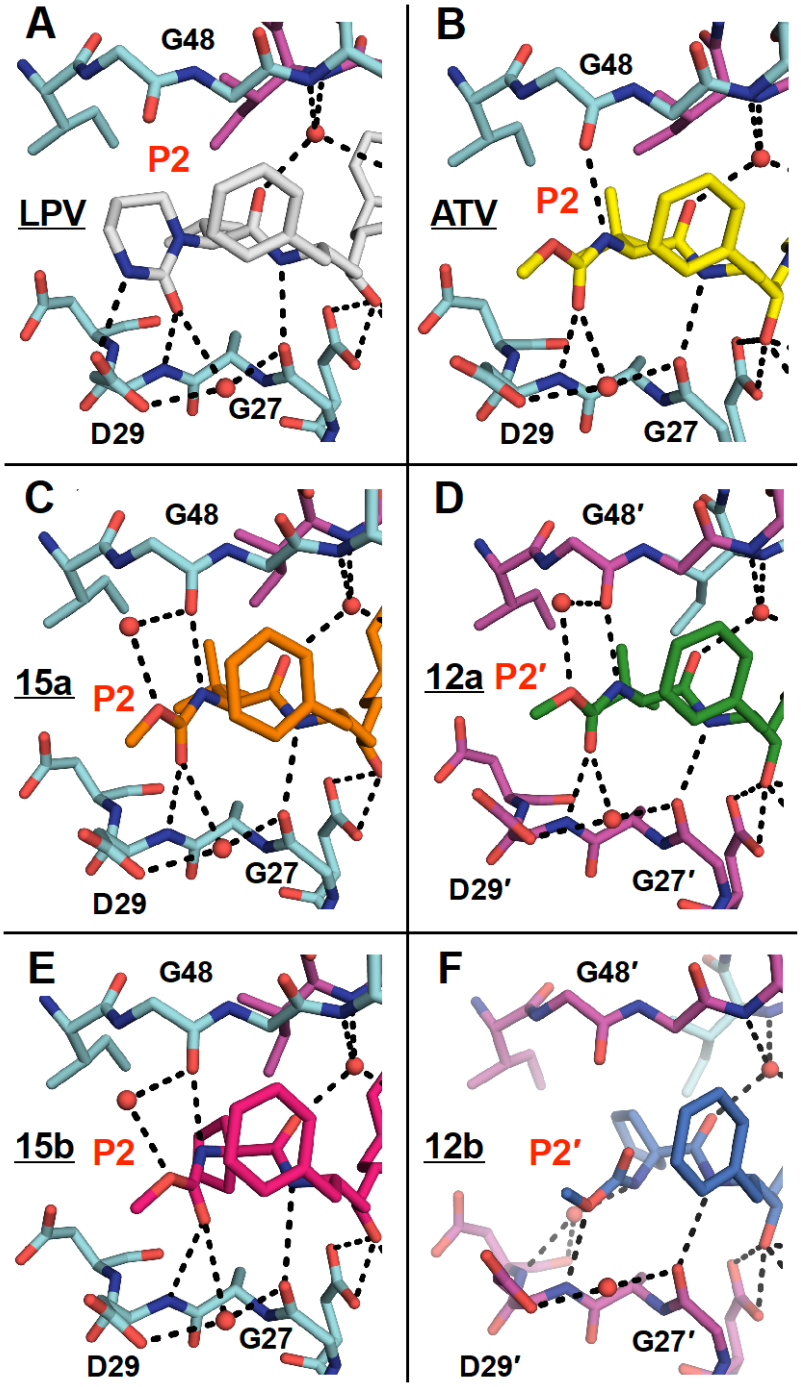
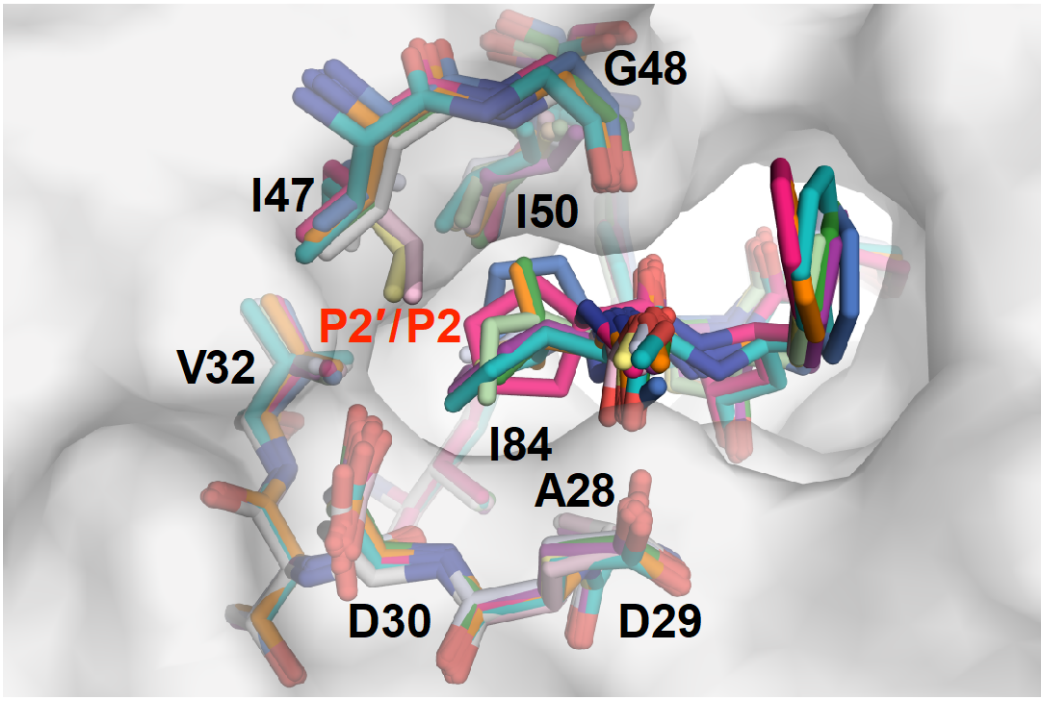


Figure 5.9 Comparison of protease-inhibitor direct and water-mediated hydrogen bond interactions in the S2/S2′ subsites for the cyclic urea moiety of LPV (A), *N*-(methoxycarbonyl)-*tert*-leucine of ATV (B), and the *N*-(methoxycarbonyl)-capped amino acid moieties of hybrid compound 15a (C), 12a (D), 15b (E), and 12b (F). Hydrogen bonds are shown as black dashes (cutoff 3.5 Å) and waters are shown as red spheres.

The carbamate of the distal compound **12b** is flipped relative to all other inhibitors losing the direct hydrogen bond to flap residue Gly48′ [**Figure 5.9F**]. This flip positions the carbonyl oxygen of the carbamate close to the carbonyl oxygen of Gly48′ (3.3 Å), creating a very unfavorable interaction and preventing compound **12b** from utilizing the same water-mediated hydrogen bonds as LPV and most hybrid compounds. Instead **12b** forms a hydrogen bond with a “backside” water in the S2′ subsite, creating a water-mediated interaction with the main-chain carbonyl and NH of Asp30′ [**Figure 5.9F**]. The marked changes in the hydrogen bonding interactions in the S2/S2′ subsite likely underlie the greatly reduced potency of **12b** compared to all other hybrid compounds. Unlike **12b**, the P2 carbamate group in compound **15b** maintained a similar conformation as in the *tert*-leucine analogue **15a** but with a slight change in the position of the carbonyl oxygen away from Asp29. This shift resulted in increased hydrogen bond distance between the P2 carbamate carbonyl and the main-chain NH of Asp29 compared to **15a** (3.3 Å versus 2.9 Å). However, the reduced potency of **15b** could not result from minor differences in hydrogen bonding interactions alone, underlying the significance of non-polar interactions for potency.

The protease-inhibitor vdW interactions in the S2/S2′ subsite varied significantly depending on the size and shape of the hydrophobic amino acid moiety [**Figure 5.10**]. Compounds **12a** and **15a** with the *tert*-leucine make highly distributed vdW contacts with a number of residues. In contrast, compared to *tert*-leucine, the cycloleucine side chain in **12b** predominantly interacts with Ile50 and I84V, while losing significant contacts with residues 28–30 [**Figure 5.11**]. The cycloleucine side chain of **15b** packs in the same hydrophobic area but makes less vdW contacts than **12b**. Compound **12c** with the

cyclopentyl glycine also has increased vdW contacts with Ile50 but without experiencing losses at other resides as observed for **12b**. In general, the cyclic hydrophobic amino acid derivatives make more localized vdW interactions in the S2/S2′ subsite while the acyclic moieties make more distributed contacts. Thus, in addition to altered polar interactions, the differences in vdW contacts due to the size and shape of the hydrophobic amino acid moiety are likely responsible for the varied inhibitor potencies.



	12b	12c	12d	12e	12f	12g	12h		15b	15c
A50	-1.67	-1.04	-0.04	-0.11	-0.21	-0.05	-0.60	B50	-0.33	1.10
B28	1.45	0.02	1.04	0.14	0.11	-0.13	0.14	A28	0.86	-0.11
B29	1.39	0.45	1.12	0.37	0.40	0.10	0.53	A29	0.61	0.22
B30	1.50	-0.29	1.63	0.27	0.19	-0.48	0.20	A30	0.55	-0.15
B32	-0.10	0.15	1.01	0.55	0.36	0.18	-0.22	A32	0.13	0.40
B47	0.76	0.00	0.15	0.76	0.12	0.20	0.13	A47	0.39	-0.09
B48	0.97	0.26	0.34	0.34	0.45	0.05	0.21	A48	0.10	0.53
B84	-0.22	-0.23	0.33	0.16	-0.12	0.03	-0.67	A84	0.05	-0.26

P2' P2

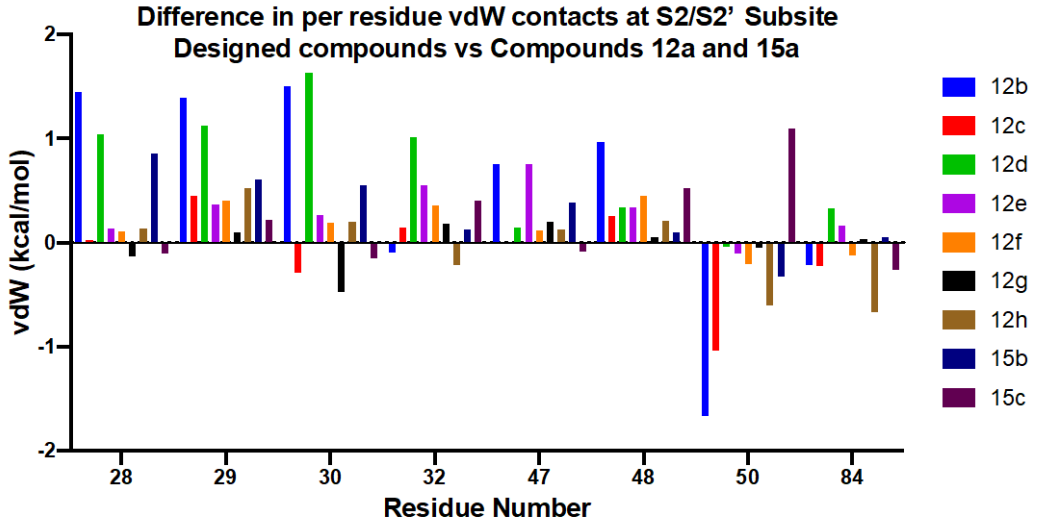


Figure 5.10 Comparison of per residue vdW contacts in the S2/S2' subsite between hybrid compounds and the best compounds (12a/15a). Negative numbers indicate that the compounds are making better contacts with those residues compared to 12a/15a.

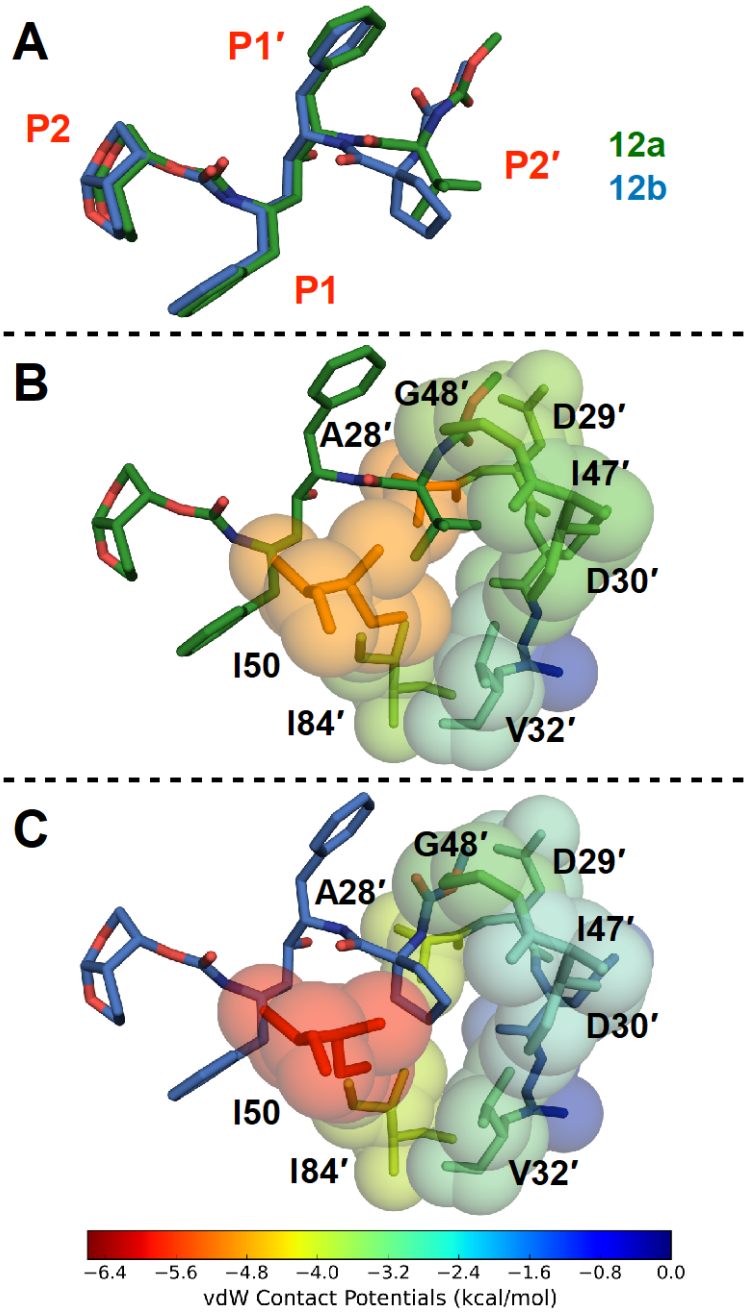


Figure 5.11 Comparison of binding conformation (A) and van der Waals packing with residues in the S2′ subsite for the most and least potent inhibitors 12a (B, green sticks) and 12b (C, blue sticks). Compound 12a makes more extensive contacts with the protease as estimated by relative vdW contact energy compared to 12b (–91 versus –86 kcal/mol), with major difference occurring at the S2′ subsite (–27 versus –23 kcal/mol). Protease residues are colored red to blue for highest to lowest vdW interactions with the inhibitor.

5.4.3 Molecular Dynamics (MD) Simulations

Starting from the cocrystal structures, molecular dynamics (MD) simulations were utilized to interrogate the stability of protease-inhibitor interactions in the dynamic ensemble of the complexes. Three replicates of fully hydrated 100 ns MD simulations were performed on each protease-inhibitor complex. All simulations reached convergence [**Figure 5.12**].

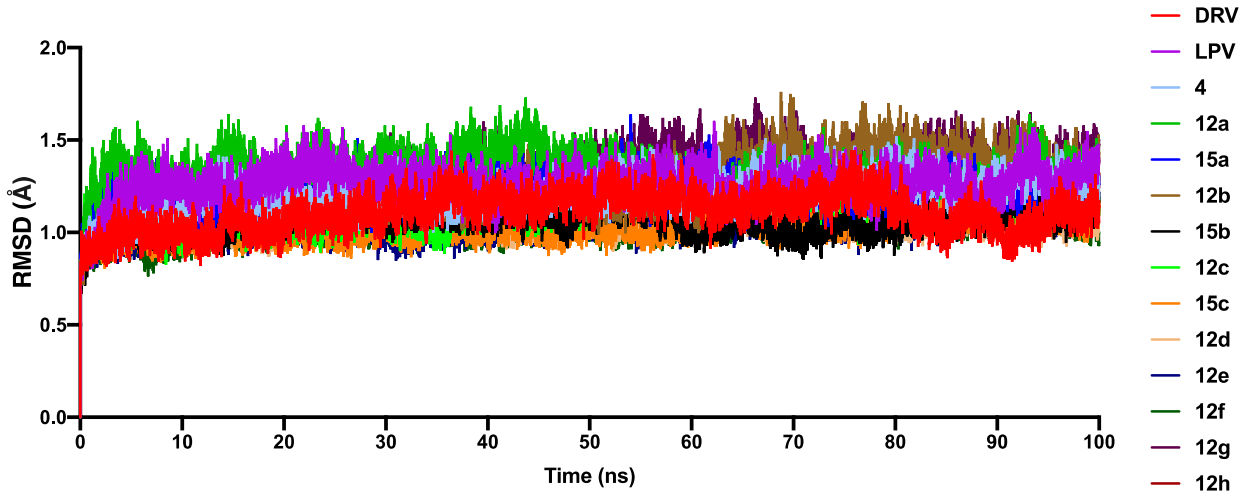


Figure 5.12 Root-mean-square deviation (RMSD) analysis of the molecular dynamics simulations revealed the systems reached convergence.

The protease active site and flaps dynamics were analyzed by measuring C_α distances between specific residues. The analysis focused on the most and least potent of both the distal and proximal compounds (**12a**, **15a**, **12b** and **15b**) compared to LPV. The cocrystal structures suggested the flaps and the 80's loops surrounding the inhibitor may be perturbed when bound to different compounds. To monitor dynamics of the flaps, distance distributions between C_α atoms of Ile50–Ile84′, Ile50′–Ile84 and Ile50–Ile50′ were calculated during the MD trajectories. Distances for Ile84–Ile84′ and Pro81–Pro81′ were used to probe the expansion or narrowing of the "lower" and "upper" side wall of the active site, respectively [**Figure 5.13**]. In protease complexes with hybrid compounds, the active site remained largely unchanged in the "lower" part of the active site as indicated by a single, narrow distribution of Ile84–Ile84′ distance centered around 15.2 Å. However, compared to LPV the shift in the Pro81–Pro81′ distance distribution with compounds **12a**, **15a**, **12b**, and **15b** indicated the active site was narrower in the "upper" portion close to the flaps. In addition, the active site was asymmetrically longer (Ile50′–Ile84 or Ile50–Ile84′), sampling bimodal distance distributions, suggesting the sampling of a semi-open conformation of the flaps. This was supported by the Ile50–Ile50′ distance distribution, with one major peak around 5.3 Å (closed flaps) and another ~10 Å corresponding to semi-open flaps with compounds **15a** and **15b** [**Figure 5.13**]. Increased distance between residues Ile50 and Ile50′ in the relatively short time (nanoseconds) during MD simulations has been previously suggested to indicate flap opening.[211-212] Thus, perturbed flaps dynamics and narrowing of the active site may indicate unfavorable binding of these compounds compared to LPV, which is supported by the *K*i data.

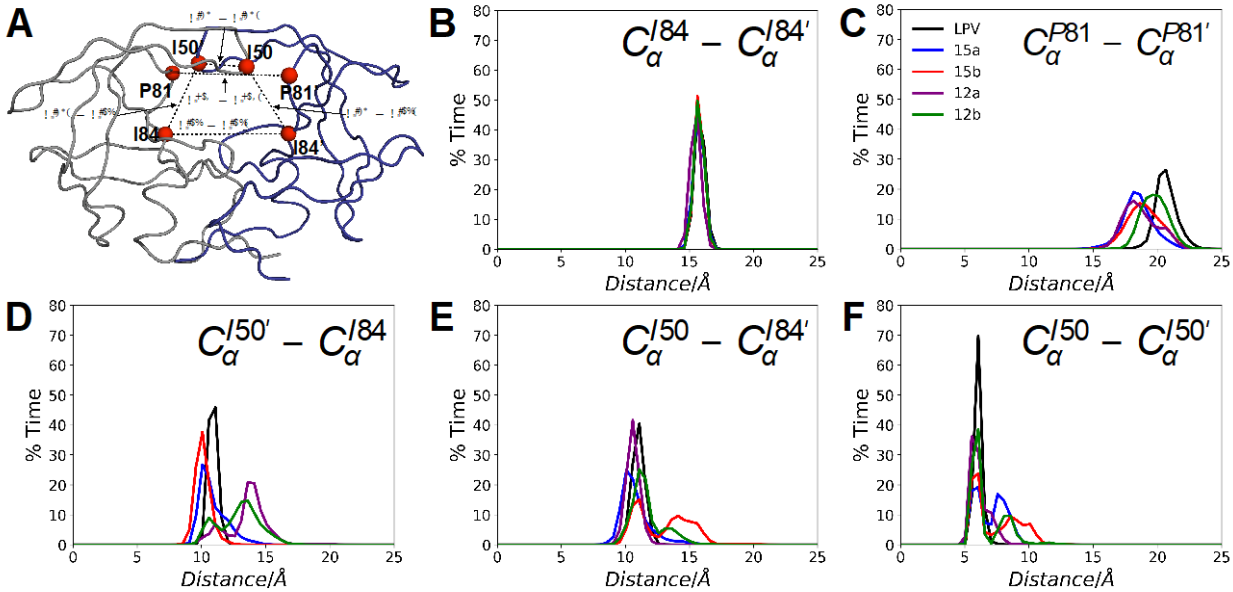


Figure 5.13 Distance distribution between (A) C_α atoms of active site residues which can be used to estimate the height, width, and diagonal distance distribution of the active site over a molecular dynamics simulation. Distance distributions of (B) Ile84–Ile84′ and (C) Pro81–Pro81′ represent the width of the bottom- and top-half of the active site, respectively. Distance distributions of (D) Ile50–Ile84′ and (E) Ile50′–Ile84 aid in visualizing diagonal motion of the flaps. Finally, the distance distribution of (F) Ile50–Ile50′ aids in visualizing the pulling apart of the flaps.

Next, the stability of inhibitor binding was analyzed by calculating root-mean-square fluctuations (RMSF) of each atom [**Figure 5.14**]. In all complexes, the *bis*-THF moiety often displayed the lowest RMSF amongst all moieties, regardless of its position on the Phe-Phe core. Notably, even in the least active compounds with nanomolar potency (**12b** and **15b**), the *bis*-THF moiety showed fluctuations similar to that observed in **12a** and **15a**. Moreover, despite the pseudo-symmetric nature of the Phe-Phe core, the phenylalanine side chain adjacent to the *bis*-THF moiety displayed lower RMSF compared to the distal phenylalanine side chain. The P2/P2′ amino acid moieties showed varying RMSF profiles, independent of the size or shape of the hydrophobic side chain. The phenylalanine adjacent to the amino acid moiety consistently displayed greater RMSF, with the exception of **12e**. The varying fluctuations of the P1/P1′ moiety depending on the size and shape of the P2/P2′ moiety indicate interdependence between the S1/S1′ and S2/S2′ subsites. The P2/P2′ amino acid moiety also affected the fluctuations of the *bis*-THF moiety but to a much lesser extent. This data suggests that in hybrid compounds the *bis*-THF moiety and the adjacent P1/P1′ moiety remain relatively stable in the protease active site regardless of the position on the Phe-Phe core. In contrast, the fluctuations of the amino acid and the adjacent P1/P1′ moieties appear to destabilize overall inhibitor binding. Thus, due to interdependence between the S1/S1′ and S2/S2′ subsites, the optimization of the *bis*-THF containing hybrid PIs, would require modification of the amino acid moiety together with the adjacent P1/P1′ group.

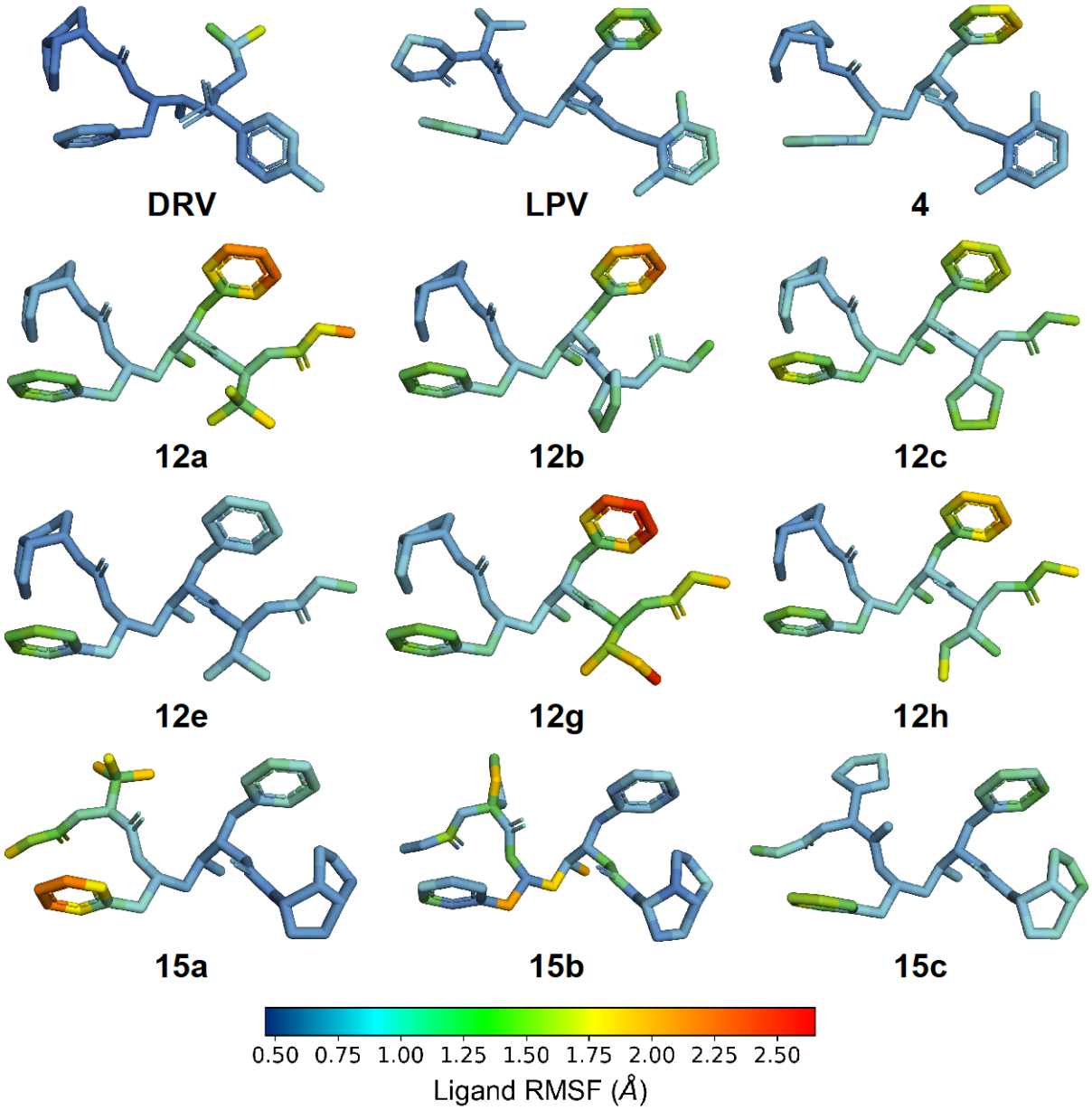


Figure 5.14 Inhibitor root-mean-square fluctuation (RMSF) of each non-hydrogen atom of DRV, LPV, and hybrid compounds mapped onto their crystal structure. Warmer colors indicate larger fluctuations.

The relationship between inhibitor and protease conformational dynamics was examined via simultaneous monitoring of inhibitor bond rotations and protease residue distances. The dihedral angles of all inhibitor rotatable bonds were calculated throughout the MD simulations. Given the bimodal sampling of the flap distances (Ile50–Ile50′, Ile50–Ile84′ and Ile50′–Ile84), we examined whether the semi-open conformation of flaps was associated with certain inhibitor bond rotations. In protease complexes with compounds **12a** and **15a**, the two most potent inhibitors, the separation of the flap tips (Ile50–Ile50′) was associated with the conformational sampling of the dihedral controlling the amino acid moiety [**Figure 5.15**]. In this semi-open flap conformation, the Ile50/Ile50′ residues at the top of the flaps also lost intra-protease vdW contacts with residues 32, 47–49 and 84 [**Figure 5.16**]. Loss of intra-protease vdW interactions could result in a less stable protease-inhibitor complex, leading to an overall decrease in inhibitor binding affinity.

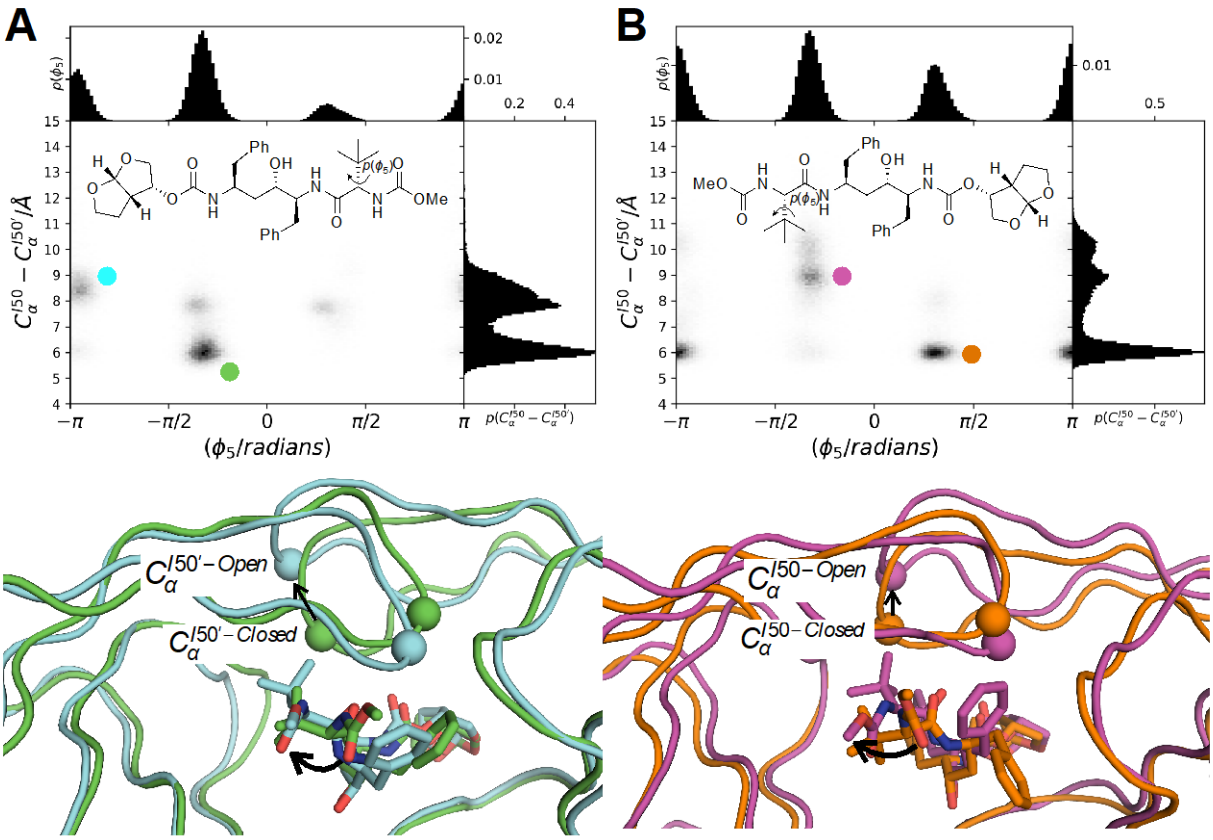


Figure 5.15 The dihedral angle (φ_5) responsible for conformational sampling of the *tert*-leucine moiety on (A) 12a and (B) 15a is associated with increased Ile50–Ile50′ distance sampling.

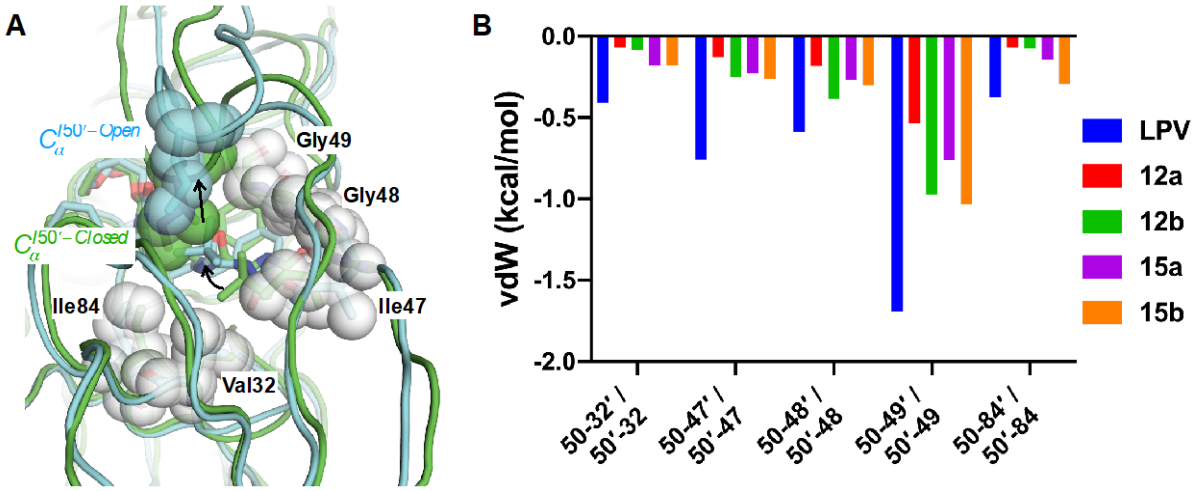


Figure 5.16 Intra-protease van der Waals interactions between Ile50/Ile50′ and neighboring residues.

In addition, protein dynamics were compared by calculating the root-mean-square fluctuations (RMSF) of C_α atoms. Consistent with our observations from the distance distributions, there were significant increases in fluctuations at the flaps and 80's loop [**Figure 5.17**]. In complex with LPV, the flaps and active site show little fluctuations, comparable to DRV. However, the proximal hybrid compounds showed increased fluctuations at both flaps, with fluctuations asymmetrically greater at the B chain flap, contacting the amino acid moiety. The increased protein fluctuation could impact the stability of protease-inhibitor hydrogen bonding interactions.

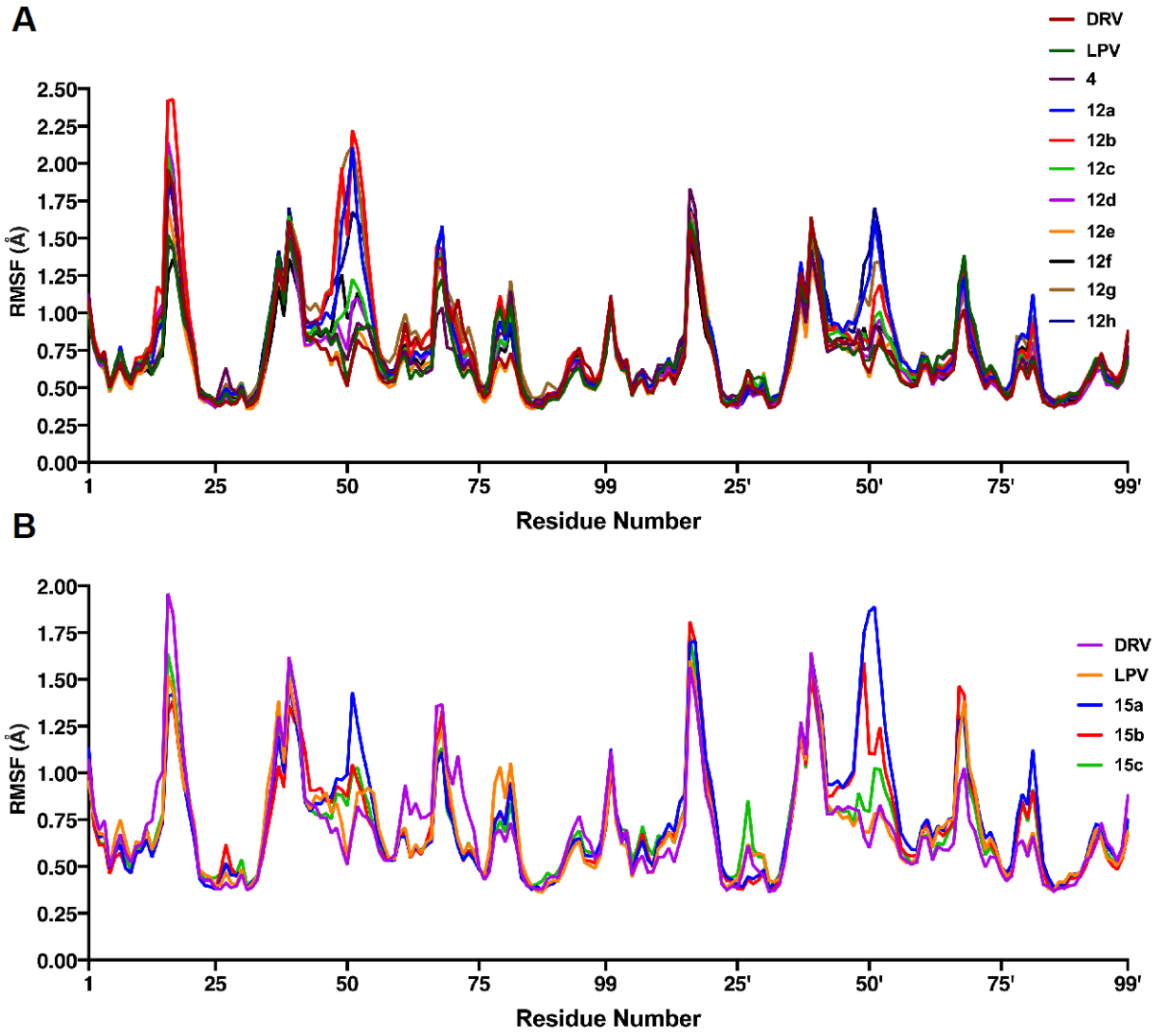


Figure 5.17 Protein root-mean-square fluctuation (RMSF) of the (A) distal and (B) proximal compounds. DRV and LPV are included for comparison.

In the cocrystal structures, the hybrid compounds make a number of direct and water-mediated hydrogen bonds with the protease active site, including the central hydroxyl with catalytic residues, coordinated water with Ile50/50′, the Phe-Phe core nitrogen atoms with Gly27, *bis*-THF with Asp29 and Asp30, carbamate with Asp29 and Gly48, and the coordinated S3′ water with Gly27′ and Asp29′. To assess if the increased protein and ligand fluctuations impact hydrogen bonding stability, the patterns and frequencies of these hydrogen bonds were monitored throughout the MD simulations [**Figure 5.18-5.19**]. In MD simulations, LPV and the hybrid compounds did not maintain all hydrogen bonds observed in the cocrystal structures. LPV and all distal compounds (**4** and **12a**–**h**) maintained a hydrogen bond with Asp25′ (66–99% frequency) but did not form a hydrogen bond with Asp25. Whereas proximal compounds **15a** and **15b** formed the hydrogen bonds with Asp25 and Asp25′ at roughly 100% and 50% frequency, respectively. In contrast, the hydrogen bond between proximal compound **15c** and Asp25 was unstable, while the hydrogen bond with Asp25′ was highly stable (92% frequency). Hybrid compounds maintained the water-mediated hydrogen bonding network with the main-chain NH of Ile50 and Ile50′ at varying stability. The interactions with Ile50 were observed to be less stable compared to the ones with Ile50′ (21–65% versus 49–78%). For most of the hybrid compounds, the hydrogen bonds between the Phe-Phe core NH groups and the main-chain carbonyls of Gly27/Gly27′ were either not observed or were very unstable.

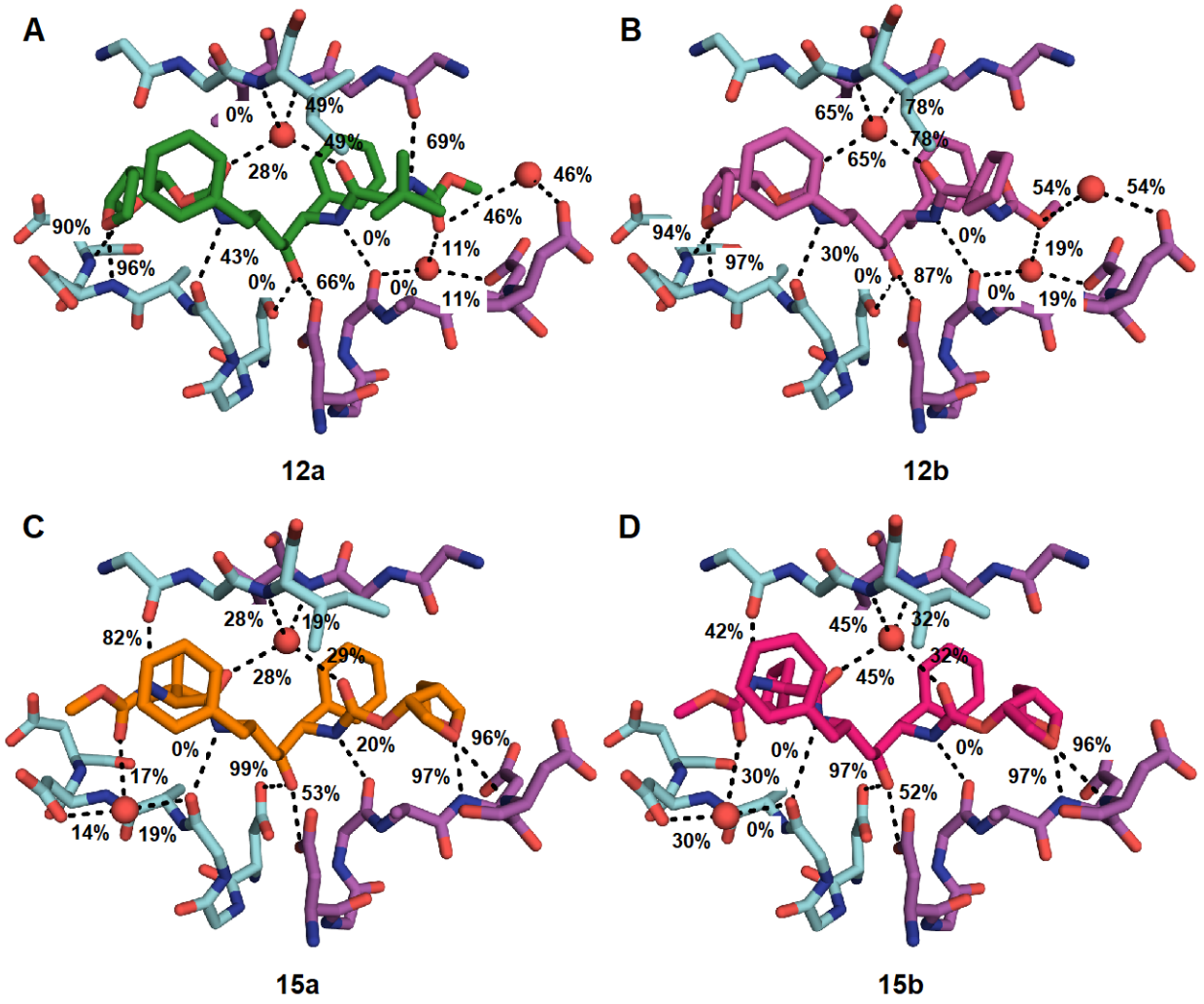


Figure 5.18 Protease-inhibitor hydrogen bond frequencies from MD simulations mapped onto the cocrystal structures of compounds (A) **12a**, (B), **12b**, (C), **15a**, and (D) **15b**. The protease side chains and inhibitors are shown as sticks; chains A and B are colored cyan and magenta, respectively.

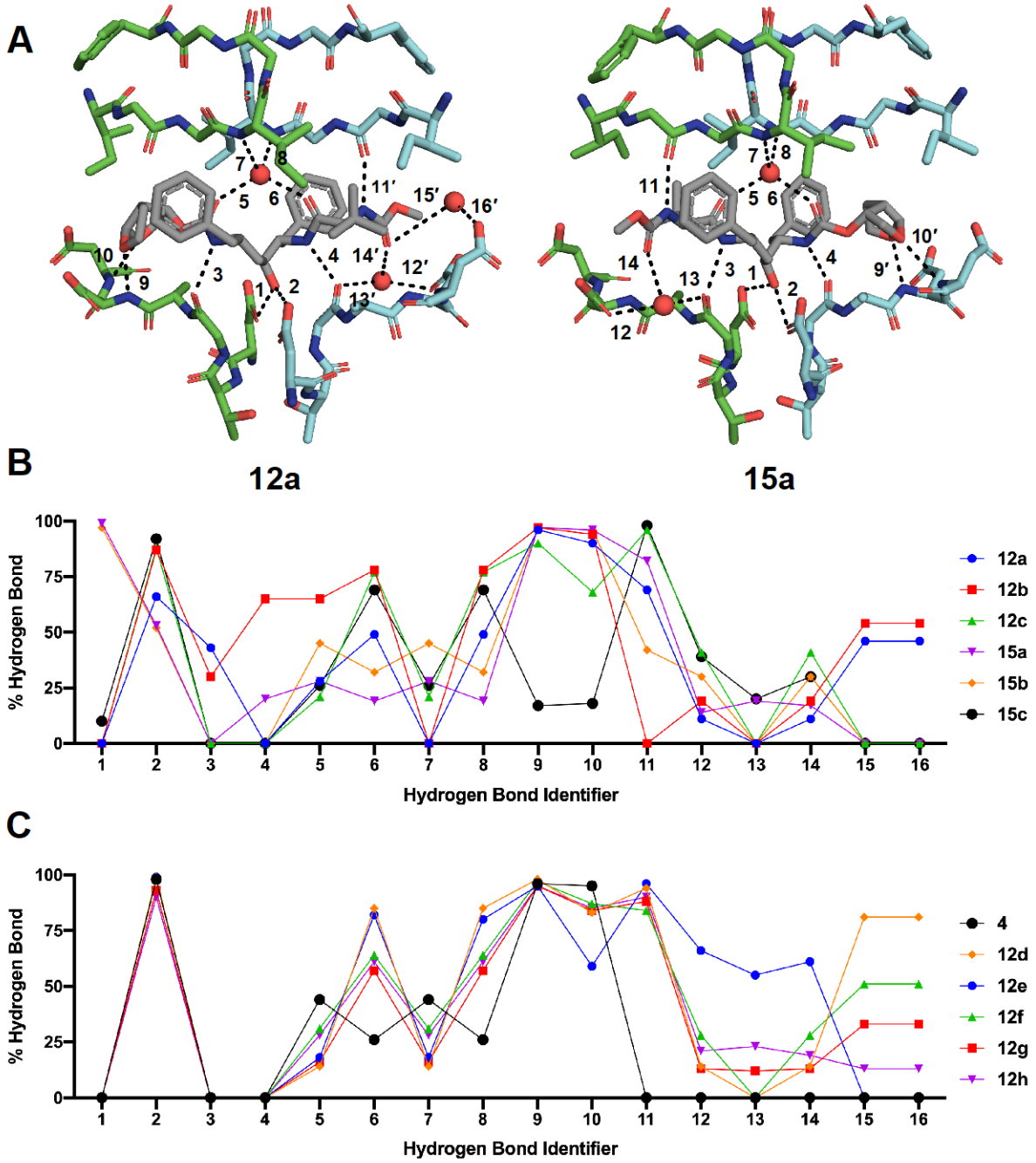


Figure 5.19 (A) Protease-inhibitor hydrogen bonds are indicated as black dashed lines on the cocrystal structures of **12a** and **15a**. Hydrogen bonding frequencies from MD simulations for (B) pairs of distal and proximal compounds and (C) the remaining hybrid compounds.

The *bis*-THF moiety in all hybrid compounds maintained two hydrogen bonds with the main-chain NH of Asp29/Asp29′ and Asp30/Asp30′ at high frequency, regardless of the position on the Phe-Phe isostere [**Figure 5.18-5.19**]. In contrast, the P2/P2′ amino acid moiety had varying effects on hydrogen bonding interactions in the S2/S2′ subsite. The carbamate nitrogen of both distal and proximal compounds formed a hydrogen bond with the backbone carbonyl of Gly48′ at moderate to high frequency, but this interaction was not observed for the least potent compound **12b** and was less stable for **15b**. The carbonyl oxygen of the P2′ amino acid moiety in distal compounds formed relatively stable water-mediated interactions with the sidechain of Asp30′. In contrast, the carbonyl oxygen of the P2 moiety in proximal compounds **15a** and **15c** formed three low-frequency (14–39%) water-mediated hydrogen bonds, bridging the main-chain carbonyl of Gly27 and the side-chain carboxylate of Asp29. These data suggest that for most hybrid compounds the key hydrogen bonding interactions of the *bis*-THF moiety and the Phe-Phe isostere with protease remain relatively stable. In addition, the size and shape of the P2/P2′ amino acid moiety not only influences the interactions in the S2/S2′ subsites but can also propagate changes in the water-mediate interactions between the inhibitors and Ile50/Ile50′.

5.5 Conclusions

We have explored hybrid HIV-1 PIs containing the *bis*-THF moiety of DRV on either side of the Phe-Phe dipeptide isostere of LPV to ascertain whether this moiety could maintain key interactions with the protease and improve potency against wild-type and primary drug-resistant variants of the enzyme. Most of the hybrid compounds retained picomolar biochemical potency irrespective of the position of *bis*-THF relative to the central hydroxyl group, likely due to the pseudo-C_2-symmetric nature of the Phe-Phe

isostere. But in both distal and proximal compound series the identity of the other P2/P2′ moiety significantly affected potency, favoring relatively flexible, hydrophobic moieties. The cocrystal structures of hybrid compounds revealed that the *bis*-THF moiety maintains the key hydrogen bonding interactions with the protease in the S2/S2′ subsite as observed in the DRV-protease complex regardless of its position on the Phe-Phe isostere. In contrast, the binding of the hydrophobic amino acid moiety in the other S2/S2′ subsite was greatly influenced by the size, shape, and flexibility of the hydrophobic group, which affected hydrogen bonding and vdW interactions with the protease. Moreover, this moiety appears to influence binding of the adjacent P1/P1′ group as well as the water-mediated interactions between the inhibitor and flap residues, indicating dynamic interdependence between the protease subsites. This is further supported by the relatively stable binding interactions of the *bis*-THF moiety and the adjacent P1/P1′ group. The dynamic interdependence between subsites in the protease active site can be exploited to optimize inhibitor potency against drug-resistant protease variants. The detailed structural characterization of hybrid HIV-1 PIs containing *bis*-THF in non-sulfonamide dipeptide isosteres offers opportunities for structure-guided optimization of these promising inhibitors.

5.6 Experimental Section

5.6.1 General

All reactions were performed in oven-dried round-bottom flasks fitted with rubber septa under argon atmosphere unless otherwise noted. All reagents and solvents, including anhydrous solvents, were purchased from commercial sources and used as received. Flash column chromatography was performed on an automated Teledyne ISCO

CombiFlash Rf+ system equipped with a UV-vis detector using disposable Redisep Gold high performance silica gel columns or was performed manually using silica gel (230−400 mesh, EMD Millipore). Thin- layer chromatography (TLC) was performed using silica gel (60 F_{254}) coated aluminum plates (EMD Millipore), and spots were visualized by exposure to ultraviolet light (UV), exposure to iodine adsorbed on silica gel, and/or staining with alcohol solutions of phosphomolybdic acid (PMA) and ninhydrin followed by brief heating. ^{1}H NMR and ^{13}C NMR spectra were acquired on Varian Mercury 400 MHz and Bruker Avance III HD 500 MHz NMR instruments. Chemical shifts are reported in ppm (δ scale) with the residual solvent signal used as a reference and coupling constant (*J*) values are reported in hertz (Hz). Data are presented as follows: chemical shift, multiplicity (s = singlet, d = doublet, dd = doublet of doublet, dd = doublet of triplet, t = triplet, m = multiplet, br s = broad singlet), coupling constant in Hz, and integration. High-resolution mass spectra (HRMS) were recorded on a Thermo Scientific Orbitrap Velos Pro mass spectrometer coupled with a Thermo Scientific Accela 1250 UPLC and an autosampler using electrospray ionization (ESI) in the positive mode. The purity of final compounds was determined by analytical HPLC and was found to be ≥95% pure. HPLC was performed on an Agilent 1200 system equipped with a multiple wavelength detector and a manual injector under the following conditions: column, Phenomenex Hypersil-BDS-5u-C18 (5 µm, 4.6 mm × 250 mm, 130 Å); solvent A, H_2O containing 0.1% trifluoroacetic acid (TFA); solvent B, CH_3CN containing 0.1% TFA; gradient, 20% B to 100% B over 15 min followed by 100% B over 5 min; injection volume, 20 µL; flow rate, 1 mL/min. The wavelengths of detection were 254 nm and 280 nm. Retention times and purity data for each target compound and intermediate can be found online.

5.6.2 Protease Gene Construction

Protease gene construction was carried out as previously described.[183-184] The NL4-3 strain has four naturally occurring polymorphisms in the protease relative to the SF2 strain. In short, the protease variant genes (I84V, I50V/A71V) were constructed using QuikChange site-directed mutagenesis (Genewiz) onto NL4-3 wild-type protease on a pET11a plasmid with codon optimization for protein expression in *Escherichia coli*. A Q7K mutation was included to prevent autoproteolysis.[185]

5.6.3 Protein Expression and Purification

The expression, isolation, and purification of WT and mutant HIV-1 proteases used for the kinetic assays and crystallization were carried out as previously described.[183-184] Briefly, the gene encoding the HIV protease was subcloned into the heat-inducible pXC35 expression vector (ATCC) and transformed into E. coli TAP-106 cells. Cells grown in 6 L of Terrific Broth were lysed with a cell disruptor and the protein was purified from inclusion bodies.[186] The inclusion body centrifugation pellet was dissolved in 50% acetic acid followed by another round of centrifugation to remove impurities. Size exclusion chromatography was used to separate high molecular weight proteins from the desired protease. This was carried out on a 2.1 L Sephadex G-75 superfine column (Millipore Sigma) equilibrated with 50% acetic acid. The cleanest fractions of HIV protease were refolded into a 10-fold dilution of 0.05 M sodium acetate at pH 5.5, 5% ethylene glycol, 10% glycerol, and 5 mM DTT. Folded protein was concentrated down to 1–2 mg/mL and stored. This stored protease was used in *K*i assays. For crystallography, a final purification was performed with a Pharmacia Superdex 75 FPLC column equilibrated with 0.05 M sodium acetate at pH 5.5, 5% ethylene glycol, 10% glycerol, and 5 mM DTT.

Protease fractions purified from the size exclusion column were concentrated to 1–2 mg/mL using an Amicon Ultra-15 10-kDa device (Millipore) for crystallization.

5.6.4 Enzyme Inhibition Assays

The enzyme inhibition assays were carried out as previously described.[90, 213] To determine the enzyme inhibition constant (*K*i), in a 96-well plate, each inhibitor was serially diluted, including a no drug control, and incubated with 0.35 nM protein for 1 hour. A 10-amino acid substrate containing an optimized protease cleavage site with an EDANS/DABCYL FRET pair was dissolved in 4% DMSO at 120 μM. Using the Envision plate reader, 5 μL of the 120 μM substrate was added to the 96-well plate to a final concentration of 10 μM. The fluorescence was observed with an excitation at 340 nm and emission at 492 nm and monitored for 200 counts, for approximately 60 min. Data was analyzed with Prism7, as described in Chapter 2. DRV was used as a control in all assays.

5.6.5 Antiviral Assays

293T and TZM-BL[181] cells (NIH AIDS Research and Reference Reagent Program) were maintained in Dulbecco's modified Eagle's medium supplemented with 10% fetal calf serum in the presence of penicillin and streptomycin at 37 °C with 5% CO_2. To determine the concentration of drugs achieving 50% inhibition of infection compared with the drug-free control, 4.5×10^6 293T cells were seeded onto a 10-cm plate 24 h before transfection. Cells were transfected with 8 μg of either the wild-type plasmid, infectious molecular clone pNL-CH derived from the pNL4-3 clone of HIV-1 using FuGENE 6 transfection reagent (Roche). The culture supernatant of 293T cells transfected with wild-type or PI-resistant HIV-1 variant was removed 18 h after transfection and the cells were washed with 1 × PBS. The 293T cells were collected and transferred to wells of a 24-well

plate. Briefly, each drug was serially diluted in the culture medium and the dilutions were added to the wells of a 24-well plate. The 293T cells (0.5×10^6 per well) collected from the transfection were added to wells containing various concentrations of drug. The culture supernatant containing virus particles was harvested 18 h after the 293T cells were reseeded in the presence of drug. This supernatant was filtered through a 0.45-μm-pore-size membrane (Millipore) to remove cell debris then used to infect 2×10^4 TZM-BL cells in a 96-well plate following a procedure previously described.[182] The culture supernatant was removed from each well 48 h post-infection, and the cells were washed with 1 × PBS. For the luciferase assay, infected TZM-BL cells were lysed in 1× reporter lysis buffer (Promega) and the cells were kept at −80 °C. After one freeze-thaw cycle, the cell lysates were transferred into a 96-well assay plate (Costar), and luciferase activity was measured using a luminometer (Promega). The culture supernatant harvested from 293T cells reseeded in the absence of drugs was used as a drug-free control. EC_{50} was determined based on a dose-response curve generated using GraphPad Prism (version 7.0).

5.6.6 Protein Crystallization

The condition reliably producing cocrystals of NL4-3 WT protease bound to PIs was discovered and optimized as previously described.[188, 213] Briefly, all cocrystals were grown at room temperature by hanging drop vapor diffusion method in a 24-well VDX hanging-drop trays (Hampton Research) with a protease concentration of 1.4–1.7 mg/mL with 3-fold molar excess of inhibitors and mixed with the precipitant solution at a 1:2 ratio. The reservoir solution was 23–27% (w/v) ammonium sulfate with 0.1 M *bis*-Tris-methane buffer at pH 5.5, and the crystallization drops were set with 2 μL of well solution and 1 μL

of protein-inhibitor solution and micro-seeded with a cat whisker. Diffraction quality crystals were obtained within 1 week. As data were collected at 100 K, cryogenic conditions contained the precipitant solution supplemented with 25% glycerol.

5.6.7 X-Ray Data Collection and Structure Solution

X-ray diffraction data were collected and solved as previously described.[184, 188, 213] Diffraction quality crystals were flash frozen under a cryostream when mounting the crystals either at our in-house Rigaku_Saturn944 X-ray system or the Chicago APS Synchrotron Beamline 23-1D-D. The cocrystal diffraction intensities from the Rigaku system were indexed, integrated, and scaled using HKL3000.[214] Structures were solved using molecular replacement with PHASER.[106] Model building and refinement were performed using Coot[107] and Phenix.[108] Ligands were designed in Maestro and the output sdf files were used in the Phenix program eLBOW[109] to generate cif files containing atomic positions and constraints necessary for ligand refinement. Iterative rounds of crystallographic refinement were carried out until convergence was achieved. To limit bias throughout the refinement process, five percent of the data were reserved for the free R-value calculation.[110] MolProbity[111] was applied to evaluate the final structures before deposition in the PDB. Structure analysis, superposition and figure generation was done using PyMOL.[215] X-ray data collection and crystallographic refinement statistics are presented in the Supporting Information [**Table 5.2**].

The cocrystal structures of all new hybrid compounds were solved in the $P2_12_12_1$ space group with one protease homodimer in the asymmetric unit and only one orientation of the bound inhibitor in the active site, which was crucial for direct comparison of inhibitor structures. The cocrystal structure of compound **4** was solved in the $P2_1$ space

group with two protease homodimers in the asymmetric unit, and one inhibitor bound to each dimer in one orientation, allowing one $P2_1$ dimer to be directly compared to the crystal structures solved in the $P2_12_12_1$ space group. Three structures (**10a**, **10b**, **13c**) had significant electron density at the flap tips (residues 50–51) indicating that the backbone atoms interacted in two conformations, as both hydrogen bond donors and acceptors. Figure generation and structural analysis calculations (distance differences, vdW, hydrogen bonds) were performed with the flap tips modeled in one conformation while the other conformation was excluded. Compound **5** formed cocrystals but were inadequate for X-ray data collection.

5.6.8 Intermolecular vdW Contact Analysis of Crystal Structures

To calculate the intermolecular vdW interaction energies the crystal structures were prepared using the Schrödinger Protein Preparation Wizard.[216] Hydrogen atoms were added, protonation states were determined, and the structures were minimized. The protease active site was monoprotonated at Asp25. Subsequently, force field parameters were assigned using the OPLS3 force field.[217] Interaction energies between the inhibitor and protease were estimated using a simplified Lennard-Jones potential, as previously described in detail.[218] Briefly, the vdW energy was calculated for pairwise interactions depending on the types of atoms interacting and the distance between them. For each protease residue, the change in vdW interactions relative to a reference complex in the same space group was also calculated for each variant structure.

5.6.9 Molecular Dynamics System Preparation

High resolution crystal structures of LPV and the designed inhibitors bound to WT protease were prepared using the Protein Preparation Wizard using Maestro within the

Schrödinger Suite[216] as previously described.[219-220] Briefly, missing atoms were added using Prime[221] and PROPKA[222] was used to determine the protonation state of the side chains at pH 7.0. The catalytic aspartic acid with a pKa higher than 7.0 was protonated whereas the one with a pKa less than 7.0 was unprotonated. Co-crystallized fragments such as phosphate ions were removed. Lastly, the structure was minimized to a convergence criterion of 0.3 Å using Impref.[223]

5.6.10 Molecular Dynamics Simulations

The prepared systems were placed in a cubic TIP3P implicit water box measuring 12Å on each side. Molecular dynamics simulations were carried out as previously described[219] using Desmond within Schrödinger Suite.[216] Briefly, chloride ions were used to neutralize the system and 0.15 M salt were added using sodium and chloride ions. The OPLS3 force field was used to parameterize the ligand and protein. Prior to starting the 100 ns MD simulations, the solvated system was minimized using the stepwise procedure described previously.[219] Triplicates of 100 ns simulations for LPV and the designed inhibitors in complex with WT protease each with a randomized velocity were started using a protocol previously developed.[219-220] The root-mean-square deviation (RMSD) and root-mean-square fluctuation (RMSF) were calculated using tools within the Schrodinger Suite.

5.7 Acknowledgements

This research used resources of the Advanced Photon Source, a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under Contract No. DE-AC02-06CH11357. GM/CA@APS has been funded in whole or in part with Federal funds from the National

Cancer Institute (ACB-12002) and the National Institute of General Medical Sciences (AGM-12006). The Eiger 16M detector was funded by an NIH–Office of Research Infrastructure Programs, High-End Instrumentation Grant (1S10OD012289-01A1). We thank the beamline specialists at 23-ID-D for their help in data collection. We thank Ellen A. Nalivaika for providing the HIV-1 protease variants, the AIDS Research and Reference Reagent Program (NIAD, NIH) for reference protease inhibitors, and Drs. William Royer, Brian Kelch and members of the Schiffer and Miller laboratories for helpful discussions. This work was supported by a grant from the National Institute of General Medical Sciences of the NIH (P01 GM109767).

CHAPTER 6:

Inhibiting HTLV-1 Protease: A Viable Target for Potent Antivirals

6.1 Abstract

Human T-cell lymphotropic virus type 1 (HTLV-1) is a retrovirus that can cause severe paralytic neurologic disease and immune disorders as well as cancer. An estimated 20 million people worldwide are infected with HTLV-1, with prevalence reaching 30% in some parts of the world. In stark contrast to HIV-1, no direct acting antivirals (DAAs) exist against HTLV-1. The aspartyl protease of HTLV-1 is a dimer similar to that of HIV-1 and processes the viral polyprotein to permit viral maturation. We report that the FDA-approved HIV-1 protease inhibitor darunavir (DRV) inhibits the enzyme with 0.8 μM potency and provides a scaffold for drug design against HTLV-1. Analogs of DRV that we designed and synthesized achieved sub-micromolar inhibition against HTLV-1 protease and inhibited Gag processing in viral maturation assays. Co-crystal structures of these inhibitors with HTLV-1 protease highlight opportunities for future inhibitor design. Our results show promise toward developing highly potent HTLV-1 protease inhibitors.

6.2 Introduction

Human T-cell lymphotropic viruses type 1 and type 2 (HTLV-1/2) were the first human retroviruses described nearly 40 years ago.[56-57] For decades HTLV-1 has been known to be highly carcinogenic and cause severe paralytic neurologic disease as well as immune disorders that can increase susceptibility to bacterial infections. HTLV-1 infection can lead to adult T-cell leukemia/lymphoma (ATL), HTLV-associated myelopathy/tropical spastic paraparesis (HAM/TSP), uveitis and infective dermatitis.[224-225] HTLV-1 is transmitted via the same routes as HIV-1 (sexually, via blood or mother-to-child) with significant HIV/HTLV co-infections reported in Europe, America, and Africa.[58] About 20 million people worldwide are infected with HTLV-1,[59] reaching endemic rates of 30% in some parts of the world. According to a recent report, up to 45% of the adult population among five Aboriginal communities in central Australia tested positive for HTLV-1.[226] With global human movement, HTLV-1 infections are starting to expand into previously non-endemic regions as recently reported for Spain from Latin American immigrants.[227] Unfortunately, there are no vaccines or direct acting antivirals (DAAs) against HTLV-1.

In the absence of preventive vaccines and DAAs, treatment options for HTLV-1 infected patients are scarce. For the estimated 3–5% of the HTLV-1 infected patients who develop ATL, chemotherapy is ineffective with very poor survival rates and relapse leading to death.[60-61] Earlier studies with HIV-1 protease inhibitor ritonavir, although not very potent against HTLV-1, were promising against HTLV-1 induced ATL in cell lines and mouse models.[228] In the clinic, a combination of zidovudine, an HIV-1 reverse transcriptase inhibitor, and IFN-α significantly improved prognosis, indicating active

HTLV-1 replication in ATL patients that can be targeted by antiretrovirals. As with HIV-1 infections, development of specific DAAs has a great potential to improve treatment outcomes. Such DAAs may also prove to be effective in all HTLV-1 associated human diseases, for pre-exposure prophylaxis (PrEP) especially in areas where the virus is endemic, and for preventing mother-to-child transmissions.[229]

HTLV-1 utilizes a similar viral machinery as HIV-1, and inhibiting the viral protease should prevent viral maturation. HIV-1 protease inhibitors are a success story of rational structure-based drug design, with the most recent FDA-approved drug, darunavir (DRV), having exceptional low-picomolar potency against wild type (WT) HIV-1 protease.[78] As peptidomimetic transition-state analogues, protease inhibitors are the most potent among HIV-1 antivirals, as enzymes bind strongest to their transition-state as compared to substrates or products.[230-231] Due to this potency and relatively high barrier to resistance, DRV has been tried as a monotherapy in patients with treatment-naïve HIV where the viral load and replication are relatively low.[232] Success against HIV-1 protease suggests similar potency can be achieved against HTLV-1 protease.

HTLV-1 protease is a 28 kDa homodimeric aspartyl protease that is 28% identical to HIV-1 protease with 45% identity between active site residues (**Figure 6.1)**,[62-63] yet HTLV-1 protease has considerably distinct substrate specificity (**Table 6.1)**.[63-67] The crystal structure of HTLV-1 protease was determined in 2005[65] and early attempts at inhibitor design involved peptide scanning with non-hydrolysable substrates.[233] Iterations of this process incorporated unnatural amino acids[234-235] and different capping moieties[236-237] leading to the identification of allophenylnorstatin-based HTLV-1 protease inhibitors with low nanomolar potency.[67] This inhibitor did not progress to clinic, likely due

to high peptidic character and large molecular weight. A few peptidic inhibitors were cocrystallized[238] and deposited into the Protein Data Bank (PDB)[69-70] providing valuable insights into inhibitor binding at the HTLV-1 active site. HIV-1 protease inhibitor indinavir (IDV), an FDA-approved drug, was also reported to weakly inhibit HTLV-1.[68] More recently, pyrrolidine-based C2-symmetric non-peptidic inhibitors were investigated and achieved low nanomolar potency.[239] Despite these efforts, there is still a need for exploring chemical scaffolds that will ultimately result in a clinically relevant HTLV-1 protease inhibitor with sub-nanomolar potency.

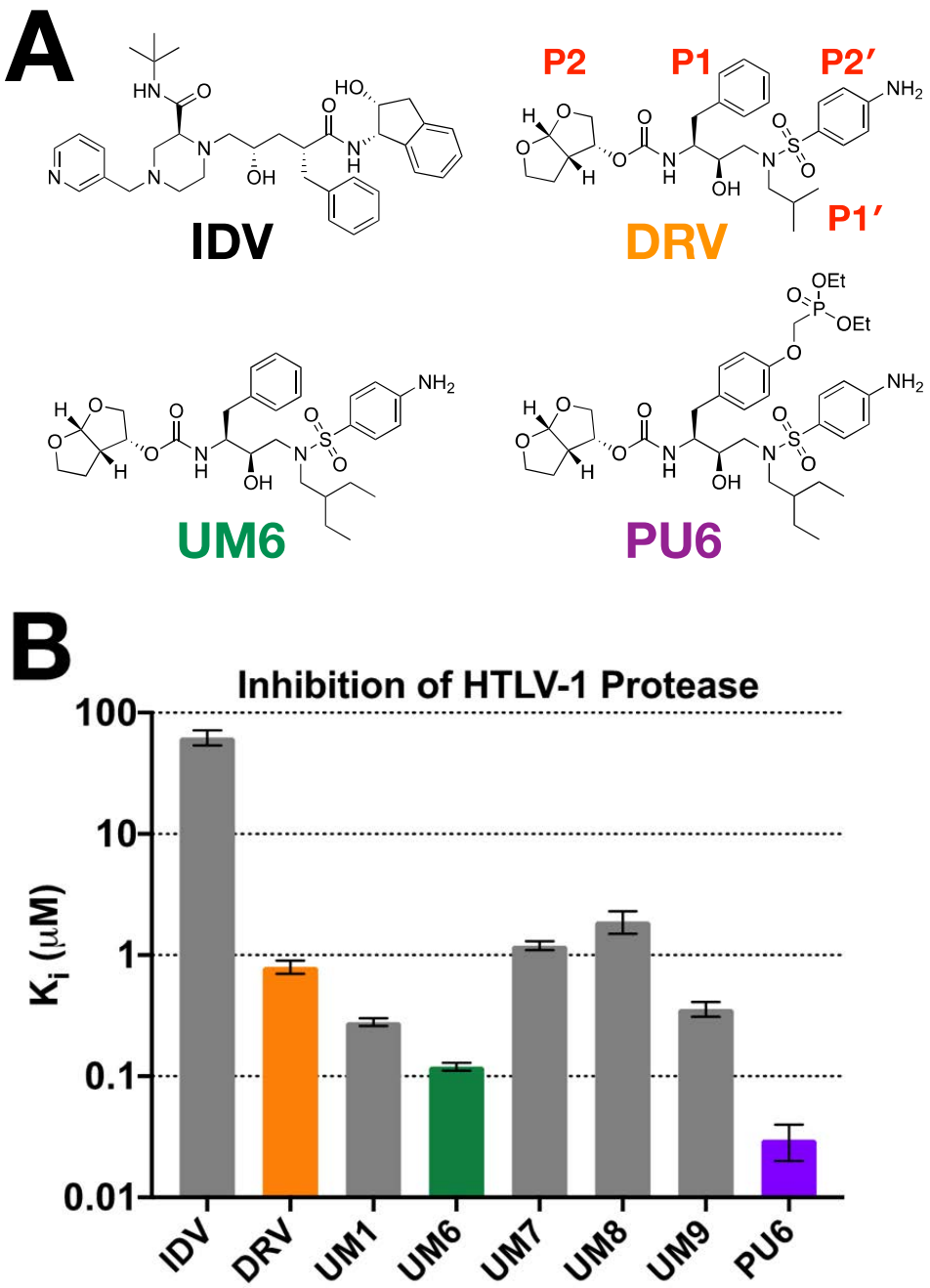


Figure 6.1 Inhibitors tested against HTLV-1 protease. (A) Chemical structure of indinavir (IDV), darunavir (DRV) with P2–P2′ moieties labeled, and DRV analogs that exhibited improved potency. (B) Enzyme inhibition constants (K_i) for IDV, DRV, and DRV analogs against HTLV-1 protease.

Table 6.1 HIV-1 protease and HTLV-1 protease substrates/cleavage sites.

Cleavage Site	HIV-1 (P4-P1/P1′-P4′)	HTLV-1 (P4-P1/P1′-P4′)
MA/CA	SQNY/PIVQ	PQVL/PVMH
CA/NC	-	TKVL/VVQP
Gag/PR	SFNF/PQIT	ASIL/PVIP
PR/Pol	TLNF/PISP	PVIL/PIQA
Pro/RT	-	PAVL/GLEL
RT-RH/IN	RKIL/FLDG	VLQL/SPAD

In this work, we tested the FDA-approved HIV-1 protease inhibitor DRV and our own novel DRV analogues against HTLV-1 protease. Several modifications to the DRV scaffold improved the inhibitors to low nanomolar potency, similar to the most potent HTLV-1 protease inhibitors to date. These compounds successfully inhibited maturation of viral-like particles (VLPs) by blocking the processing of viral Gag. Through the determination of co-crystal structures and comparison with analogous complexes in HIV-1 protease we have additional insights into how to further increase potency and potentially develop the first direct acting antiviral against HTLV-1.

6.3 Results and Discussion

6.3.1 DRV and Analogs Inhibit HTLV-1 Protease

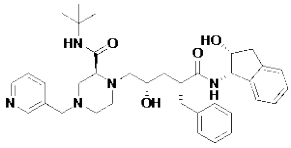
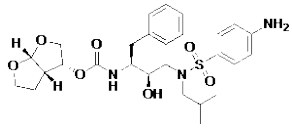
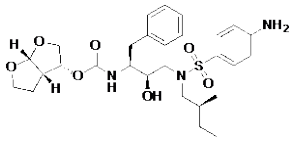
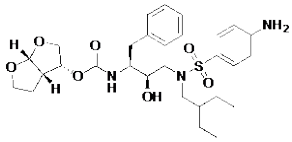
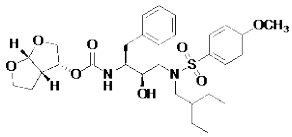
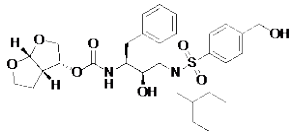
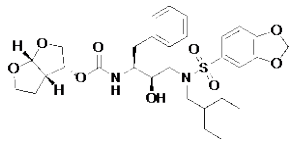
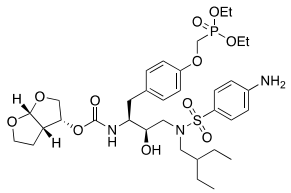
We expressed and purified the 116-amino-acid construct of HTLV-1 protease for enzyme inhibition assays and crystallization. This protease construct has 9 amino acids removed from the C-terminus to facilitate crystallization,[65] which does not affect catalytic activity.[240] Utilizing a FRET-based enzymatic assay, the activity of HTLV-1 protease was tested using two peptide substrates based on HTLV-1 matrix/capsid (MA/CA) and capsid/nucleocapsid (CA/NC) cleavage sites (**Table 6.2**). The Michaelis-Menten constant (K_M) was measured as 101.3 ± 1.9 μM and 31.6 ± 5.9 μM, consistent with previously reported values.[241] HTLV-1 catalyzed cleavage of the MA/CA substrate an order of magnitude faster than that of CA/NC, with a catalytic efficiency of 0.21± 0.02 $\mu M^{-1}s^{-1}$. Thus, the purified HTLV-1 protease is folded and active in our conditions.

Table 6.2 Enzymatic activity of HTLV-1 protease measured using natural substrate sequences.

Cleavage Site	Sequence (P4-P1/P1′-P4’)	K_M (μM)	k_{cat} (s^{-1})	k_{cat} / K_M ($\mu M^{-1}s^{-1}$)
MA/CA	PQVL/PVMH	101.3 ± 1.9	21.6 ± 0.1	0.21 ± 0.02
CA/NC	TKVL/VVQP	31.6 ± 5.9	1.9 ± 0.1	0.06 ± 0.06

The FDA-approved HIV-1 protease inhibitors IDV and DRV were tested against HTLV-1 protease (**Figure 6.1** and **Table 6.3)**. IDV weakly inhibited HTLV-1 protease (K_i = 62.7 μM), while DRV was two orders of magnitude more potent with a K_i of 0.8 μM. We previously found that modifications at the P1′ and P2′ positions of DRV improved potency against resistant variants of HIV-1 protease.[45] These DRV analogues with larger hydrophobic P1′ moieties, increasing from an isobutyl to *sec*-butyl and isohexyl, were also tested against HTLV-1 protease. Some of the analogs had significantly improved potency, the best being UM6 with a K_i of 0.12 ± 0.01 μM, which is ~7-fold lower compared to DRV. In both UM1 and UM6, increasing the hydrophobicity at P1′ improved potency. However, modifications at the P2′ position did not further increase potency, indicating the aniline moiety of DRV is better than phenoxymethyl, hydroxymethylbenzene, or benzo(1,3)dioxole.

Table 6.3 Inhibition constant (K_i) against HTLV-1 protease.

Code	Inhibitor	2D Structure	K_i (μM)
IDV	Indinavir		63 ± 9
DRV	**Darunavir**		**0.8 ± 0.1**
UM1	UMass1		0.28 ± 0.02
UM6	**UMass6**		**0.12 ± 0.01**
UM7	UMass7		1.2 ± 0.1
UM8	UMass8		1.9 ± 0.4
UM9	UMass9		0.4 ± 0.1
PU6	**P-UMass6**		**0.03 ± 0.01**

Finally, we added a diethylphosphonate group via a methyloxy (O-CH_2-O) linker to UM6 at the *para* position of the P1 phenyl (PU6) as this moiety had improved potency in a similar scaffold against highly drug resistant HIV-1 proteases.[242] With this additional phosphonate modification, PU6 inhibited HTLV-1 protease 4-fold better than UM6 (K_i = 0.03 ± 0.01 μM). Therefore, through exploration of a small panel of DRV analogues an HTLV-1 protease inhibitor with low-nM potency was identified with over 26-fold improvement over the FDA-approved inhibitor DRV.

6.3.2 Inhibition of Gag Processing in HTLV-1 VLP's

In addition to the biochemical enzymatic assays, DRV and the two analogs UM6/PU6 were tested for their ability to prevent Gag maturation (**Figure 6.2**). HTLV-1 Gag expressed in Hek293T cells produced viral-like particles (VLPs), which were harvested, resolved by SDS-PAGE, and visualized using anti-capsid antibodies (see Methods for details). In the absence of any inhibitor, HTLV-1 protease rapidly processed Gag in VLPs to yield smaller fragments corresponding to cleavage products. DRV and both analogs effectively inhibited Gag cleavage and VLP maturation at as low as 1:1 ratio. Decreasing inhibitor concentration resulted in a dose-dependent response, as expected. Therefore, DRV analogs inhibit HTLV-1 protease thus Gag processing, which is required for viral maturation.

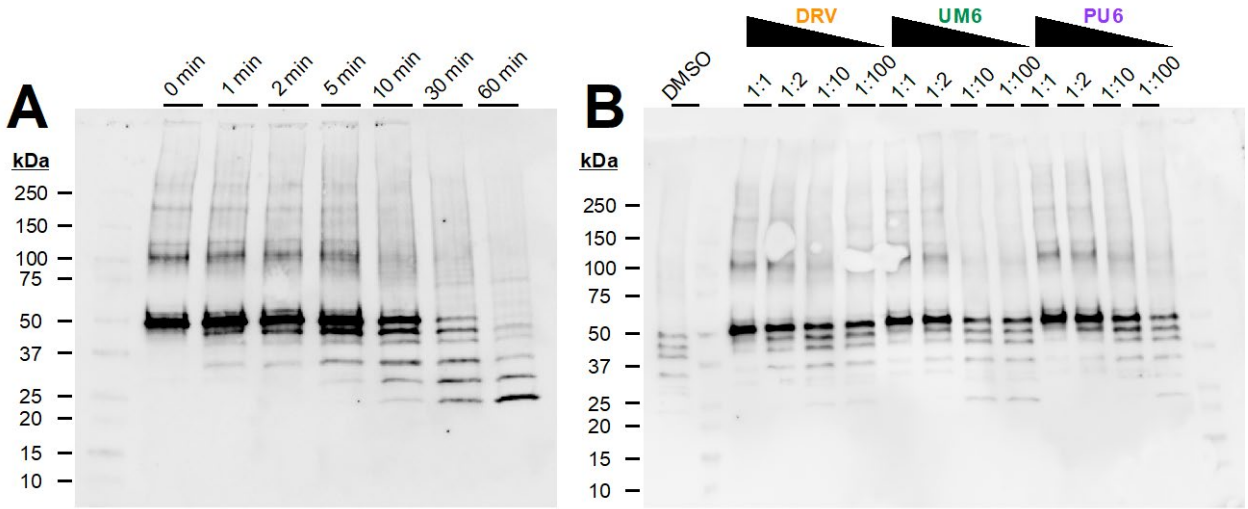


Figure 6.2 HTLV-1 Gag cleavage by protease in VPLs. (A) Gag cleavage by HTLV-1 protease (1 μM) at 37° C without inhibitor as a function of time. Cleavage products visualized with anti-HTLV-1 p24 (capsid). (B) Gag cleavage after 60 min with decreasing molar ratio of inhibitor to protease. At a 1:1 ratio (1 μM), no Gag cleavage products are observed for any inhibitor, while below 1:10 ratio (100 nM inhibitor) some p24 capsid is observed after 60 min.

6.3.3 Cocrystal Structures of DRV and Analogs Bound to HTLV-1 Protease

To elucidate how the DRV scaffold binds to HTLV-1 protease, high-resolution cocrystal structures with DRV, UM6, and PU6 (2.05–2.29 Å) were determined (**Table 6.4**). The three complexes crystallized in the same space group ($P6_322$), containing one homodimer in the asymmetric unit. For comparison, the structure of PU6 bound to HIV-1 protease was also determined to 1.84 Å resolution (**Table 6.4**). Following established convention, the monomer that contacts the aniline side of DRV (P2′) is denoted the prime (′) or B chain, while the A chain that interacts with the bis-THF (P2) will remain non-prime. Starting from these crystal structures, a series of molecular dynamics (MD) simulations were performed to assess the stability of the inhibitor–protease interactions observed in the crystal structures.

Table 6.4 X-ray data collection and crystallographic refinement statistics.

Protease-Inhibitor	***HTLV-DRV***	***HTLV-UM6***	***HTLV-PU6***	***HIV-PU6***
PDB ID	6W6Q	6W6R	6W6S	6W6T
Data Collection				
Space group	$P6_322$	$P6_322$	$P6_322$	$P2_12_12_1$
Cell dimensions:				
a (Å)	78.5	77.9	76.6	51.1
b (Å)	78.5	77.9	76.6	58.0
c (Å)	160.6	160.1	157.3	61.7
α, β, γ (°)	90, 90, 120	90, 90, 120	90, 90, 120	90, 90, 90
Resolution (Å)	40.1 - 2.10 (2.18 - 2.10)	41.9 - 2.05 (2.12 - 2.05)	39.3 - 2.29 (2.37 - 2.29)	21.9-1.84 (1.91-1.84)
Unique reflections	17823 (1728)	18768 (1804)	12954 (1244)	16226 (1420)
Total reflections	162887 (16621)	173500 (15456)	114508 (11155)	104007 (4828)
R_{merge} [a]	0.09 (2.00)	0.10 (3.49)	0.24 (3.89)	0.06 (0.29)
R_{pim}	0.03 (0.68)	0.04 (1.25)	0.09 (1.36)	-
CC1/2	1.00 (0.37)	1.00 (0.23)	0.99 (0.15)	-
*CC**	1.00 (0.73)	1.00 (0.62)	1.00 (0.50)	-
Completeness (%)	99.8 (100)	99.8 (99.6)	99.9 (99.9)	98.4 (88.0)
Redundancy	9.1 (9.6)	9.2 (8.6)	8.8 (9.0)	6.4 (3.4)
Average I/σ	12.6 (1.2)	12.2 (0.7)	5.9 (0.9)	22.7 (3.3)
Wilson B-factors (Å²)	55.2	50.6	61.3	25.2
Refinement				
R_{factor} (%)[c]	22.9	22.1	26.6	20.0
R_{free} (%)[d]	26.2	24.3	30.5	22.7
RMSD[b] in:				
Bond lengths (Å)	0.007	0.012	0.006	0.003
Bond angles (°)	0.81	1.32	0.82	0.59
Ramachandran:				
Favored	96.05	96.49	96.05	99.48
Allowed	3.95	3.07	3.95	0.52
Outliers	0.00	0.44	0.00	0.00

[a] $R_{sym} = \Sigma \mid I - \langle I \rangle \mid / \Sigma I$, where I = observed intensity, $\langle I \rangle$ = average intensity over symmetry equivalent; values in parentheses are for the highest resolution shell.
[b] RMSD, root mean square deviation.
[c] $R_{factor} = \Sigma \mid\mid F_o\mid - \mid F_c\mid\mid / \Sigma \mid F_o\mid$.
[d] R_{free} was calculated from 5% of reflections, chosen randomly, which were omitted from the refinement process.

6.3.4 Interactions Lost in the Binding of DRV to HTLV-1 Versus HIV-1 Protease

In the cocrystal structure determined, DRV was bound at the active site of HTLV-1 protease with an overall conformation similar to that in HIV-1 protease (**Figure 6.3**). However, because of variations in the active site, evident in dramatically altered substrate specificity, certain key protease–inhibitor interactions were either lost or weakened in HTLV-1 protease. As with all HIV-1 protease inhibitors, DRV is a transition state analogue: the central hydroxyl moiety interacts with the side chain oxygens of both catalytic aspartates at a distance of 2.5–3.2 Å between heavy atoms. In HTLV-1 protease, these distances increased to 2.8–3.5 Å. In addition, the more open active site in HTLV-1 protease resulted limited packing of the inhibitor and the flaps, with no evidence of coordinated waters in the electron density maps. Finally, DRV is not large enough to keep optimal distance to the catalytic residues while coordinating bis-THF moiety interactions at the S2 subsite. In HIV-1 protease, the bis-THF oxygens have hydrogen bonds with the backbone nitrogen atoms of D29 and D30 [**Figure 3B**] within 2.8–3.2 Å while remaining 3.5 Å away from the repulsive force of the carbonyl oxygen of D30. This binding also allows the NH atom of the carbamate linker to interact with the carbonyl oxygen of Gly27 at a distance of 3.0 Å. In HTLV-1 protease, the bis-THF had a single hydrogen bond with the backbone nitrogen of D36 at a favorable distance of 2.9 Å. The nitrogen of M37 and carbonyl oxygen of Gly34 were both beyond hydrogen bonding distance (3.4–3.6 Å). Thus, these three major disruptions of hydrogen bonds in HTLV-1 protease (**Figure 3B**) between (1) the central hydroxyl and catalytic aspartates (which centers the inhibitor within the active site); (2) the loss of the tetrahedral network coordinating the inhibitor to the flaps through a conserved water; (3) the bis-THF moiety and backbone nitrogens in

the S2′ subsite, weaken the affinity with DRV as previously observed in highly DRV-resistant HIV-1 protease variants.[243]

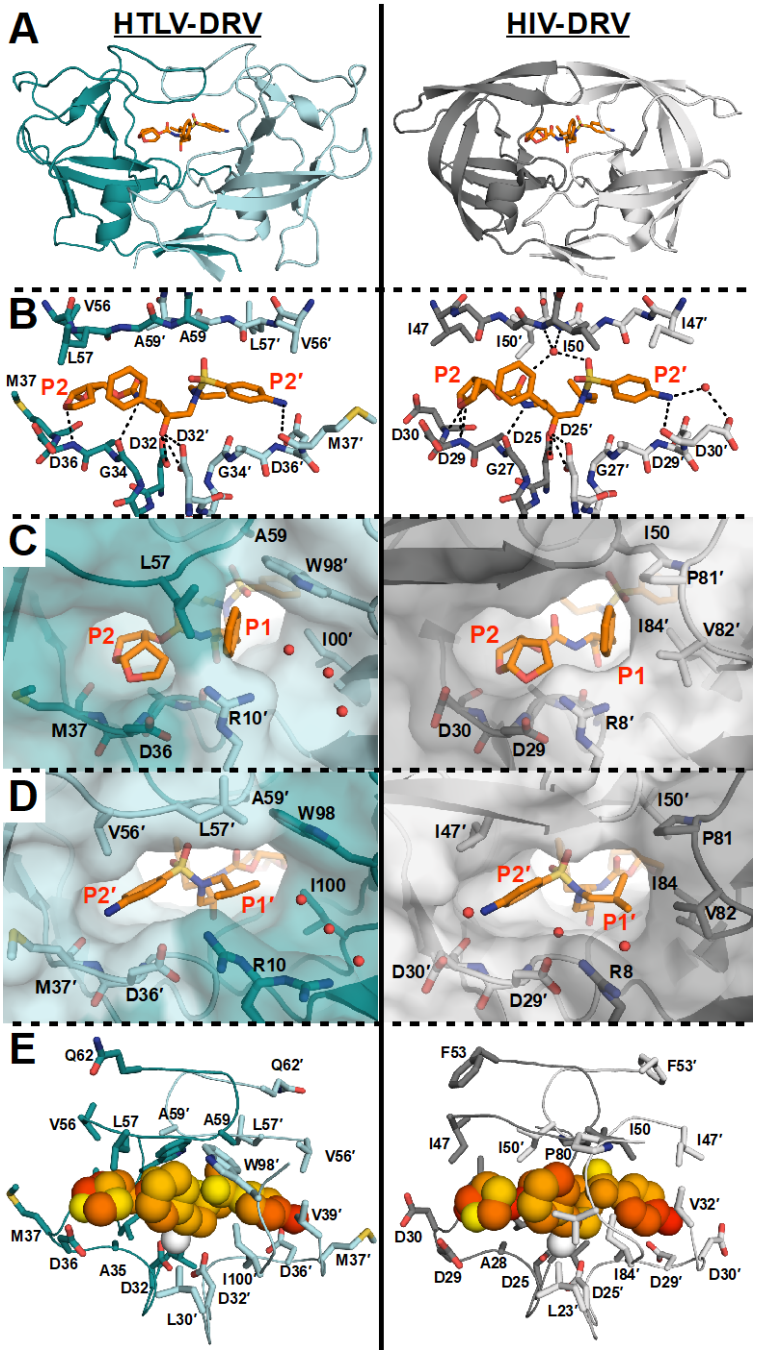


Figure 6.3 Comparison of DRV binding to HTLV-1 versus HIV-1 protease. (A) Cocrystal structures of DRV bound to HTLV-1 and HIV-1 protease (PDB IDs 6W6Q and 6DGX, respectively); (B) close-up view of bis-THF moiety in the S2 subsite; (C) close-up view of aniline moiety in the S2′ subsite; (D) inter-molecular hydrogen bonds between DRV and protease active site; (E) packing around DRV at the active site visualized through mean inter-molecular vdW interactions with protease during MD simulations, where yellow to red indicates increased packing. In all panels, the prime side monomer (chain A) interacting with the bis-THF moiety is in darker shade, and DRV is depicted as orange sticks in panels A-D.

The packing of DRV is dramatically different in HTLV-1 compared with HIV-1 protease. In HTLV-1, the phenyl ring of P1 moiety does not pack well against the hydrophobic residues L30′, W98′, I100′, and L57 in the S1 subsite (**Figure 6.3C**) and the exposed ring coordinates a channel filled with water molecules. In contrast in HIV-1 protease, the phenyl ring of P1 packs against the hydrophobic side chains in the 80s loop, especially V82 (**Figure 6.3C)**. This reflects the substrate specificity of HIV protease where residues with aromatic rings (Phe and Tyr) naturally occur at the P1 position of cleavage sites [**Table 6.1**] and modifications to increase van der Waals (vdW) contacts at this site can greatly improve potency against resistant proteases.[203, 242, 244] The cocrystal structure with HTLV-1 protease indicates modifications to the P1 moiety to either stack against the hydrophobic side chains in the S1 subsite or exploit the available space in the channel may increase potency.

In homodimeric proteases, the S2′ subsite is symmetrically related to the S2 subsite and made up of identical residues. In the S2′ subsite of HIV-1 protease, DRV P2′ aniline makes a direct hydrogen bond with the backbone oxygen of D30′ and a water-mediated interaction with the side chain of D30′ [**Figure 6.3D**]. In our DRV analogs (UM7-9), modifications to P2′ –including phenoxymethyl, hydroxymethylbenzene, benzo(1,3)dioxole– improved potency against resistant variants by establishing more favorable interaction with the backbone nitrogen of D30′.[45] In contrast, DRV and analogs harboring the aniline moiety had the best potency among those tested against HTLV-1 protease (**Figure 6.1, Table 6.3**). The cocrystal structure (**Figure 6.3D**) shows the aniline nitrogen is roughly equidistant from both the backbone oxygen and backbone nitrogen of residue M37′ but not within hydrogen bonding distance. The aniline benzene ring had a

slightly different rotation angle and S2' subsite had a deeper hydrophobic pocket beyond V56′ toward F67′. These differences may underlie why the P2′ modifications did not increase potency toward HTLV-1 protease.

When DRV bound to HTLV-1 protease, both P1 and P1′ moieties were largely exposed to channels occupied by water molecules (**Figure 6.3C,D**). These moieties are nestled in hydrophobic pockets in HIV-1 protease with favorable hydrophobic contacts. In HTLV-1 protease the favorable packing around these moieties was substantially decreased, which was also the case for the overall inhibitor due to the larger active site (**Figure 6.3E**).

6.3.5 Structural Basis of Improved Potency of DRV Analogs Toward HTLV-1 Protease

Cocrystal structures of DRV analogs that had improved potency were determined with HTLV-1 protease. Both UM6 and PU6 have a larger isohexyl moiety at the P1′ position compared to DRV's isobutyl. When bound to the protease, these aliphatic moieties had diverging conformations (**Figure 6.4A**) which could all be accommodated within the relatively large hydrophobic S1′ subsite (**Figure 6.4B**). The conformation of the P1′ moiety seemed to impact that of the P2′ aniline, suggesting subsite interdependence and the need to optimize these moieties simultaneously. The P1′ isohexyl moiety was not large enough to displace any water molecules in the extended S1' channel. However, hydrophobic packing against S1′ residues (Ala59′ and Trp98) were enhanced (**Figure 6.1**), likely underlying the improved potency of analogs with the larger P1′ moiety. Even larger P1′ moieties that pack against these hydrophobic moieties while extending into the channel with a polar face to interact with Arg10 may further increase the binding affinity.

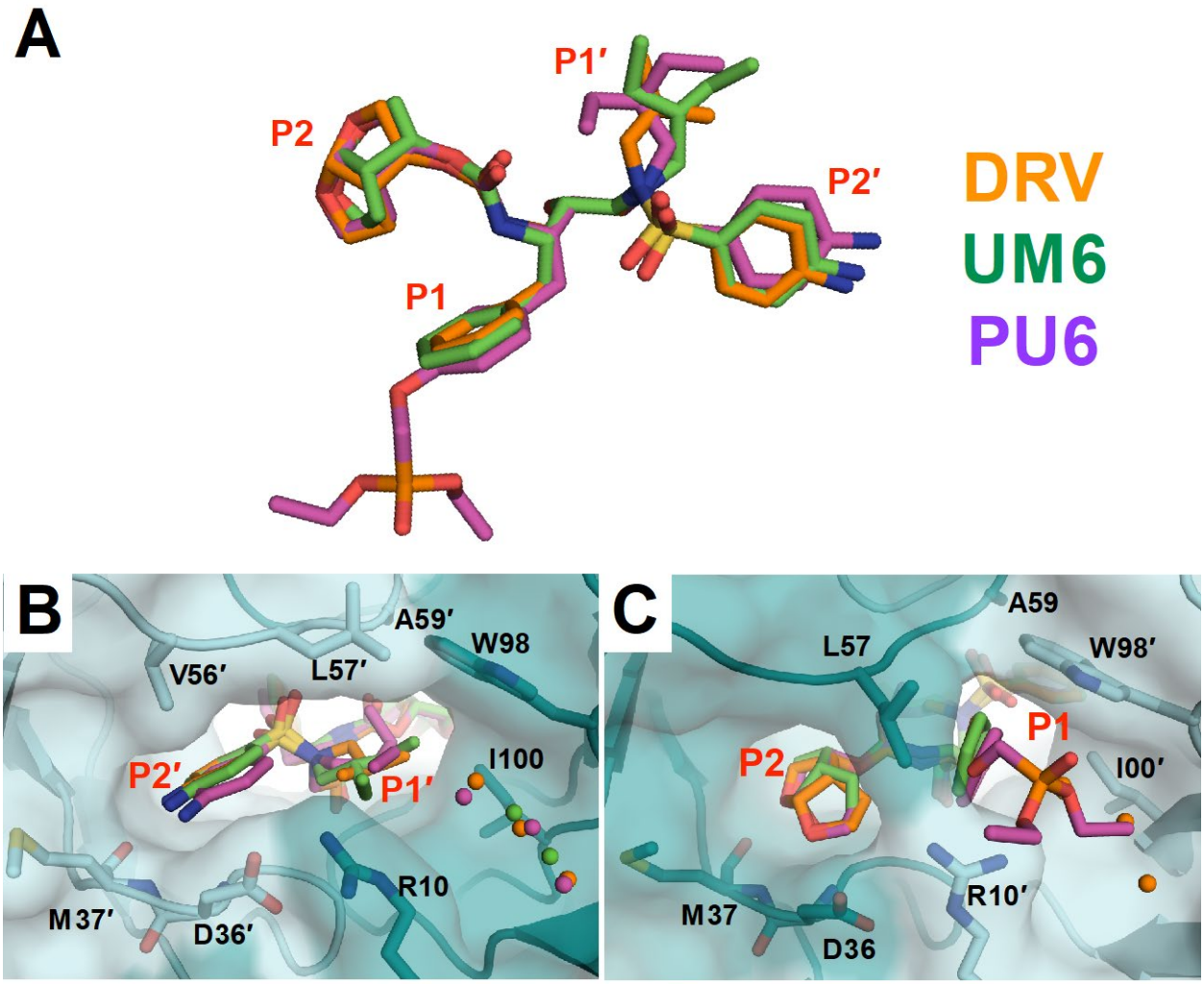


Figure 6.4 Comparison of DRV and DRV analogs when bound to HTLV-1 protease. (A) Alignment of inhibitors. (B) Close-up view of P2′-P1′ moiety in the S2′-S1′ subsite. The P1′ moiety does not have a conserved binding conformation. (C) Close-up view of P2-P1 moieties in the S2-S1 subsite. The phosphonate moiety extends into the S1 subsite, displacing conserved crystallographic waters.

On the flip side of the active site, the invariant P2 bis-THF and P1 benzene ring superimposed well with only minor conformational divergence (**Figure 6.4A, C**). The phosphonate added to the P1 ring in PU6 extended into the S1 channel, substantially increasing vdW contacts with Leu57, Trp98′, Arg10′ (**Figure 6.5**). These account for the order of magnitude increase in the measured inhibition constant (**Figure 6.1**).

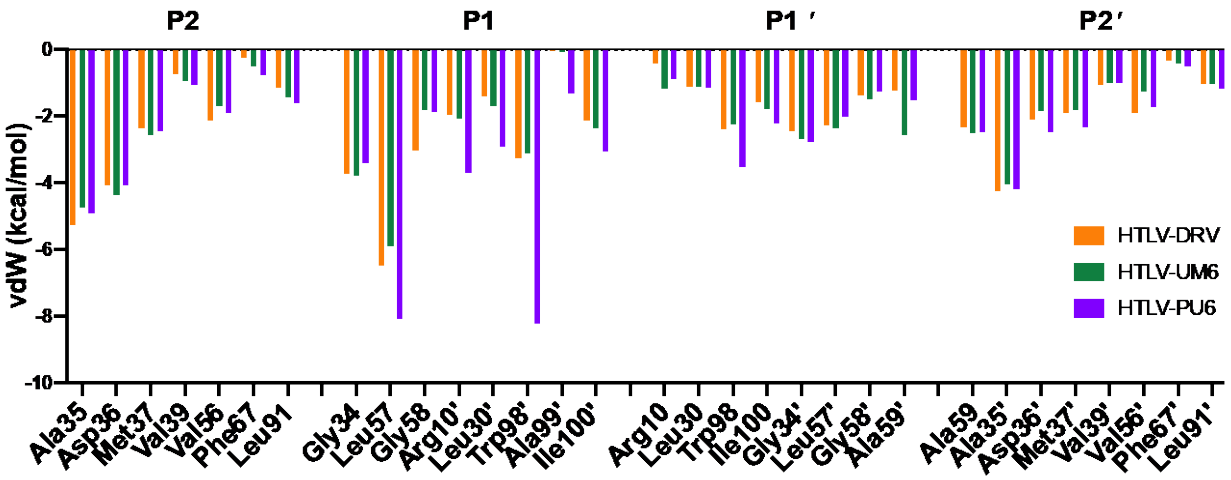


Figure 6.5 Per-residue vdW contacts between protease and inhibitor, grouped by inhibitor moiety.

The phosphonate addition to the P1′ benzene ring has been reported for other HIV-1 protease inhibitor scaffolds to increase potency against resistant variants by “solvent anchoring”.[242] We determined the crystal structure of our PU6 inhibitor bound to HIV-1 protease to both investigate the mechanism of increased potency and to compare with HTLV-1 protease (**Figure 6.6**). The conformation of the phosphonate moiety was completely different in the two cocrystal structures. In HIV-1 protease, one branch of the phosphonate moiety interacted with Pro81′ in the 80s loop while the other extended to the flaps. This suggests that rather than solvent interactions, the mechanism of improved potency against resistant HIV-1 protease variants is through interactions with the invariant Pro81′ and stabilization of the closed conformation of the flaps. In HTLV-1 protease, the phenyl ring intercalates between Leu57 and Trp98′, while Leu57 prevents the phosphonate group from interacting with the protease flaps. Instead, the phosphonate addition extends into the S1 channel, as detailed above.

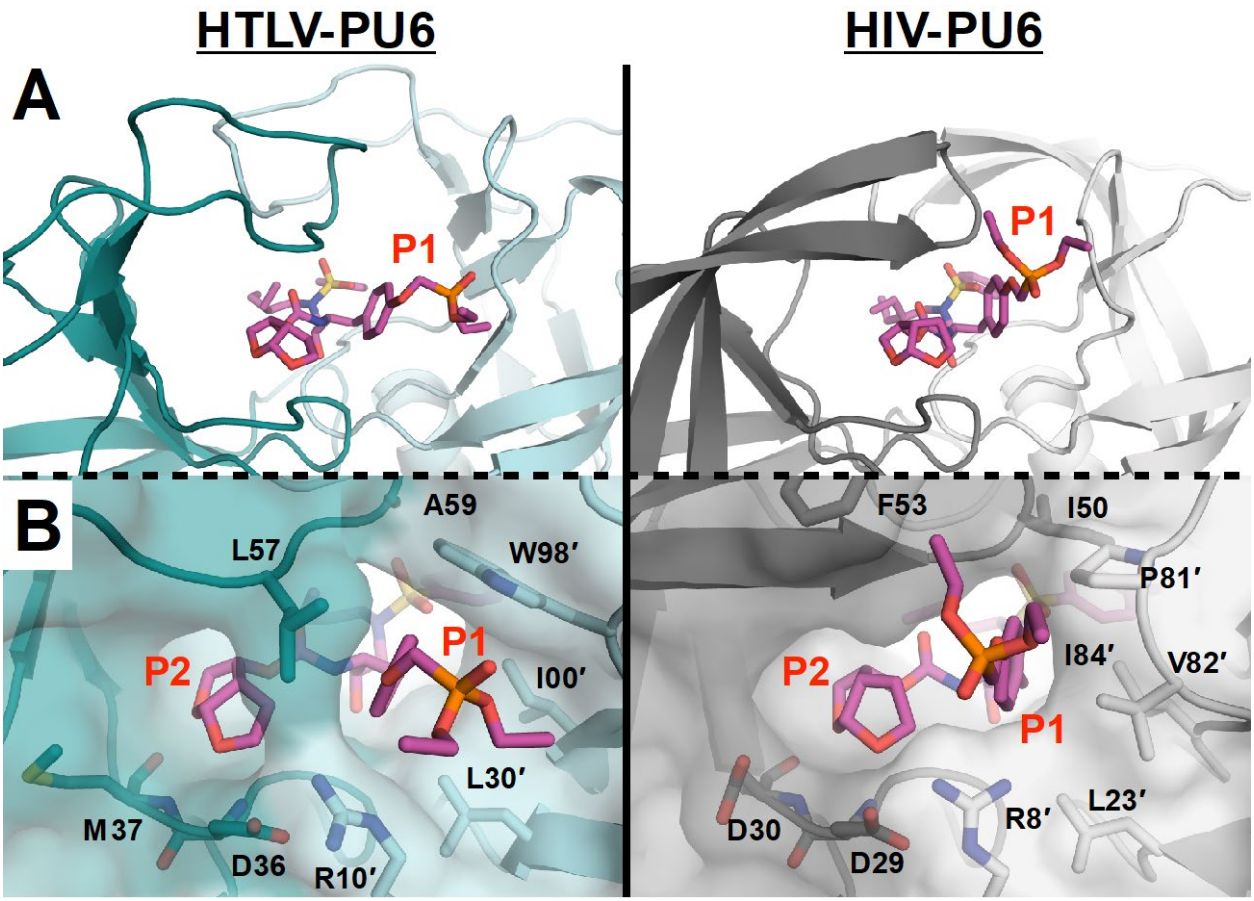


Figure 6.6 Comparison of PU6 when bound to HTLV-1 and HIV-1 protease. (A) The phosphonate moiety of PU6 sticks out into the S1 subsite in HTLV-1, whereas it binds up against the flaps in HIV-1 protease. (B) Residue L57 prevents the phosphonate from interacting with the flaps of HTLV-1 protease.

The crystal structures indicate that exploration of larger P1′ and P1 moieties to fill the water-occupied channels is highly promising. Our MD simulations indicate Ala59/59′ and Trp98/98′ are highly flexible (**Figure 6.7**). Modifications that increase interactions with Ala59/59′ to stabilize the flaps and with Trp98/98′ to lock the side chain conformer will likely further increase inhibitor potency. These tryptophan residues and phenylalanine (Phe67/Phe67′) in the S2/S2′ subsites also provide an opportunity for π-π stacking interactions between protease and inhibitor (**Figure 6.8**). Previously reported C2-symmetric pyrrolidine-based inhibitors with P1/P1′ aromatic rings had π-π stacking interactions with Trp98, which improved potency.[239] In our MD simulations, π-π stacking of P1 phenylalanine with Trp98 was maintained only about half of the time, and only PU6 formed edge-to-face π-π stacking with Phe67′ again for about half of the simulation time. Stabilizing these stacking interactions may further increase potency. Overall, analysis of crystal structures suggests that DRV offers a promising scaffold for HTLV-1 protease inhibitor design, and pinpoints interactions that can be optimized to achieve potent inhibition.

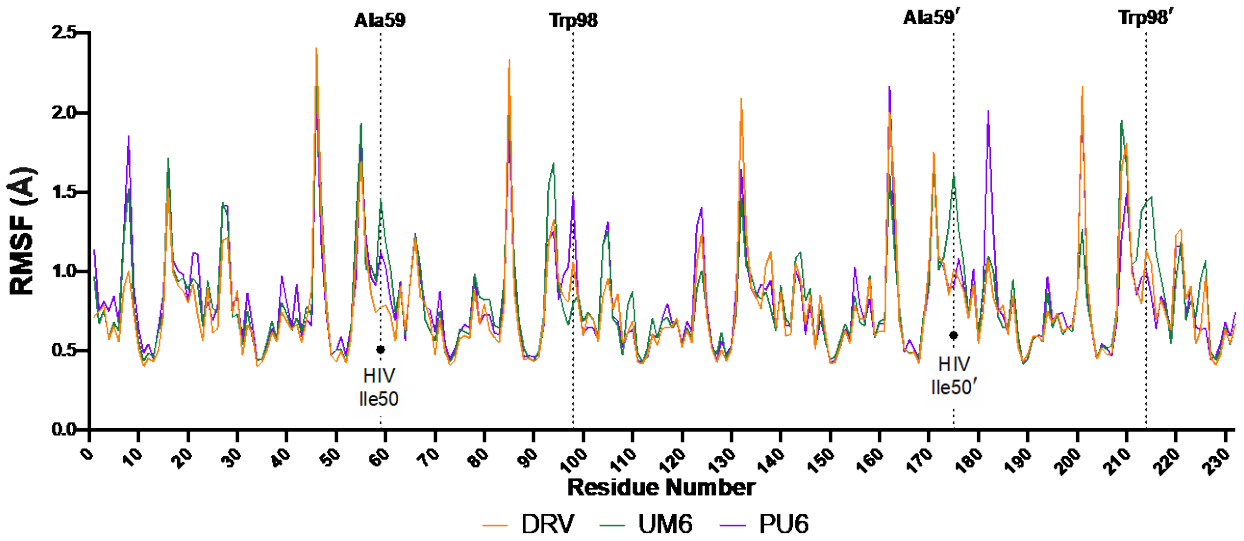


Figure 6.7 Root-Mean-Square Fluctuation (RMSF) of Cα atoms for DRV, UM6 and PU6 in complex with HTLV-1 protease from molecular dynamics simulations.

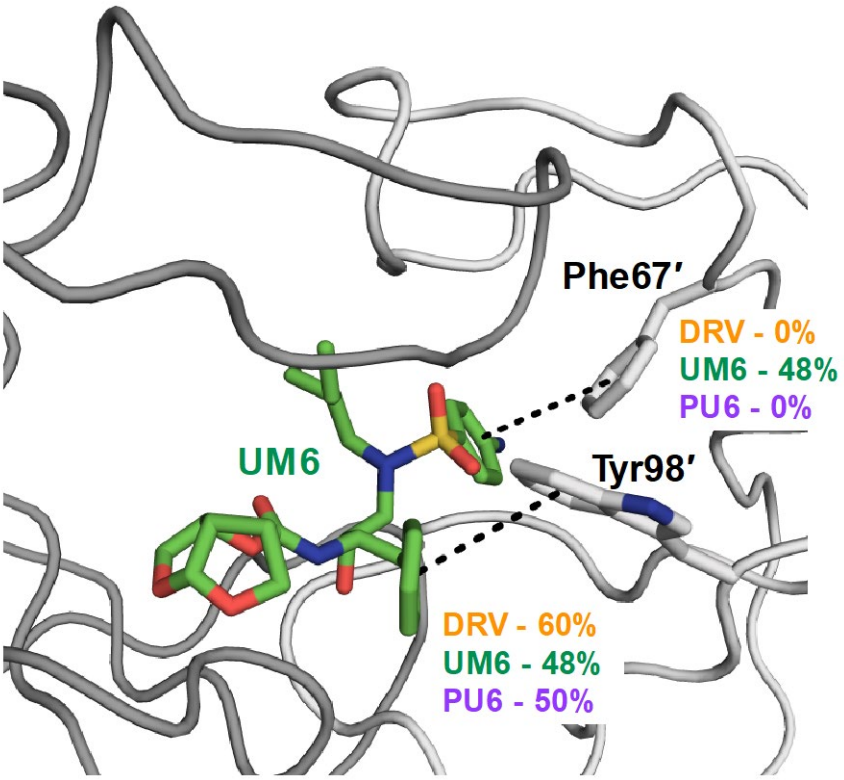


Figure 6.8 Aromatic side chains in HTLV-1 active site can form π-π stacking interactions, primarily in edge-to-face configuration, with the P1 phenylalanine and P2′ aniline of DRV analogs. Frequency of the π-π stacking interactions (dashed lines) during MD simulations is indicated.

6.4 Conclusions

HTLV-1 is a growing global threat with no DAAs or vaccines to prevent the spread of the disease. We sought to inhibit HTLV-1 protease using DRV, the most potent FDA-approved HIV-1 protease inhibitor. Our results show that DRV is a promising scaffold for inhibitor design and antiviral development against HTLV-1 protease. With modifications to two moieties of DRV, the lead compound (PU6) achieved substantially (~27-fold) improved potency *in vitro* and inhibited Gag processing by HTLV-1 protease. The structural insights gained here will guide the design of highly potent HTLV-1 protease inhibitors.

6.5 Methods

6.5.1 Construction of the HTLV-1 Protease Expression Plasmid

The 116 amino acid HTLV-1 variant (Uniprot Accession ID: Q82134) was ordered from Genscript on a pET11a plasmid with codon optimization for protein expression in Escherichia coli. A L40I mutation was included to prevent autoproteolysis.[241]

6.5.2 Protein Expression and Purification

The expression, isolation, and purification of HTLV-1 protease used for the kinetic assays and crystallization was carried out as previously described for HIV-1 protease.[40, 99] Briefly, the gene encoding the HTLV-1 protease was subcloned into the heat-inducible pXC35 expression vector (ATCC) and transformed into *E. coli* TAP-106 cells. Cells grown in 6 L of Terrific Broth were lysed with a cell disruptor and the protein was purified from inclusion bodies.[245] The inclusion body centrifugation pellet was dissolved in 50% acetic acid followed by another round of centrifugation to remove impurities. Size exclusion chromatography was used to separate high molecular weight proteins from the desired

protease. This was carried out on a 2.1 L Sephadex G-75 superfine column (Millipore Sigma) equilibrated with 50% acetic acid. The cleanest fractions of HTLV-1 protease were refolded into a 10-fold dilution of 0.05 M sodium acetate at pH 5.5, 5% ethylene glycol, 10% glycerol, and 5 mM DTT. Folded protein was concentrated down to 1–2 mg/mL and stored at -80°C. Protein aliquots were thawed, concentrated to ~10 mg/mL, and a final purification was performed with a Pharmacia Superdex 75 FPLC column equilibrated with 0.05 M sodium acetate at pH 5.5, 5% ethylene glycol, 10% glycerol, and 5 mM DTT. Protease fractions purified from the size exclusion column were concentrated to 1–2 mg/mL using an Amicon Ultra-15 10-kDa device (Millipore). Freshly purified protease was used for crystallization and the rest was stored at -80°C. This stored protease was used for K_M and K_i assays.

6.5.3 Enzyme Binding Assays to Determine K_M

To determine the enzyme activity, in a 96-well plate, purified protease was provided a 10-amino acid substrate containing the natural cleavage site (MA/CA or CA/NC) with an EDANS/DABCYL FRET pair dissolved in 8% DMSO was 2/3 serially diluted from 0 to 40 μM for MA/CA and 0 to 100 μM to CA/NC. HTLV-1 protease was diluted to 12 μM and, using a PerkinElmer Envision plate reader, and 5 μL were added to the 96-well plate to obtain a final concentration of 1 μM. Fluorescence was observed with an excitation at 340 nm and emission at 492 nm and monitored for 200 counts. A FRET inner filter effect correction was applied as previously described.[103] Data corrected for the inner filter effect was analyzed with Prism8, as described in Chapter 3.

6.5.4 Enzyme Inhibition Assays to Determine K_i

To determine the enzyme inhibition constant (K_i), in a 96-well plate, each inhibitor was 2/3 serially diluted from 400 μM to 6.9 μM for IDV, 30 μM to 0.5 μM for DRV and UM6 or 2 μM to 0.03 μM for PU6. All assays included a 0 μM control, and incubated with 1 μM protein for 1 hour. A 10-amino acid substrate containing a solubility enhanced HTLV-1 MA/CA protease cleavage site with an EDANS/DABCYL FRET pair (BAChem) was dissolved in 4% DMSO at 120 μM. Using the Envision plate reader, 5 μL of the 120 μM substrate was added to the 96-well plate to a final concentration of 10 μM. The fluorescence was observed with an excitation at 340 nm and emission at 492 nm and monitored for 200 counts. Data was analyzed with Prism8, as described in Chapter 2. UM6 was used as a control in all assays.

6.5.5 Protein Crystallization

Many crystallization conditions produced HTLV-1 protease cocrystals with a hexagonal plate morphology, but the condition most reliably producing larger crystals in three dimensions was discovered using the JCSG+ sparse screen, well C6, containing 40% (v/v) PEG 300, 0.1M Phosphate/Citrate pH 4.2. All cocrystals were grown at room temperature by hanging drop vapor diffusion method in a 24-well VDX hanging-drop trays (Hampton Research) with a protease concentration of 4.0 mg/mL with 5-fold molar excess of inhibitor and mixed with the precipitant solution at a 1:1 ratio. The precipitant solution consisted of 39–41% (v/v) PEG 300 with 0.1 M Phosphate/Citrate buffer at pH 4.2 and the crystallization drops were set with 1 μL of precipitant solution and 1 μL protein-inhibitor solution and micro-seeded with a cat whisker and dried over a well solution of 3.0-4.0 M NaCl. Diffraction quality crystals were obtained within 2 weeks. Data were collected at

100 K and due to the high percentage of low molecular weight PEG there was no need for supplemental cryoprotectant. The cocrystal of PU6 bound to HIV-1 protease was generated as previously described.[37, 213, 243, 246]

6.5.6 Data Collection and Structure Solution

The three HTLV-1 protease cocrystals were shot at the Chicago APS Synchrotron Beamline 23-ID-D using beamline control software JBluIce.[247] The diffraction images were indexed, integrated, and scaled using the GM/CA autoprocessing pipeline which utilizes XDS,[248] POINTLESS,[249] and AIMLESS.[250] X-rays diffracted through a cocrystal of PU6 bound to HIV-1 protease were collected by our in-house Rigaku Saturn944 system and intensities were indexed, integrated, and scaled using HKL3000.[105]

Structures were solved by molecular replacement in the program PHASER[106] using an HTLV-1 protease monomer (PDB: 3WSJ) or a WT HIV-1 protease monomer (PDB: 6DGX). Model building and refinement was performed using Coot[107] and Phenix.[108] Ligands were designed in Maestro and the output sdf file was used in the Phenix program eLBOW[109] to generate the cif file containing atomic positions and constraints necessary for ligand refinement. Iterative rounds of crystallographic refinement were carried out until convergence was achieved. To limit bias throughout the refinement process, five percent of the data were reserved for the free R-value calculation.[110] MolProbity was applied to evaluate the final structures before deposition in the PDB.[111] Structure analysis, superposition and figure generation was done using PyMOL.[251] X-ray data collection and crystallographic refinement statistics are presented in the Supporting Information [**Table 6.4**].

6.5.7 Structural Analysis

To calculate the intermolecular vdW interaction energies the crystal structures were prepared using the Schrödinger Protein Preparation Wizard.[216] Hydrogen atoms were added, protonation states were determined, and the structures were minimized. The protease active site was monoprotonated at Asp25. Subsequently, force field parameters were assigned using the OPLS3 force field.[217] Interaction energies between the inhibitor and protease were estimated using a simplified Lennard-Jones potential, as previously described in detail.[218] Briefly, the vdW energy was calculated for pairwise interactions depending on the types of atoms interacting and the distance between them.

6.5.8 Molecular Dynamics Simulations

High resolution HTLV-1 cocrystal structures with DRV, UM6, and PU6 (PDB: 6W6Q, 6W6R, 6W6S) were used as starting coordinates for molecular dynamics simulations. All starting structures were prepared using the Protein Preparation Wizard from Schrodinger.[216] Crystallographic water molecules were retained, missing atoms were added using Prime,[221] and PROPKA[143-144] was used to determine the protonation state of side chains at pH 7.0. The resulting structure was minimized under restraint to a convergence criterion of 0.3 Å using Impref.[145]

All molecular dynamics simulations were carried out using Desmond, within Schrodinger, [216] with the OPLS3e force field for the inhibitor and protein. The systems were prepared as previously discussed.[85] Briefly, the cocrystal structures were placed within a cubic TIP3P water box measuring 15Å on each side. Chloride ions were first used to neutralize the system and sodium and chloride atoms were added to reach a physiological 0.15 M salt concentration. Prior to simulation, each solvated system was

relaxed using a series of restrained minimization stages as previously described.[85] These stages consisted of successive minimizations with restraints on i) the heavy protein atoms, ii-iii) the protein backbone atoms and finally iv) no restraints. The restraining force constants were 1000, 1000 and 5 kcal mol^{-1} $Å^{-2}$ for stages i), ii) and iii), respectively and the minimization was done using steepest descent followed with the limited-memory BFGS method to a tolerance of 0.5 kcal mol^{-1} $Å^{-1}$. During unrestrained minimization, this tolerance was further reduced to 0.05 kcal mol^{-1} $Å^{-1}$. Molecular dynamics for each system was carried out in triplicate, with each of the three 100 ns simulations starting with different randomized velocities. The root-mean-square deviation (RMSD) from the starting structures indicates that the simulations reached equilibrium after 10 ns.

6.5.9 Analysis from MD Simulations

vdW packing about the inhibitor was calculated as previously described.[40, 147] Briefly, force field parameters are not optimized for sulfurs violating the octet rule (e.g. the sulfur of the P2' moiety of DRV, UM6, and PU6). As such, the packing around this sulfur atom was calculated by averaging the vdW packing of the four adjacent atoms.

π-π stacking interactions were analyzed utilizing the simulation interaction diagram within Maestro from the Schrodinger Suite.[216] A π-π stacking interaction in the edge-to-face configuration was identified when the distance between the ring centroids was less than 5.5 Å and the angle between the rings was between 60° and 120°. A π-π stacking interaction in the edge-to-face configuration was identified when the distance between the ring centroids was less than 5.5 Å and the angle between the rings was between 60° and 120°. A π-π stacking interaction in the face-to-face configuration was identified when

the distance between the ring centroids was less than 4.4 Å and the angle between the rings was less than 30°.

6.6 Acknowledgements

This research used resources of the Advanced Photon Source, a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under Contract No. DE-AC02-06CH11357. GM/CA@APS has been funded in whole or in part with Federal funds from the National Cancer Institute (ACB-12002) and the National Institute of General Medical Sciences (AGM-12006). The Eiger 16M detector was funded by an NIH–Office of Research Infrastructure Programs, High-End Instrumentation Grant (1S10OD012289-01A1). We thank the beamline specialists at 23-ID-D for their help in data collection.

7 CHAPTER 7:

Optimal Refinement of Pseudomerohedrally Twinned $P6_1$ HIV-1 Protease–Inhibitor Cocrystal Structures

7.1 Abstract

Twinning is a crystal growth anomaly in which protein monomers exist in different orientations but are related in a specific way, causing diffraction reflections to overlap. Twinning imposes additional symmetry on the data, often leading to assignment of a higher symmetry space group. Specifically, in pseudomerohedral twinning, reflections imperfectly overlap and require a twin law to separate unique structural data from overlapping reflections. Neglecting twinning in the crystallographic analysis of quasi-rotationally symmetric homo-oligomeric protein structures can mask the degree of structural non-identity between monomers. In particular, any deviations from perfect symmetry will be lost if higher than appropriate symmetry is applied during crystallographic analysis. Such cases warrant choosing between the highest symmetry space group possible, or determining if the monomers have distinguishable structural asymmetries and thus require a lower symmetry space group and a twin law. Using hexagonal cocrystals of HIV-1 protease, a C_2 symmetric homo-dimer whose symmetry is broken by the bound ligand, we show that both assigning a lower symmetry space group and applying a twin law during refinement is critical to achieving the most accurate structural model. By re-analyzing three recently published HIV-1 protease structures, we demonstrate improvements in nearly every crystallographic metric. Most importantly, we find that the inhibitor can be modeled in a single orientation, allowing for more precise structural analysis. This protocol may be applicable to many other homo-oligomers in the PDB.

7.2 Introduction

Macromolecular structures determined through X-ray crystallography allow for high-resolution measurements and structural analyses. The accuracy of structures is especially important for protein–inhibitor complexes, as relatively small errors can greatly mislead drug design and structure-activity relationship studies. A common problem in crystallography is that disordered regions of proteins cannot reliably be modeled due to the averaging out of electron density. Less appreciated, data averaging can also occur when determining structures that are twinned due to pseudosymmetry.

Twinning is a crystal growth anomaly in which the specimen is composed of separate domains whose orientations differ but are related in a specific way (Yeates, 1997). The domains may be rotated, reflected, or inverted with respect to each other causing diffraction pattern reflections to overlap. Because amino acids are chiral, proteins only exhibit rotation symmetry (Parsons, 2003) which may cause twinning. The type of twinning is categorized by the number of dimensions the crystal lattices overlap. Merohedral twinning describes cases in which the lattices of two or more distinct domains coincide exactly in three dimensions, whereas nonmerohedral twinning is when the overlap occurs in fewer than three dimensions. Lattices imperfectly overlapping in three dimensions is referred to as pseudomerohedral (Yeates, 1997). Depending on the space group, twin laws are necessary to correct for crystallographic twinning and refine the structure to a high level of confidence (Zwart et al., 2008). Due to chirality, proteins may crystallize in 65 of the 230 possible crystallographic space groups, and twinning occurs only in lower symmetry space groups of the trigonal, hexagonal, tetrahedral, or cubic systems (Nespolo and Aroyo, 2016). Twinning imposes additional symmetry on the

diffraction pattern and may bias the data to be processed in a space group higher than appropriate (Zwart et al., 2008). In cases of pseudomerohedral twinning, where lattices do not perfectly overlap, processing data in the highest possible space group will average out asymmetric data, without allowing a twin law to later recover asymmetric information (Hamdane et al., 2009, Donovan et al., 2016, Campeotto et al., 2018). Of the ~145,000 structures in the Protein Data Bank (PDB) (Berman et al., 2000, Berman et al., 2003) about 55,000 are homo-oligomers, 96% of which are rotationally symmetric (Tuvi-Arad and Alon, 2019). Pseudomerohedral twinning in such structures may lead to the assumption of perfect symmetry for distinct monomers of homo-oligomers.

HIV-1 protease is a 99 amino acid C_2 symmetric homodimer (Wlodawer et al., 1989). In the apo form, the two monomers are highly symmetric, but binding an inhibitor or substrate induces asymmetry in the protease, especially in the flaps [**Figure 7.1A-B**] (Rose et al., 1998, Prabu-Jeyabalan et al., 2000). HIV-1 protease is among the proteins with the most published structures in the PDB. When this manuscript was submitted, a BLAST search of WT HIV-1 Protease (UniProtID: Q8ULI9) returned 729 structures with at least 50% sequence identity and inhibitor or substrate bound at the active site. Of these 729 cocrystal structures, most were solved in the primitive orthorhombic space groups $P2_12_12$ or $P2_12_12_1$, accounting for 260 and 229 structures, respectively [**Figure 7.2A-B**]. The $P2_12_12_1$ space group is particularly fortuitous as it represents the biologically functional unit, containing two monomers in the asymmetric unit and the inhibitor bound in a single orientation. The $P2_12_12$ space group also contains two monomers in the asymmetric unit, but about a third of the time the inhibitor is modeled in two orientations with split occupancies due to observed C_2 symmetric electron density in the active site.

In these primitive orthorhombic space groups, unless two crystal axes are the same length, twinning is not possible (Yeates, 1997).

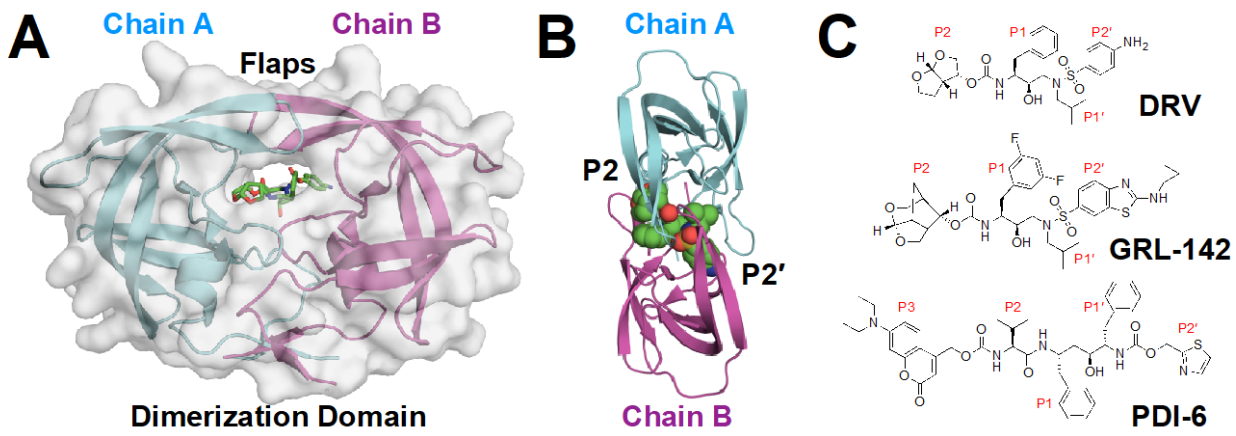


Figure 7.1 A) Side view: HIV-1 protease structure shown in cartoon representation with transparent surface. Inhibitor (DRV) shown as green sticks. B) Top view: HIV-1 protease structure in cartoon representation with DRV shown as green spheres. C) Inhibitors bound to HIV-1 protease whose cocrystal structures were investigated and re-refined in this research.

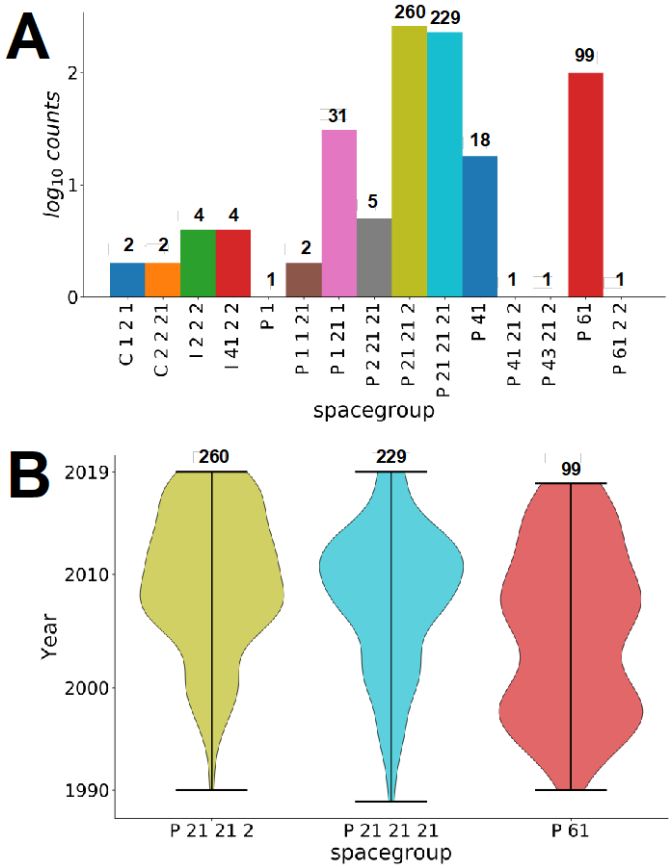


Figure 7.2 A) Number of HIV-1 protease-like cocrystals structures deposited into the PDB by space group. B) Violin plot showing the distribution of cocrystal structures published over time in the three most common space groups.

For HIV-1 protease cocrystals, the next most common space group is $P6_1$ with 99 published structures [**Figure 7.2**]. Hexagonal cocrystals solved in $P6_1$ space group have two monomers in the asymmetric unit, and 50 of 99 structures have asymmetric inhibitors modeled in two orientations at roughly 50% occupancy. When HIV-1 protease crystallizes in this hexagonal form, the monomers remain highly symmetric resulting in pseudomerohedral twinning. This near-perfect symmetry causes the hexagonal data to appear as the higher symmetry space group $P6_122$ (Erickson et al., 1990, Dreyer et al., 1992). If the data were processed in $P6_122$, all overlapping reflections derived from the two monomers would be combined and averaged to model the single monomer per asymmetric unit. Although choosing to process the pseudomerohedrally twinned data in $P6_1$ space group correctly models a dimer, by not applying a twin law when refining the structure, the asymmetries remain averaged out, as if modeling two identical $P6_122$ monomers. As previously described, a twin law is necessary to account for asymmetries in the reflections of overlapping diffraction patterns. Only once was a twin law operator used to refine a high-resolution structure of an apo HIV protease dimer in the $P6_1$ space group (PDB: 3IXO). In this case, the twin law was shown to decrease the R-factors and significantly improve the electron density of the asymmetric flaps (Robbins et al., 2010). Despite this precedent, twin laws have not been applied during refinement of hexagonal HIV-1 cocrystal structures.

To evaluate and quantify the benefit of applying a twin law when refining a pseudomerohedrally twinned homodimer, we refined three HIV-1 protease cocrystal structures with increasingly asymmetric inhibitors [**Figure 7.1C**]. To determine any improvement with the twin law, our recently published HIV-1 protease structure with

darunavir (DRV) bound, both with and without the twin law applied, was refined. Next the twin law was applied to two recently published HIV-1 protease cocrystal structures from the PDB with bound inhibitors that were more asymmetric than DRV. We found that applying a twin law operator during refinement improves the structure and allows the inhibitor to be modeled in a single orientation. This case study may inform hundreds of other structures involving pseudomerohedrally twinned homo-oligomers with or without inhibitors bound.

7.3 Materials and Methods

7.3.1 Protein Preparation and Crystallization

Details involving HIV-1 protease gene construction, protein expression and purification, and cocrystallization steps were previously published (Henes et al., 2019, Aoki et al., 2017, Schimer et al., 2015).

7.3.2 Data Collection and Processing

For our data set (PDB: 6OPV), cocrystal diffraction intensities were indexed, integrated, and scaled using HKL3000 (Otwinowski and Minor, 1997). For the previously published data sets (PDB: 5TYS (Aoki et al., 2017) and 4U7Q (Schimer et al., 2015)), the MTZ files were downloaded from the PDB. Data quality was checked with phenix.xtriage (Adams et al., 2010), which also provided the twin fraction and the suggested twin law. All three data sets were flagged as merohedrally twinned, with twin fractions estimated to be above 0.48 by the cumulative distance of H or above 0.45 by Britton analysis (Britton, 1972, Yeates, 1988). The twin law (h,-h-k,-l) was detected for these hexagonal data sets.

7.3.3 Molecular Replacement

Structures were solved using molecular replacement with PHASER (McCoy et al., 2007). A WT HIV-1 protease monomer from a DRV bound cocrystal structure solved in $P6_1$ was used for molecular replacement (PDB: 6OPS) (Henes et al., 2019).

7.3.4 Refinement

After molecular replacement, model building and refinement was performed using Coot (Emsley and Cowtan, 2004) and Phenix (Adams et al., 2010). During refinement, optimized stereochemical weights and torsion-angle non-crystallographic symmetry operators were utilized. The twin law (h,-h-k,-l) was applied where specified. Ligands were modeled in Maestro and the sdf files were used in the Phenix program eLBOW (Moriarty et al., 2009) to generate the cif file containing atomic positions and constraints necessary for ligand refinement. Iterative rounds of crystallographic refinement were carried out until convergence was achieved. To limit bias throughout the refinement process, five percent of the data were reserved for the free R-value calculation (Brunger, 1992). MolProbity (Davis et al., 2007) was utilized to evaluate the quality of the final structures.

7.3.5 Visualization of Structures

Alignment by superposition and figure generation was done using PyMOL (DeLano, 2002). Throughout this analysis, chain A is depicted as cyan sticks and chain B as light magenta sticks. To depict the electron density and difference map for each structure, Phenix was used to output maps as ccp4 files using map coefficients. Specifically, files used were: the MTZ file with R-free flags, final PDB file, and final MTZ file. Throughout this analysis, $2F_o$-F_c direct maps are depicted as grey mesh contoured at 1.0 σ while the

F_o-F_c difference maps have positive density depicted as green mesh contoured at 3.0 σ and negative density as red mesh contoured at -3.0 σ.

7.4 Results and Discussion

7.4.1 Evaluating Twin Law and Inhibitor Orientation

To determine the benefit of the application of a twin law, the first dataset evaluated was that of a high-resolution (1.91 Å) HIV-1 protease cocrystal structure with DRV bound, which we recently determined (PDB: 6OPV) (Henes et al., 2019). As the data was processed in the $P6_1$ space group, molecular replacement placed two monomers in the asymmetric unit. The dimer was refined to convergence with and without the twin law applied for the duration of the refinement process. In addition, to determine if the inhibitor orientation imposed any structural bias, the structure was refined with bis-THF moiety of DRV interacting with either the A chain, the B chain, or in both conformations at 50% occupancy [**Figure 7.3**]. This resulted in six separate models from the same dataset.

All six structures were refined to convergence and finished with similar R-factor statistics. Comparing similar cocrystal structures, application of the twin law reduced R_{work} by about 0.5% and R_{free} by about 1.5%. The twin law visibly improved the electron density maps surrounding the inhibitor and flap tip residues, apparent by the reduction of density in the F_o-F_c difference map [**Figure 7.3**].

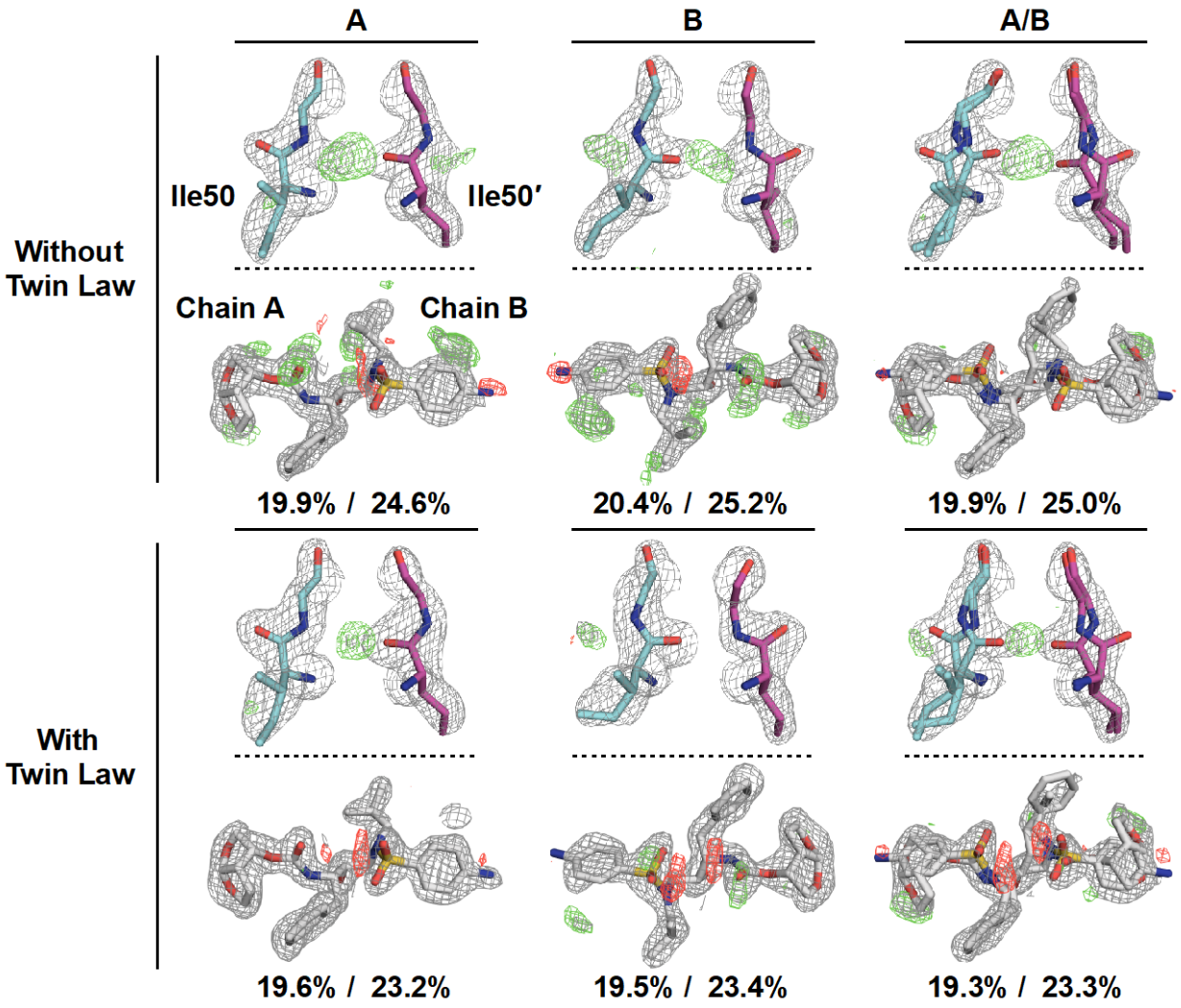


Figure 7.3 Electron density maps surrounding the flap tip residues, Ile50 and Gly51(top), and the inhibitor, DRV (bottom), shown as sticks. A, B, and A/B indicate the orientation of the modeled inhibitor. R_{work}/R_{free} values are indicated below each of the 6 structures refined in the $P6_1$ space group.

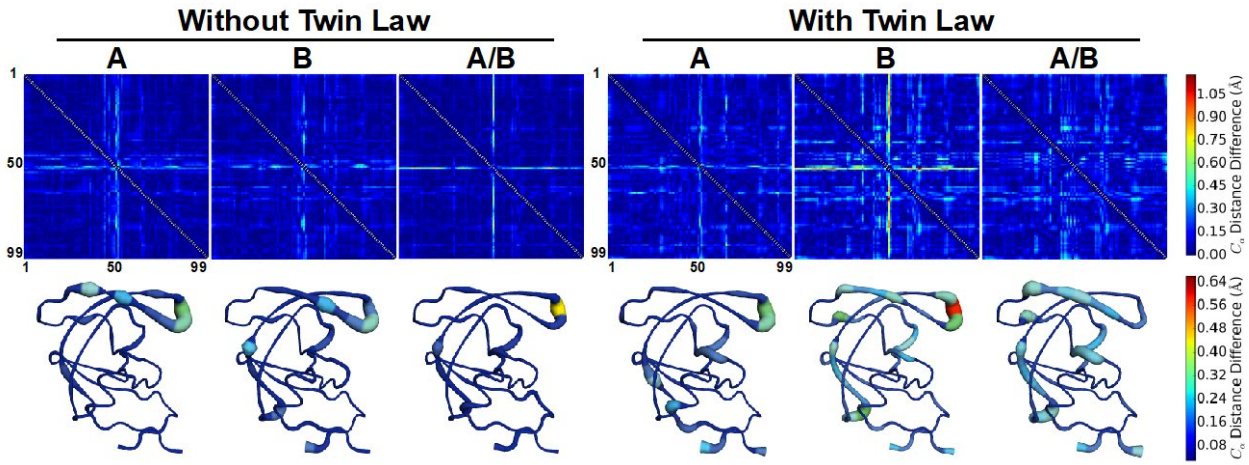


Figure 7.4 Internal distance differences between C_α atoms of chain A and chain B of all re-solved structures of DRV bound HIV-1 protease (PDB: 6OPV) with and without the twin law applied. Thicker tubes and warmer colors indicate larger differences.

When the inhibitor was modeled in two orientations, the refinement began with exactly 50/50 split occupancy between inhibitor orientation A and B. During refinement, when the occupancies were allowed to change without any restriction, but no twin law was applied, the occupancies never diverged beyond a 55/45 split. Interestingly, when the twin law was applied, the inhibitor occupancies converged to an 80/20 split. Comparing successive rounds of refinement, under these parameters, the electron density map around the inhibitor became increasingly asymmetric. This demonstrates the twin law's ability to, without bias, select the preferred single orientation of the inhibitor. In **Figures 7.3-7.4**, the structure with two inhibitors and the twin law applied had the inhibitor occupancies locked at 50/50 for the final round of refinement. This returned C_2 symmetry to the electron density around the inhibitor and allowed for direct comparison to the structure with two inhibitor orientations without the twin law applied.

While all structures refined to similar statistics, the application of the twin law enabled capturing asymmetric details of the structure of the complex, as highlighted by internal distance differences between the two chains. Without the twin law, the structures were highly symmetric, especially in the core [**Figure 7.4**]. Modeling the inhibitor in two orientations induces more symmetry between monomers than observed in the single orientation models, as any actual differences are averaged out. With the twin law, the monomers had more internal variations throughout the protease. This was especially true around the active site where the flaps and 80's loop adopt different conformations between the two monomers, which has been characterized in other space groups when binding asymmetric inhibitors.

This first set of tests showed that the application of a twin law allows a near-perfectly twinned cocrystal structure to be solved as a homodimer with the inhibitor and flaps modeled in a single orientation without losing any confidence in the model. The twin law, in fact, teases out known asymmetries between the monomers, resulting in a more accurate model.

7.4.2 Re-Solving Structure with a More Asymmetric Inhibitor

DRV is a fairly symmetric inhibitor when rotated about the C_2 axis, as its P1 and P2 moieties roughly occupy the same three-dimensional space as its P1′ and P2′ moieties [**Figure 7.1C**]. To further test this protocol, we chose to investigate an HIV-1 protease cocrystal structure with a more asymmetric DRV analog, GRL-142, bound to the active site (PDB: 5TYS) (Aoki et al., 2017). This recently published structure was solved in the $P6_1$ space group with a dimer in the asymmetric unit and the highly potent inhibitor (GRL-$142_{published}$) modeled in two conformations with occupancies of 0.53 and 0.47 [**Figure 7.5A**]. The 2.0 Å resolution structure was published with R_{work} and R_{free} at 19.5% and 23.7%, respectively, and an overall model-map correlation coefficient of 0.877. We took the published $P6_1$ data from the PDB and solved the structure with molecular replacement, but refined the model with a twin law applied. In the re-solved structure, the inhibitor (GRL-142_{twin}) was modeled in a single orientation with the “crown-like” tetrahydropyranofuran group interacting with chain A. The R-factors decreased to 17.9 % and 20.5% with improved electron density around the inhibitor [**Figure 7.5B**]. The re-solved structure also exhibited improved protein B-factors, bond lengths, bond angles, and Ramachandran angles [**Table 7.1**]. Overall, the model-map correlation coefficient improved to 0.891.

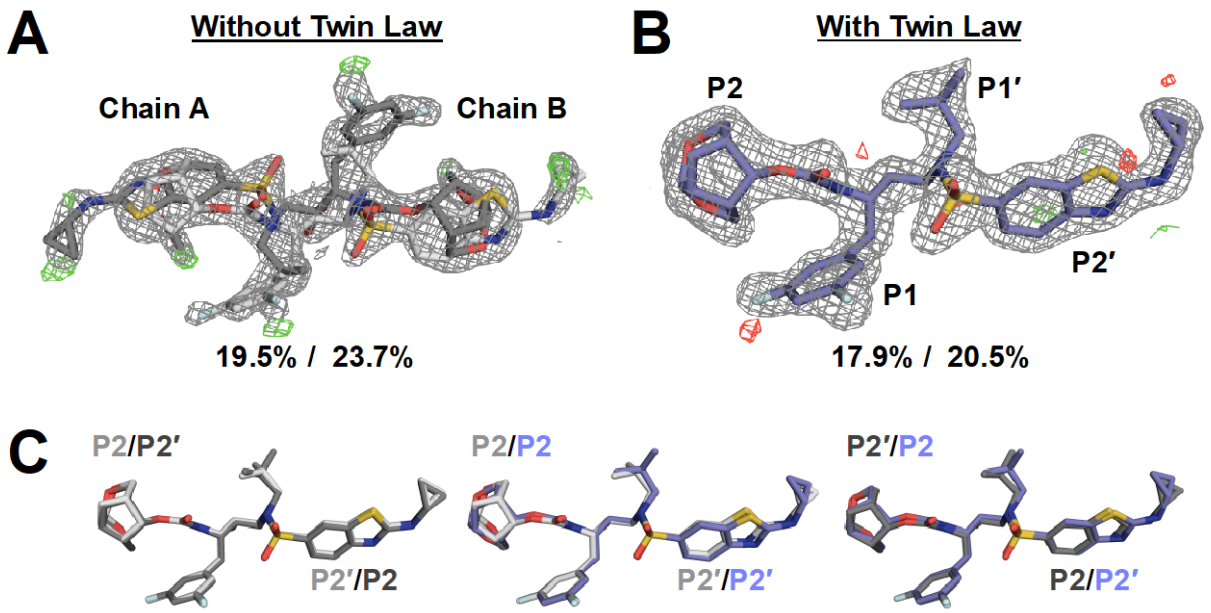


Figure 7.5 A) Electron density maps surrounding the published inhibitor, GRL-142$_{published}$, shown as sticks colored light grey (orientation A) and dark grey (orientation B). B) Electron density maps surrounding the re-solved inhibitor, GRL-142$_{twin}$, shown as sticks colored slate (orientation A). C) Alignment of the two GRL-142$_{published}$ inhibitor orientations and alignment of each GRL-142$_{published}$ orientation against the re-solved inhibitor, GRL-142$_{twin}$.

Table 7.1 X-ray data collection and crystallographic refinement statistics of the published and re-solved structures.

	PUBLISHED	TWIN LAW	PUBLISHED	TWIN LAW
INHIBITOR	GRL-142$_{published}$	GRL-142$_{twin}$	PDI-6$_{published}$	PDI-6$_{twin}$
PDB ID	5TYS	-	4U7Q	-
RESOLUTION (Å)	2.01	2.01	1.70	1.70
SPACE GROUP	$P6_1$	$P6_1$	$P6_1$	$P6_1$
CELL DIMENSIONS:				
***A* (Å)**	63.13	63.13	62.65	62.65
***B* (Å)**	63.13	63.13	62.65	62.65
***C* (Å)**	82.23	82.23	80.12	80.12
A (°)	90	90	90	90
B (°)	90	90	90	90
Γ (°)	120	120	120	120
UNIQUE REFLECTIONS	12,475	12,475	23,062	23,062
REDUNDANCY	10.2 (9.7)	10.2 (9.7)	5.1 (5.0)	5.1 (5.0)
COMPLETENESS (%)	99.9 (100)	99.9 (100)	99.7 (99.8)	99.7 (99.8)
R_{SYM} (%)^A	0.10 (0.46)	0.10 (0.46)	0.045 (0.711)	0.045 (0.711)
AVERAGE I/Σ	26.6 (4.8)	26.6 (4.8)	20.8 (2.57)	20.8 (2.57)
TWIN FRACTION:				
H-TEST	0.477	0.477	0.491	0.491
BRITTON	0.453	0.453	0.452	0.452
TWIN LAW		(h,-h-k,-l)		(h,-h-k,-l)
B-FACTORS:				
WILSON B (Å^2)			31.48	
MEAN B VALUE (Å^2)		27.59	34.9	30.70
PROTEIN (Å^2)	28.36	27.53		30.73
MAIN CHAINS (Å^2)	25.84			
SIDE CHAINS (Å^2)	31.11			
LIGAND (Å^2)	20.05	23.57		32.34
WATERS (Å^2)	34.9			
RMSD^B IN:				
BOND LENGTHS (Å)	0.008	0.003	0.016	0.006
BOND ANGLES (°)	1.022	0.634	2.191	0.884
RAMACHANDRAN PLOT:				
FAVORED (%)	97.94	98.45	95.5	98.97
ALLOWED (%)	2.06	1.55	4.0	1.03
R_{FACTOR} (%)^C	19.49	17.9	27.5	24.1
R_{FREE} (%)^D	23.66	20.5	32.6	27.5

[a]$R_{sym} = \Sigma \mid I - \langle I \rangle \mid / \Sigma I$, where I = observed intensity, $\langle I \rangle$ = average intensity over symmetry equivalent; values in parentheses are for the highest resolution shell.
[b]RMSD, root mean square deviation.
[c]$R_{factor} = \Sigma \mid\mid F_o\mid - \mid F_c\mid\mid / \Sigma \mid F_o\mid$.
[d]R_{free} was calculated from 5% of reflections, chosen randomly, which were omitted from the refinement process.

When aligned, the GRL-142$_{published}$ and GRL-142$_{twin}$ structures highly overlap with subtle differences throughout. To quantify the differences between the inhibitors, pairwise root-mean-square deviation (RMSD) was calculated between common atoms. The RMSD of the two GRL-142$_{published}$ inhibitor orientations was 0.27 Å, while the RMSD between the GRL-142$_{twin}$ inhibitor and published orientations were 0.75 Å and 0.74 Å [**Figure 7.5C**]. The asymmetry between the monomers in each structure (GRL-142$_{published}$ and GRL-142$_{twin}$) was also quantified. A simple alignment of the GRL-142$_{published}$ monomers returned an RMSD of 0.09 Å with nearly identical C_α differences between monomers except one outlier (residue Gly40) [**Figure 7.6**]. The GRL-142$_{twin}$ structure showed slightly larger C_α differences, mainly in the flaps and distal “40’s loop”, with an overall RMSD of 0.13 Å.

This test case showed that structural models from high quality data can be improved with the application of a twin law. Not only did R_{work} and R_{free} decrease by 1.6% and 3.2%, respectively, but the inhibitor could be modeled in a single orientation, known asymmetric structural features were extracted, and the protease structure improved in nearly every stereochemical metric.

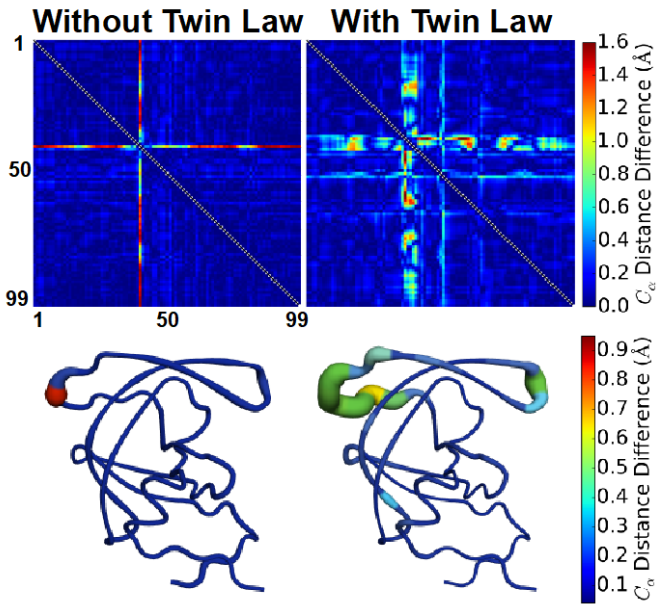


Figure 7.6 Internal distance differences of C_α atoms of chain A compared to chain B of the published GRL-142 bound HIV-1 protease without the twin law (PDB: 5TYS) and of the re-solved structure with the twin law applied. Thicker tubes and warmer colors indicate larger differences.

7.4.3 Re-Solving Structure with a Highly Asymmetric Inhibitor

Lastly, this protocol was applied to another HIV-1 protease cocrystal structure, with a larger and even more asymmetric photolyzable inhibitor (PDI-6) bound to the active site (PDB: 4U7Q) (Schimer et al., 2015). This structure was also solved in the $P6_1$ space group with a dimer in the asymmetric unit and the inhibitor (PDI-$6_{published}$) modeled in two conformations at 50% occupancy [**Figure 7.7A**]. This structure had lower quality data, with R_{work} and R_{free} at 27.5% and 32.6%, respectively, which are uncharacteristically high for a 1.7 Å resolution structure. The overall model-map correlation coefficient was 0.805. We again took the published $P6_1$ data from the PDB and solved the structure with molecular replacement, but refined the model to convergence with a twin law applied. The inhibitor, PDI-6_{twin}, was modeled in a single orientation, with the larger half of the inhibitor interacting with chain A and labeled P3-P1 [**Figure 7.7B**]. The R-factors decreased by several percent to 24.1 % and 27.5%, which significantly improved the electron density around the inhibitor. Again, the protease of PDI-6_{twin} had improved B-factors, bond lengths, bond angles, and Ramachandran angles [**Table 7.1**], while the overall model-map correlation coefficient improved to 0.834.

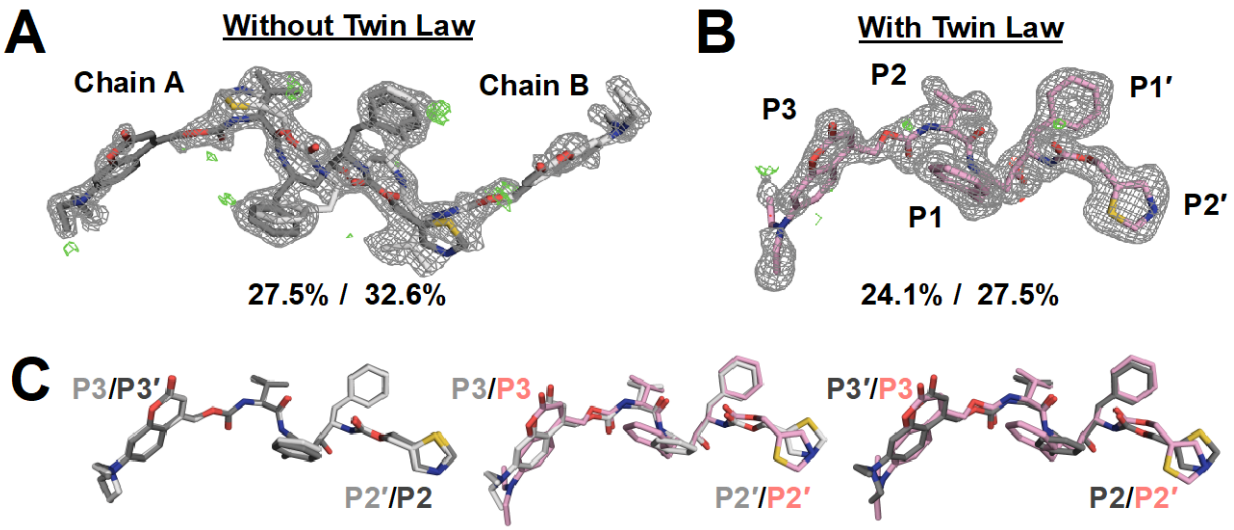


Figure 7.7 A) Electron density maps surrounding the published inhibitor, PDI-6$_{published}$, shown as sticks colored light grey (orientation A) and dark grey (orientation B). B) Electron density maps surrounding the re-solved inhibitor, PDI-6$_{twin}$, shown as sticks colored pink (orientation A). C) Alignment of the two PDI-6$_{published}$ inhibitor orientations and alignment of each PDI-6$_{published}$ orientation against the re-solved inhibitor, PDI-6$_{twin}$.

When aligned, the PDI-$6_{published}$ and PDI-6_{twin} inhibitor structures had some differences throughout, particularly at the P1 and P2′ moieties [**Figure 7.7C**]. Pairwise atomic RMSD of the two PDI-$6_{published}$ inhibitor orientations was 0.93 Å, while the RMSD between the PDI-6_{twin} inhibitor and published orientations were much larger, at 1.48 Å and 1.57 Å. As for the protease, the PDI-$6_{published}$ structure only exhibited differences in internal distances involving the flaps, with an overall RMSD of 0.03 Å between the two monomers [**Figure 7.8**]. The PDI-6_{twin} structure revealed that the twin law allows asymmetric structural features to be resolved throughout the protease, as should be expected due to the extremely asymmetric nature of the inhibitor. The RMSD between the monomers in the PDI-6_{twin} structure was an order of magnitude larger, at 0.33 Å.

This test case demonstrated that structures refined from lower quality data can also improve with the application of a twin law. The PDI-6_{twin} structure had improved electron density around the inhibitor and some moieties exhibited notably different binding conformations. Examining this cocrystal structure highlights that SAR studies based on the published model could be misled.

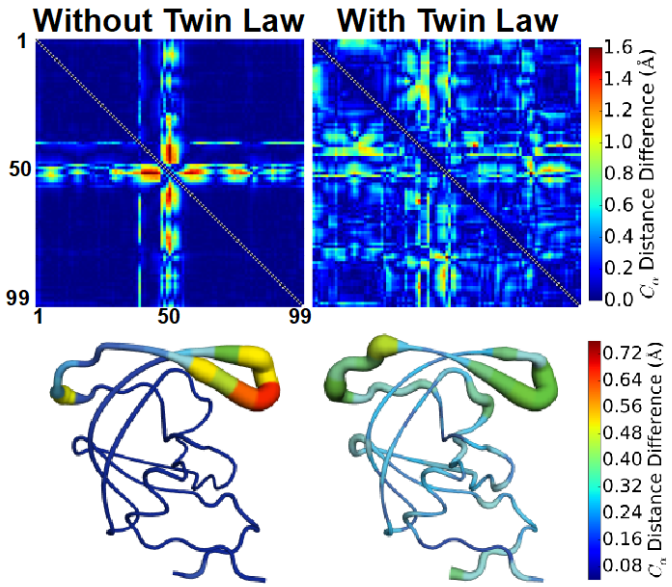


Figure 7.8 Internal distance differences of C_α atoms of chain A compared to chain B of published PDI-6 bound HIV-1 protease without the twin law (PDB: 4U7Q) and of the resolved structure with the twin law applied. Thicker tubes and warmer colors indicate larger differences.

7.5 Conclusion

The application of a twin law improves cocrystal structures of psuedomerohedrally twinned homo-oligomers with inhibitors bound. We demonstrated this for HIV-1 protease where we resolved and refined crystal structures determined in $P6_1$, but this rule should be generally applicable to any potentially asymmetric homo-oligomeric complexes. Refining the crystal structure in the appropriate space group with a twin law results in structures with inhibitors modeled in a single orientation and extracts asymmetric structural differences between the monomers. In addition, the final model has clearer $2F_o$-F_c electron densities, fewer F_o-F_c difference densities, and better statistics overall. This protocol is immediately applicable to about a dozen hexagonal HIV-1 protease structures that are released each year, and may benefit hundreds of other symmetrical homo-oligomeric structures in the PDB. Overall, when applicable, processing data in a lower symmetry space group while applying the relevant twin law will result in better models, therefore allowing for more accurate structural analyses.

8 CHAPTER 8:

Discussion

8.1 Structural and Dynamic Analyses are Essential to Elucidate Complex Mechanisms of Resistance

The prevalence of small molecule inhibitors used as antimicrobial drugs, pesticides, and herbicides has made drug resistance a common phenomenon in many systems. Different organisms have evolved diverse mechanisms of resistance against various drug pressures. Whether in a single- or multi-celled organism, the mechanisms of resistance can be classified as “target-site” or “non-target-site”.[252-253] Target-site resistance includes changes to the target (usually an enzyme) that reduce inhibitor binding affinity or potency. This includes single-site mutations in the target or horizontal gene transfer, as well as improved drug tolerance caused by increasing gene copy numbers and enzyme overexpression.[252-253] Non-target-site resistance includes altering drug uptake and efflux, improved drug sequestration, and chemical alterations to the drug to increase metabolism.[252-253] Our lab and my work focuses on target-site drug resistance, specifically target-site mutations that reduce inhibitor binding affinity.

The epidemic of target-site drug resistance suggests that our current drug design paradigms, focused solely on disrupting the activity of WT targets, are insufficient for long-term efficacy. Greater success against target-site resistance requires a deeper understanding the underlying molecular mechanisms: the structural and dynamic alterations in the enzyme–inhibitor system. Understanding drug resistance in HIV-1 protease, one of the most mutable and genetically diverse enzymes, can shed light on other rapidly evolving disease targets.

Resistance occurs when a mutation causes a reduction in the affinity for an inhibitor with minimal effect on substrate processing. The primary mutations of resistance usually

occur near the active site. More robust inhibitors incorporate strategies to withstand these primary resistance mutations. In Chapter 2, I looked at a few common proximal mutations in HIV-1 protease and how they altered the binding of a small panel of DRV analogues with modifications at the P1′ moiety. Using static structures and employing PCA we were able to extract why the I50V mutation greatly reduced affinity for the DRV analogs while the I84V mutation did not.

Unlike proximal mutations, distal mutations reduce inhibitor affinity by propagating changes to the active site. Analyzing the dynamics of the protease-inhibitor complex is critical for elucidating the mechanism distal mutations use to propagate resistance throughout the protease. Chapter 3 involved quantifying the reduction in potency due to the accumulation of mutations and how distal mutations affected inhibitor binding. This analysis evaluated the potency of one inhibitor (DRV) across many variants. This research showed that distal mutations do not significantly change the size and shape of the active site, but dramatically alter the dynamics. Analyzing which protease-inhibitor interactions were conserved over time indicates where inhibitor modifications could potentially improve potency against resistant variants.

8.2 Elucidating Conserved Static and Dynamic Protein-Inhibitor Interactions is Key to Inhibitor Optimization

The large number of HIV-1 protease resistant variants, generated in lab or in patients, provide many proteases with active sites of different sizes, shapes, and dynamics. These diverse targets allow us to test which inhibitor modifications improve potency against mildly and highly resistant proteases. In Chapter 4, we modified the P2′ moiety of DRV to emulate the optimal interactions of the P2 bis-THF moiety by

accommodating additional hydrogen bonding interactions with NH groups of the protease backbone. That research showed modest improvements against WT and primary resistant mutants but showed greater improvements against highly mutated variants compared to the unmodified DRV.

SBDD involves improving existing scaffolds and exploring new scaffolds. We have learned that designing inhibitors on a static structure does not always translate to success *in vitro* and evaluating the dynamic ensemble better predicts if the protease-inhibitor interactions are maintained over time. In Chapter 5, we made a novel inhibitor scaffold by combining components of three FDA-approved inhibitors. This research showed that the static structures revealed enhanced protease-inhibitor hydrogen bonding interactions, but the dynamics showed the hydrogen bonds were non-optimal, with key interactions lost or at low frequency throughout the simulations.

8.3 Recipe for Robustness Against Resistance

My research on drug resistance in HIV-1 protease and inhibitor design provides a template for designing robust inhibitors against other genetically diverse systems. Inhibitor design is a long, iterative process which requires a multidisciplinary collaboration between biochemists, enzymologists, medicinal chemists, cellular biologists, structural biologists, and computational biologists. Although many of the techniques we utilized are common, it’s how our expertise is combined that creates a product greater than the sum of its parts.

The first step of inhibitor design is to identify a therapeutic target, specifically an enzyme that a small molecule can bind to and inhibit its function to cure or alleviate the disease. Choosing a good target enzyme is not a trivial step. The target must be a

necessary enzyme, preferably with high evolutionary conservation, to minimize resistant heterogeneity, and low homology to host enzymes, to minimize off-target effects. It is also favorable if an enzyme has many substrates to interact with, so that it will have a well-defined substrate envelope and more evolutionary constraints.

After choosing an enzyme, it is important to optimize expression and purification of properly folded, active protein and to develop an activity assay that can also function as a precise inhibition assay. If there are no lead compounds for the target of interest, the inhibition assay should be miniaturized for lead discovery via high-throughput screens. If compounds are found to inhibit the enzyme, you may need to confirm they are competitive inhibitors, assuming that is the design strategy.

Structural analysis of high-resolution enzyme-inhibitor complexes is critical to the success of robust inhibitor design. Not only does structure-based drug design (SBDD) allow planning of inhibitor modifications in three-dimensional space, structures also demonstrate exactly which residues the inhibitor interacts with when bound to the enzyme. It is possible to create highly potent inhibitors through structure activity relationship (SAR) studies alone, but that methodology lacks unambiguous binding information. Even if an unbound (apo) structure of the enzyme is available, numerous studies have shown that inhibitor docking software is unreliable at predicting binding conformations. Also, if the structure of the target enzyme is not known, it can be modeled from a homologous structure *in silico*, but that comes with assumptions and uncertainty.

Knowing enzyme-inhibitor interactions on a per residue level is necessary to design inhibitors that minimize interactions with variable residues and maximize interactions with invariant residues. Residue variability can be inferred by comparing homologous

enzymes which can be mapped onto a structure. If possible, it is best to force resistance by evolutionary selection through sub-optimal drug pressure and to determine heterogeneity through genetic sequencing or protein mass spectrometry. This process is not always performed, but should be a standard experiment if minimizing drug resistance is the ultimate goal. Also, forcing resistance is one of the only ways to identify which mutations, especially distal, contribute to reductions in inhibitor binding.

Lastly, it is extremely important to take the additional step beyond structural analysis into dynamic analysis by utilizing molecular dynamics (MD) simulations. My research has shown that while SBDD is able to generate highly potent inhibitors, it is unable to anticipate dynamic mechanisms of resistance. MD simulations on the protein-inhibitor complexes, across several resistant variants and different inhibitors, elucidate which protein-inhibitor interactions are sustained dynamically over time. Optimizing these conserved interactions increases the probability of maintaining potency against all evolutionarily accessible variants and should therefore improve rates of resistance.

8.4 Future Directions: HTLV-1 Protease

Unlike HIV, which has an abundance of data, HTLV inhibition is in its infancy. Luckily, we can use the HIV model to quickly design a potent protease inhibitor which may improve patient overall survival rates. In Chapter 6, I show that the FDA-approved HIV-1 protease inhibitor DRV can inhibit HTLV-1 protease at 800 nM potency. Evaluating a small panel of inhibitors, we discovered that modifications at two moieties improved potency 27-fold, to 30 nM. If one were to continue this research, they could continue to test additional modifications to the DRV scaffold. A more thorough SAR study would likely improve potency several-fold more. They could also synthesize novel scaffolds that better mimic

HTLV-1 protease's natural substrates. This should create more accurate transition state analogues which may improve potency and would likely reduce susceptibility to resistance.

While there is no resistance data for HTLV-1 protease, we should expect resistance to occur and should design inhibitors accordingly. To orient SBDD with 3D constraints in mind, it would be useful to determine the static substrate envelope for HTLV-1 protease. It would also be useful to thoroughly process sequencing data from patients with HIV/HTLV co-infections who are also taking HIV-1 protease inhibitors/antiretrovirals. This could tell us if drugs like DRV are causing primary resistance mutations. It would also be informative to generate drug resistance mutations via passaging and sequencing. Our collaborators have an assay for generating virus-like particles (VPLs) which allows for a more *in vivo* quantification of if our inhibitors are crossing the cellular membrane and inhibiting HTLV-1 protease. However, there are currently no assays for propagating infectious virus, therefore there can be no evolutionary selection through drug pressure. For now, the active site identity between HIV and HTLV protease can give us clues to what residues should be variant and avoided or invariant and exploited.

8.5 Other Rapidly Evolving Disease Targets that Could Benefit From this Methodology

Although HIV-1 protease is the optimal model system to study resistance, with many variants, inhibitors, structures, and simulations, our methods can likely be applicable to most other rapidly evolving systems. For any given drug target where you would like to improve potency and reduce resistance, you really just need some known inhibitors with inhibition data and a structure of the target in complex with one of the inhibitors. Of course,

it would be best to have a lot of inhibitors with a good range of potencies and many high-resolution structures. With those parameters in mind, our lab has identified some candidates of disease targets that could benefit from our methodologies. Instead of viral targets, we have chosen to explore antibiotic and oncology targets, where inhibition data with corresponding cocrystal structures are common.

One class of targets are dihydrofolate reductases (DHFRs). A variant of this enzyme is present in all organisms as DHFR is essential in the synthesis of purine, thymidine, certain amino acids and thus cell growth. Specific targeting of DHFR has permitted development of chemotherapy agents in oncology such as methotrexate (MTX) and the clinical antibiotic, trimethoprim (TMP). Unfortunately, TMP resistant strains of bacteria have emerged in clinic, due to both mutation and horizontal gene transfer. The extensive variation that exists within DHFR variants provide a rich dataset to unravel molecular mechanisms key to inhibition and resistance [**Figure 8.1**].[254]

WT	34 nM
	Fold
T14A+P26Q	2.8
N23S	1.2
S31F	11.0
F36C	100.0
L65P	22.3
F36C+L65P	397.4
S37T	0.6
A67V	9.1
A67V+C166Y	3.7
R59G+A67V	45.0
V79I	3.6
S106P	1.6
S106P+E127G	1.6
T144A	5.1
T144A+K171E	1.9
D153V	3.0
I158V	6.1
C166Y	4.1
R170F	1.2

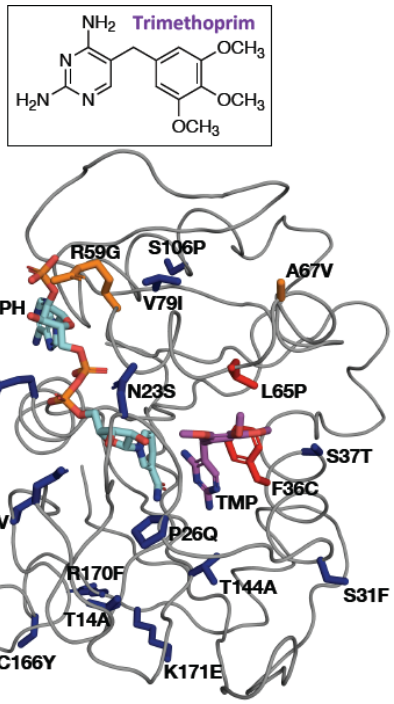
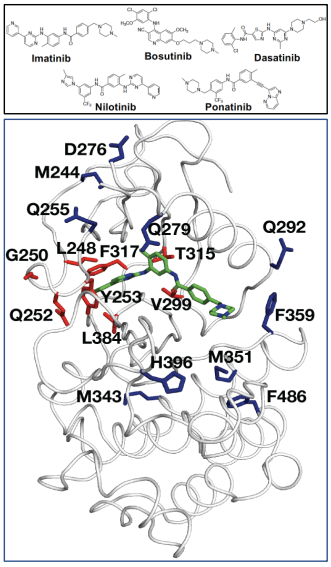


Figure 8.1 Resistant variants of DHFR to TMP with mutations mapped on the structure. Decrease in potency relative to WT (fold-change) is listed and colored according to severity of resistance.[254]

Kinases are another class of enzymes with clinical resistance, targeted in cancer treatment. Drug resistance is a major problem in oncology, with mutations in kinases obliterating effectiveness of chemotherapy. The FDA has approved 48 small molecule kinase inhibitors to date,[255] and 25 of those drugs are tyrosine kinase inhibitors (TKI), a mainstay therapy in many cancers. Approximately 25% of patients need to alter their TKI usage due to molecular resistance. The TKI targets we would like to explore include BCR-ABL and epidermal growth factor receptor (EGFR). Both proteins have many resistance mutations, both proximal and distal [**Figure 8.2**].[256-261] Although the inhibitors are potent (low-nanomolar IC50), resistance still occurs. It's possible that evaluating these resistant variants and designing inhibitors for a dynamic target could improve resistance rates.

	Fold Change (WT = 1)				
	Imatinib	Bosutinib	Dasatinib	Nilotinib	Ponatinib
Inhibitor	Type 2	Type 1	Type 1	Type 2	Type 2
WT IC50 nM	527.0	41.6	1.8	17.7	2.1
WT	1	1	1	1	1
L248R	14.6	22.9	12.5	30.2	6.2
L248V	3.5	3.5	5.1	2.8	3.4
G250E	6.9	4.3	4.4	4.6	6.0
Q252H	1.4	0.8	3.1	2.6	6.1
Y253F	3.6	1.0	1.6	3.2	3.7
Y253H	8.7	0.6	2.6	26.8	2.6
V299L	1.5	26.1	8.7	1.3	0.6
T315A	1.7	6.0	58.9	2.7	0.4
T315I	17.5	45.4	75.0	39.4	3.0
T315V	12.2	29.3	738.8	57.0	2.1
F317L	2.6	2.4	4.5	2.2	0.7
F317R	2.3	33.5	114.8	2.3	4.9
F317V	0.4	11.5	21.3	0.5	2.3
L384M	1.3	0.5	2.2	2.3	2.2
L248R+F359I	11.7	39.3	13.7	96.2	17.7
M244V	0.9	0.9	2.0	1.2	3.2
E255K	6.0	9.5	5.6	6.7	8.4
E255V	17.0	5.5	3.4	10.3	12.9
D276G	2.2	0.6	1.4	2.0	2.1
E279K	3.6	1.0	1.6	2.0	3.0
E292L	0.7	1.1	1.3	1.8	2.0
M343T	1.2	1.1	0.9	0.8	0.9
M351T	1.8	0.7	0.9	0.4	1.2
F359I	6.0	2.9	3.0	16.3	2.9
F359V	2.9	0.9	1.5	5.2	4.4
H396P	2.4	0.4	1.1	2.4	1.4
H396R	3.9	0.8	1.6	3.1	5.9
F486S	8.1	2.3	3.0	1.9	2.1
PDB Code	2HYY	3UE4	4XEY	3CS9	3IK3



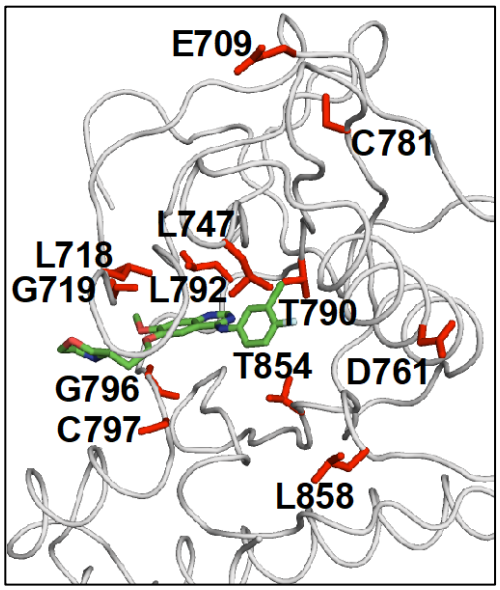


Figure 8.2 Left) BCR-ABL mutations and resistance to FDA-approved inhibitors. Mapped onto BRC-ABL structure (PDB: 2HYY) with proximal and distal mutations shown as red and blue sticks.[256] Right) EGFR kinase domain with resistance mutation sites shown as sticks and gefitinib bound to the active site (green) (PDB: 2ITY).[257-261]

8.6 The Future of SBDD

Structures are integral to drug design (SBDD) and for reliable MD simulations, which are critical for the methodologies our lab promotes. Every structural determination method has its pros, cons, and limitations. X-ray crystallography is well established with over 140,000 structures in the PDB, but some proteins do not crystallize. Solution state NMR has around 13,000 structures, but they are often very small proteins or short strands of DNA/RNA, which rarely have a ligand bound. Cryo Electron Microscopy (CryoEM) is an emerging technology with about 4,500 published structures. CryoEM was limited to very large proteins at low resolution, but has undergone a “resolution revolution” over the past few years and is constantly improving.

The latest structural technique is called electron crystallography, which blends electron microscopy and protein crystallography. Because this technique uses protein crystals that are a billionth the size needed for X-ray crystallography, it is also known as microcrystal electron diffraction (MicroED). Although this structural determination method still requires crystallization of target protein, when done correctly, MicroED can generate high resolution (< 3.0 Å) structures in a very short amount of time. Since 2013, there are only about 100 structures on the PDB solved by MicroED, but like CryoEM, the more structures that are derived from this technique, the more it will be utilized.

8.7 Concluding Remarks

Drug resistance remains a significant threat and the currently employed drug design techniques are not equipped to solve this problem. Although resistance will occur any time evolution exists under the selective pressure of an imperfect inhibitor in a heterogeneous population, this research outlines the types of analyses necessary to

design more robust inhibitors. More specifically, this thesis outlines methodologies to diagnose various mechanisms of drug resistance and how to identify and optimize critical enzyme-inhibitor interactions. While this thesis focused on the ideal HIV-1 protease model system, our protocols are compatible with nearly any drug target. For example, we demonstrated how quickly we could improve potency against HTLV-1 protease, with many ideas for future optimization. Hopefully this thesis work directly contributes to the development of a HTLV-1 drug which will improve overall survival for those suffering from leukemia/lymphoma. More broadly, I hope these strategies guide future drug design efforts that will one day lead to more robust therapeutics for a variety of rapidly evolving diseases.

REFERENCES

1. World Health Organization. Fact Sheet: Antimicrobial Resistance. **2018**.
2. Editors, P. M., Antimicrobial Resistance: Is the World UNprepared? *PLoS Med.* **2016,** *13* (9), e1002130.
3. O'Neill, J. Antimicrobial Resistance: Tackling a crisis for the health and wealth of nations. http://amr-review.org/sites/default/files/AMR%20Review%20Paper%20-%20Tackling%20a%20crisis%20for%20the%20health%20and%20wealth%20of%20nations_1.pdf (accessed July 9, 2016).
4. Collier, P.; O'Neill, L. J., Two years on: an update on achievement towards the recommendations of the antimicrobial resistance report. *J. Appl. Microbiol.* **2018,** *125* (2), 308-312.
5. Korber, B.; Gaschen, B.; Yusim, K.; Thakallapally, R.; Kesmir, C.; Detours, V., Evolutionary and immunological implications of contemporary HIV-1 variation. *Br. Med. Bull.* **2001,** *58*, 19-42.
6. Ho, D. D.; Neumann, A. U.; Perelson, A. S.; Chen, W.; Leonard, J. M.; Markowitz, M., Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection. *Nature* **1995,** *373* (6510), 123-6.
7. Wei, X.; Ghosh, S. K.; Taylor, M. E.; Johnson, V. A.; Emini, E. A.; Deutsch, P.; Lifson, J. D.; Bonhoeffer, S.; Nowak, M. A.; Hahn, B. H.; et al., Viral dynamics in human immunodeficiency virus type 1 infection. *Nature* **1995,** *373* (6510), 117-22.
8. Perelson, A. S.; Neumann, A. U.; Markowitz, M.; Leonard, J. M.; Ho, D. D., HIV-1 dynamics in vivo: virion clearance rate, infected cell life-span, and viral generation time. *Science* **1996,** *271* (5255), 1582-6.
9. Abram, M. E.; Ferris, A. L.; Shao, W.; Alvord, W. G.; Hughes, S. H., Nature, position, and frequency of mutations made in a single cycle of HIV-1 replication. *J. Virol.* **2010,** *84* (19), 9864-78.
10. Mansky, L. M.; Temin, H. M., Lower in vivo mutation rate of human immunodeficiency virus type 1 than that predicted from the fidelity of purified reverse transcriptase. *J. Virol.* **1995,** *69* (8), 5087-94.
11. Onafuwa-Nuga, A.; Telesnitsky, A., The remarkable frequency of human immunodeficiency virus type 1 genetic recombination. *Microbiol. Mol. Biol. Rev.* **2009,** *73* (3), 451-80, Table of Contents.
12. Sadler, H. A.; Stenglein, M. D.; Harris, R. S.; Mansky, L. M., APOBEC3G contributes to HIV-1 variation through sublethal mutagenesis. *J. Virol.* **2010,** *84* (14), 7396-404.
13. Vajpayee, M.; Mohan, T., Current practices in laboratory monitoring of HIV infection. *Indian J. Med. Res.* **2011,** *134* (6), 801-22.
14. Noguera-Julian, M.; Edgil, D.; Harrigan, P. R.; Sandstrom, P.; Godfrey, C.; Paredes, R., Next-Generation Human Immunodeficiency Virus Sequencing for Patient Management and Drug Resistance Surveillance. *J. Infect. Dis.* **2017,** *216* (suppl_9), S829-S833.
15. Rhee, S. Y.; Gonzales, M. J.; Kantor, R.; Betts, B. J.; Ravela, J.; Shafer, R. W., Human immunodeficiency virus reverse transcriptase and protease sequence database. *Nucleic Acids Res.* **2003,** *31* (1), 298-303.

16. Shafer, R. W., Rationale and uses of a public HIV drug-resistance database. *J. Infect. Dis.* **2006,** *194 Suppl 1*, S51-8.
17. Wensing, A. M.; Calvez, V.; Ceccherini-Silberstein, F.; Charpentier, C.; Gunthard, H. F.; Paredes, R.; Shafer, R. W.; Richman, D. D., 2019 update of the drug resistance mutations in HIV-1. *Top. Antivir. Med.* **2019,** *27* (3), 111-121.
18. Meek, T. D.; Dayton, B. D.; Metcalf, B. W.; Dreyer, G. B.; Strickler, J. E.; Gorniak, J. G.; Rosenberg, M.; Moore, M. L.; Magaard, V. W.; Debouck, C., Human immunodeficiency virus 1 protease expressed in Escherichia coli behaves as a dimeric aspartic protease. *Proc. Natl. Acad. Sci. U. S. A.* **1989,** *86* (6), 1841-5.
19. Wlodawer, A.; Miller, M.; Jaskolski, M.; Sathyanarayana, B. K.; Baldwin, E.; Weber, I. T.; Selk, L. M.; Clawson, L.; Schneider, J.; Kent, S. B., Conserved folding in retroviral proteases: crystal structure of a synthetic HIV-1 protease. *Science* **1989,** *245* (4918), 616-21.
20. Navia, M. A.; Fitzgerald, P. M.; McKeever, B. M.; Leu, C. T.; Heimbach, J. C.; Herber, W. K.; Sigal, I. S.; Darke, P. L.; Springer, J. P., Three-dimensional structure of aspartyl protease from human immunodeficiency virus HIV-1. *Nature* **1989,** *337* (6208), 615-20.
21. Kurt Yilmaz, N.; Swanstrom, R.; Schiffer, C. A., Improving Viral Protease Inhibitors to Counter Drug Resistance. *Trends Microbiol.* **2016,** *24* (7), 547-557.
22. Brik, A.; Wong, C. H., HIV-1 protease: mechanism and drug discovery. *Org. Biomol. Chem.* **2003,** *1* (1), 5-14.
23. Hyland, L. J.; Tomaszek, T. A., Jr.; Roberts, G. D.; Carr, S. A.; Magaard, V. W.; Bryan, H. L.; Fakhoury, S. A.; Moore, M. L.; Minnich, M. D.; Culp, J. S.; et al., Human immunodeficiency virus-1 protease. 1. Initial velocity studies and kinetic characterization of reaction intermediates by 18O isotope exchange. *Biochemistry* **1991,** *30* (34), 8441-53.
24. Szeltner, Z.; Polgar, L., Rate-determining steps in HIV-1 protease catalysis. The hydrolysis of the most specific substrate. *J. Biol. Chem.* **1996,** *271* (50), 32180-4.
25. Arts, E. J.; Hazuda, D. J., HIV-1 antiretroviral drug therapy. *Cold Spring Harb. Perspect. Med.* **2012,** *2* (4), a007161.
26. Palmisano, L.; Vella, S., A brief history of antiretroviral therapy of HIV infection: success and challenges. *Ann. Ist. Super. Sanita* **2011,** *47* (1), 44-8.
27. Ali, A.; Bandaranayake, R. M.; Cai, Y.; King, N. M.; Kolli, M.; Mittal, S.; Murzycki, J. F.; Nalam, M. N.; Nalivaika, E. A.; Ozen, A.; Prabu-Jeyabalan, M. M.; Thayer, K.; Schiffer, C. A., Molecular Basis for Drug Resistance in HIV-1 Protease. *Viruses* **2010,** *2* (11), 2509-35.
28. Organization, W. H., Updated recommendations on first-line and second-line antiretroviral regimens and post-exposure prophylaxis and recommendations on early infant diagnosis of HIV: interim guidance. **2018**.
29. Back, D.; Sekar, V.; Hoetelmans, R. M., Darunavir: pharmacokinetics and drug interactions. *Antivir. Ther.* **2008,** *13* (1), 1-13.
30. Wensing, A. M.; van Maarseveen, N. M.; Nijhuis, M., Fifteen years of HIV Protease Inhibitors: raising the barrier to resistance. *Antiviral Res.* **2010,** *85* (1), 59-74.
31. Saskova, K. G.; Kozisek, M.; Rezacova, P.; Brynda, J.; Yashina, T.; Kagan, R. M.; Konvalinka, J., Molecular characterization of clinical isolates of human immunodeficiency virus resistant to the protease inhibitor darunavir. *J. Virol.* **2009,** *83* (17), 8810-8.

32. Kozisek, M.; Lepsik, M.; Grantz Saskova, K.; Brynda, J.; Konvalinka, J.; Rezacova, P., Thermodynamic and structural analysis of HIV protease resistance to darunavir - analysis of heavily mutated patient-derived HIV-1 proteases. *FEBS J.* **2014,** *281* (7), 1834-47.
33. Agniswamy, J.; Shen, C. H.; Aniana, A.; Sayer, J. M.; Louis, J. M.; Weber, I. T., HIV-1 protease with 20 mutations exhibits extreme resistance to clinical inhibitors through coordinated structural rearrangements. *Biochemistry* **2012,** *51* (13), 2819-28.
34. Nakashima, M.; Ode, H.; Suzuki, K.; Fujino, M.; Maejima, M.; Kimura, Y.; Masaoka, T.; Hattori, J.; Matsuda, M.; Hachiya, A.; Yokomaku, Y.; Suzuki, A.; Watanabe, N.; Sugiura, W.; Iwatani, Y., Unique Flap Conformation in an HIV-1 Protease with High-Level Darunavir Resistance. *Front. Microbiol.* **2016,** *7*, 61.
35. Yedidi, R. S.; Garimella, H.; Aoki, M.; Aoki-Ogata, H.; Desai, D. V.; Chang, S. B.; Davis, D. A.; Fyvie, W. S.; Kaufman, J. D.; Smith, D. W.; Das, D.; Wingfield, P. T.; Maeda, K.; Ghosh, A. K.; Mitsuya, H., A conserved hydrogen-bonding network of P2 bis-tetrahydrofuran-containing HIV-1 protease inhibitors (PIs) with a protease active-site amino acid backbone aids in their activity against PI-resistant HIV. *Antimicrob. Agents Chemother.* **2014,** *58* (7), 3679-88.
36. Agniswamy, J.; Shen, C. H.; Wang, Y. F.; Ghosh, A. K.; Rao, K. V.; Xu, C. X.; Sayer, J. M.; Louis, J. M.; Weber, I. T., Extreme multidrug resistant HIV-1 protease with 20 mutations is resistant to novel protease inhibitors with P1'-pyrrolidinone or P2-tris-tetrahydrofuran. *J. Med. Chem.* **2013,** *56* (10), 4017-27.
37. Lockbaum, G. J.; Leidner, F.; Rusere, L. N.; Henes, M.; Kosovrasti, K.; Nachum, G. S.; Nalivaika, E. A.; Ali, A.; Yilmaz, N. K.; Schiffer, C. A., Structural Adaptation of Darunavir Analogues against Primary Mutations in HIV-1 Protease. *ACS Infect Dis* **2019,** *5* (2), 316-325.
38. Rhee, S. Y.; Taylor, J.; Fessel, W. J.; Kaufman, D.; Towner, W.; Troia, P.; Ruane, P.; Hellinger, J.; Shirvani, V.; Zolopa, A.; Shafer, R. W., HIV-1 protease mutations and protease inhibitor cross-resistance. *Antimicrob. Agents Chemother.* **2010,** *54* (10), 4253-61.
39. Henderson, G. J.; Lee, S. K.; Irlbeck, D. M.; Harris, J.; Kline, M.; Pollom, E.; Parkin, N.; Swanstrom, R., Interplay between single resistance-associated mutations in the HIV-1 protease and viral infectivity, protease activity, and inhibitor sensitivity. *Antimicrob. Agents Chemother.* **2012,** *56* (2), 623-33.
40. Ozen, A.; Lin, K. H.; Kurt Yilmaz, N.; Schiffer, C. A., Structural basis and distal effects of Gag substrate coevolution in drug resistance to HIV-1 protease. *Proc. Natl. Acad. Sci. U. S. A.* **2014,** *111* (45), 15993-8.
41. Tozser, J.; Blaha, I.; Copeland, T. D.; Wondrak, E. M.; Oroszlan, S., Comparison of the HIV-1 and HIV-2 proteinases using oligopeptide substrates representing cleavage sites in Gag and Gag-Pol polyproteins. *FEBS Lett.* **1991,** *281* (1-2), 77-80.
42. Konnyu, B.; Sadiq, S. K.; Turanyi, T.; Hirmondo, R.; Muller, B.; Krausslich, H. G.; Coveney, P. V.; Muller, V., Gag-Pol processing during HIV-1 virion maturation: a systems biology approach. *PLoS Comput. Biol.* **2013,** *9* (6), e1003103.
43. Prabu-Jeyabalan, M.; Nalivaika, E.; Schiffer, C. A., Substrate shape determines specificity of recognition for HIV-1 protease: analysis of crystal structures of six substrate complexes. *Structure* **2002,** *10* (3), 369-81.
44. Lefebvre, E.; Schiffer, C. A., Resilience to resistance of HIV-1 protease inhibitors: profile of darunavir. *AIDS Rev.* **2008,** *10* (3), 131-42.

45. Nalam, M. N.; Ali, A.; Reddy, G. S.; Cao, H.; Anjum, S. G.; Altman, M. D.; Yilmaz, N. K.; Tidor, B.; Rana, T. M.; Schiffer, C. A., Substrate envelope-designed potent HIV-1 protease inhibitors to avoid drug resistance. *Chem Biol* **2013,** *20* (9), 1116-24.
46. Lockbaum, G. J.; Leidner, F.; Royer, W. E.; Kurt Yilmaz, N.; Schiffer, C. A., Optimizing the refinement of merohedrally twinned P61 HIV-1 protease-inhibitor cocrystal structures. *Acta Crystallogr D Struct Biol* **2020,** *76* (Pt 3), 302-310.
47. Rose, R. B.; Craik, C. S.; Stroud, R. M., Domain flexibility in retroviral proteases: structural implications for drug resistant mutations. *Biochemistry* **1998,** *37* (8), 2607-21.
48. Prabu-Jeyabalan, M.; Nalivaika, E.; Schiffer, C. A., How does a symmetric dimer recognize an asymmetric substrate? A substrate complex of HIV-1 protease. *J. Mol. Biol.* **2000,** *301* (5), 1207-20.
49. Erickson, J.; Neidhart, D. J.; VanDrie, J.; Kempf, D. J.; Wang, X. C.; Norbeck, D. W.; Plattner, J. J.; Rittenhouse, J. W.; Turon, M.; Wideburg, N.; et al., Design, activity, and 2.8 A crystal structure of a C2 symmetric inhibitor complexed to HIV-1 protease. *Science* **1990,** *249* (4968), 527-33.
50. Dreyer, G. B.; Lambert, D. M.; Meek, T. D.; Carr, T. J.; Tomaszek, T. A., Jr.; Fernandez, A. V.; Bartus, H.; Cacciavillani, E.; Hassell, A. M.; Minnich, M.; et al., Hydroxyethylene isostere inhibitors of human immunodeficiency virus-1 protease: structure-activity analysis using enzyme kinetics, X-ray crystallography, and infected T-cell assays. *Biochemistry* **1992,** *31* (29), 6646-59.
51. Cai, Y.; Yilmaz, N. K.; Myint, W.; Ishima, R.; Schiffer, C. A., 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.* **2012,** *8* (10), 3452-3462.
52. Foulkes-Murzycki, J. E.; Scott, W. R.; Schiffer, C. A., Hydrophobic sliding: a possible mechanism for drug resistance in human immunodeficiency virus type 1 protease. *Structure* **2007,** *15* (2), 225-33.
53. Mittal, S.; Cai, Y.; Nalam, M. N.; Bolon, D. N.; Schiffer, C. A., Hydrophobic core flexibility modulates enzyme activity in HIV-1 protease. *J. Am. Chem. Soc.* **2012,** *134* (9), 4163-8.
54. Ragland, D. A.; Nalivaika, E. A.; Nalam, M. N.; Prachanronarong, K. L.; Cao, H.; Bandaranayake, R. M.; Cai, Y.; Kurt-Yilmaz, N.; Schiffer, C. A., Drug resistance conferred by mutations outside the active site through alterations in the dynamic and structural ensemble of HIV-1 protease. *J. Am. Chem. Soc.* **2014,** *136* (34), 11956-63.
55. Paulsen, J. L.; Leidner, F.; Ragland, D. A.; Kurt Yilmaz, N.; Schiffer, C. A., Interdependence of Inhibitor Recognition in HIV-1 Protease. *J Chem Theory Comput* **2017,** *13* (5), 2300-2309.
56. Poiesz, B. J.; Ruscetti, F. W.; Gazdar, A. F.; Bunn, P. A.; Minna, J. D.; Gallo, R. C., Detection and isolation of type C retrovirus particles from fresh and cultured lymphocytes of a patient with cutaneous T-cell lymphoma. *Proc. Natl. Acad. Sci. U. S. A.* **1980,** *77* (12), 7415-9.
57. Kalyanaraman, V. S.; Sarngadharan, M. G.; Robert-Guroff, M.; Miyoshi, I.; Golde, D.; Gallo, R. C., A new subtype of human T-cell leukemia virus (HTLV-II) associated with a T-cell variant of hairy cell leukemia. *Science* **1982,** *218* (4572), 571-3.
58. Beilke, M. A., Retroviral coinfections: HIV and HTLV: taking stock of more than a quarter century of research. *AIDS Res. Hum. Retroviruses* **2012,** *28* (2), 139-47.
59. Willems, L.; Hasegawa, H.; Accolla, R.; Bangham, C.; Bazarbachi, A.; Bertazzoni, U.; Carneiro-Proietti, A. B.; Cheng, H.; Chieco-Bianchi, L.; Ciminale, V.; Coelho-Dos-Reis, J.; Esparza,

J.; Gallo, R. C.; Gessain, A.; Gotuzzo, E.; Hall, W.; Harford, J.; Hermine, O.; Jacobson, S.; Macchi, B.; Macpherson, C.; Mahieux, R.; Matsuoka, M.; Murphy, E.; Peloponese, J. M.; Simon, V.; Tagaya, Y.; Taylor, G. P.; Watanabe, T.; Yamano, Y., Reducing the global burden of HTLV-1 infection: An agenda for research and action. *Antiviral Res.* **2017,** *137*, 41-48.
60. Bazarbachi, A.; Hermine, O., Treatment of adult T-cell leukaemia/lymphoma: current strategy and future perspectives. *Virus Res.* **2001,** *78* (1-2), 79-92.
61. Nasr, R.; El-Sabban, M. E.; Karam, J. A.; Dbaibo, G.; Kfoury, Y.; Arnulf, B.; Lepelletier, Y.; Bex, F.; de The, H.; Hermine, O.; Bazarbachi, A., Efficacy and mechanism of action of the proteasome inhibitor PS-341 in T-cell lymphomas and HTLV-I associated adult T-cell leukemia/lymphoma. *Oncogene* **2005,** *24* (3), 419-30.
62. Tozser, J.; Zahuczky, G.; Bagossi, P.; Louis, J. M.; Copeland, T. D.; Oroszlan, S.; Harrison, R. W.; Weber, I. T., Comparison of the substrate specificity of the human T-cell leukemia virus and human immunodeficiency virus proteinases. *Eur. J. Biochem.* **2000,** *267* (20), 6287-95.
63. Kadas, J.; Weber, I. T.; Bagossi, P.; Miklossy, G.; Boross, P.; Oroszlan, S.; Tozser, J., Narrow substrate specificity and sensitivity toward ligand-binding site mutations of human T-cell Leukemia virus type 1 protease. *J. Biol. Chem.* **2004,** *279* (26), 27148-57.
64. Shuker, S. B.; Mariani, V. L.; Herger, B. E.; Dennison, K. J., Understanding HTLV-I protease. *Chem. Biol.* **2003,** *10* (5), 373-80.
65. Li, M.; Laco, G. S.; Jaskolski, M.; Rozycki, J.; Alexandratos, J.; Wlodawer, A.; Gustchina, A., Crystal structure of human T cell leukemia virus protease, a novel target for anticancer drug design. *Proc. Natl. Acad. Sci. U. S. A.* **2005,** *102* (51), 18332-7.
66. Naka, H.; Teruya, K.; Bang, J. K.; Aimoto, S.; Tatsumi, T.; Konno, H.; Nosaka, K.; Akaji, K., Evaluations of substrate specificity and inhibition at PR/p3 cleavage site of HTLV-1 protease. *Bioorg. Med. Chem. Lett.* **2006,** *16* (14), 3761-4.
67. Nguyen, J. T.; Kato, K.; Hidaka, K.; Kumada, H. O.; Kimura, T.; Kiso, Y., Design and synthesis of several small-size HTLV-I protease inhibitors with different hydrophilicity profiles. *Bioorg. Med. Chem. Lett.* **2011,** *21* (8), 2425-9.
68. Kuhnert, M.; Steuber, H.; Diederich, W. E., Structural basis for HTLV-1 protease inhibition by the HIV-1 protease inhibitor indinavir. *J. Med. Chem.* **2014,** *57* (14), 6266-72.
69. Berman, H.; Henrick, K.; Nakamura, H., Announcing the worldwide Protein Data Bank. *Nat. Struct. Biol.* **2003,** *10* (12), 980.
70. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E., The Protein Data Bank. *Nucleic Acids Res.* **2000,** *28* (1), 235-42.
71. World Health Organization. World health statistics 2016: monitoring health for the SDGs, sustainable development goals. **2016**, 136.
72. Agniswamy, J.; Louis, J. M.; Roche, J.; Harrison, R. W.; Weber, I. T., Structural Studies of a Rationally Selected Multi-Drug Resistant HIV-1 Protease Reveal Synergistic Effect of Distal Mutations on Flap Dynamics. *PLoS One* **2016,** *11* (12), e0168616.
73. Ghosh, A. K.; Osswald, H. L.; Prato, G., Recent Progress in the Development of HIV-1 Protease Inhibitors for the Treatment of HIV/AIDS. *J. Med. Chem.* **2016,** *59* (11), 5172-208.
74. de Meyer, S.; Vangeneugden, T.; van Baelen, B.; de Paepe, E.; van Marck, H.; Picchio, G.; Lefebvre, E.; de Bethune, M. P., Resistance profile of darunavir: combined 24-week results from the POWER trials. *AIDS Res. Hum. Retroviruses* **2008,** *24* (3), 379-88.

75. Weber, I. T.; Kneller, D. W.; Wong-Sam, A., Highly resistant HIV-1 proteases and strategies for their inhibition. *Future Med. Chem.* **2015,** *7* (8), 1023-38.
76. Nijhuis, M.; Schuurman, R.; de Jong, D.; Erickson, J.; Gustchina, E.; Albert, J.; Schipper, P.; Gulnik, S.; Boucher, C. A., Increased fitness of drug resistant HIV-1 protease as a result of acquisition of compensatory mutations during suboptimal therapy. *AIDS* **1999,** *13* (17), 2349-59.
77. Mittal, S.; Bandaranayake, R. M.; King, N. M.; Prabu-Jeyabalan, M.; Nalam, M. N.; Nalivaika, E. A.; Yilmaz, N. K.; Schiffer, C. A., Structural and thermodynamic basis of amprenavir/darunavir and atazanavir resistance in HIV-1 protease with mutations at residue 50. *J. Virol.* **2013,** *87* (8), 4176-84.
78. Surleraux, D. L.; Tahri, A.; Verschueren, W. G.; Pille, G. M.; de Kock, H. A.; Jonckers, T. H.; Peeters, A.; De Meyer, S.; Azijn, H.; Pauwels, R.; de Bethune, M. P.; King, N. M.; Prabu-Jeyabalan, M.; Schiffer, C. A.; Wigerinck, P. B., Discovery and selection of TMC114, a next generation HIV-1 protease inhibitor. *J. Med. Chem.* **2005,** *48* (6), 1813-22.
79. Ghosh, A. K.; Parham, G. L.; Martyr, C. D.; Nyalapatla, P. R.; Osswald, H. L.; Agniswamy, J.; Wang, Y. F.; Amano, M.; Weber, I. T.; Mitsuya, H., Highly potent HIV-1 protease inhibitors with novel tricyclic P2 ligands: design, synthesis, and protein-ligand X-ray studies. *J. Med. Chem.* **2013,** *56* (17), 6792-802.
80. Ghosh, A. K.; P, R. N.; Kovela, S.; Rao, K. V.; Brindisi, M.; Osswald, H. L.; Amano, M.; Aoki, M.; Agniswamy, J.; Wang, Y. F.; Weber, I. T.; Mitsuya, H., Design and Synthesis of Highly Potent HIV-1 Protease Inhibitors Containing Tricyclic Fused Ring Systems as Novel P2 Ligands: Structure-Activity Studies, Biological and X-ray Structural Analysis. *J. Med. Chem.* **2018,** *61* (10), 4561-4577.
81. Ghosh, A. K.; Fyvie, W. S.; Brindisi, M.; Steffey, M.; Agniswamy, J.; Wang, Y. F.; Aoki, M.; Amano, M.; Weber, I. T.; Mitsuya, H., Design, Synthesis, Biological Evaluation, and X-ray Studies of HIV-1 Protease Inhibitors with Modified P2' Ligands of Darunavir. *ChemMedChem* **2017,** *12* (23), 1942-1952.
82. Amano, M.; Salcedo-Gomez, P. M.; Zhao, R.; Yedidi, R. S.; Das, D.; Bulut, H.; Delino, N. S.; Sheri, V. R.; Ghosh, A. K.; Mitsuya, H., A Modified P1 Moiety Enhances In Vitro Antiviral Activity against Various Multidrug-Resistant HIV-1 Variants and In Vitro Central Nervous System Penetration Properties of a Novel Nonpeptidic Protease Inhibitor, GRL-10413. *Antimicrob. Agents Chemother.* **2016,** *60* (12), 7046-7059.
83. Hohlfeld, K.; Tomassi, C.; Wegner, J. K.; Kesteleyn, B.; Linclau, B., Disubstituted Bis-THF Moieties as New P2 Ligands in Nonpeptidal HIV-1 Protease Inhibitors. *ACS Med. Chem. Lett.* **2011,** *2* (6), 461-5.
84. Cai, Y.; Schiffer, C. A., Decomposing the energetic impact of drug resistant mutations in HIV-1 protease on binding DRV. *J. Chem. Theory Comput.* **2010,** *6* (4), 1358-1368.
85. Leidner, F.; Kurt Yilmaz, N.; Paulsen, J.; Muller, Y. A.; Schiffer, C. A., Hydration Structure and Dynamics of Inhibitor-Bound HIV-1 Protease. *J Chem Theory Comput* **2018,** *14* (5), 2784-2796.
86. Khan, S. N.; Persons, J. D.; Paulsen, J. L.; Guerrero, M.; Schiffer, C. A.; Kurt-Yilmaz, N.; Ishima, R., Probing Structural Changes among Analogous Inhibitor-Bound Forms of HIV-1 Protease and a Drug-Resistant Mutant in Solution by Nuclear Magnetic Resonance. *Biochemistry* **2018,** *57* (10), 1652-1662.

87. King, N. M.; Prabu-Jeyabalan, M.; Nalivaika, E. A.; Schiffer, C. A., Combating susceptibility to drug resistance: lessons from HIV-1 protease. *Chem. Biol.* **2004,** *11* (10), 1333-8.
88. Johnson, V. A.; Calvez, V.; Gunthard, H. F.; Paredes, R.; Pillay, D.; Shafer, R.; Wensing, A. M.; Richman, D. D., 2011 update of the drug resistance mutations in HIV-1. *Top. Antivir. Med.* **2011,** *19* (4), 156-64.
89. Matayoshi, E. D.; Wang, G. T.; Krafft, G. A.; Erickson, J., Novel fluorogenic substrates for assaying retroviral proteases by resonance energy transfer. *Science* **1990,** *247* (4945), 954-8.
90. Windsor, I. W.; Raines, R. T., Fluorogenic Assay for Inhibitors of HIV-1 Protease with Sub-picomolar Affinity. *Sci Rep* **2015,** *5*, 11286.
91. Meher, B. R.; Wang, Y., Interaction of I50V mutant and I50L/A71V double mutant HIV-protease with inhibitor TMC114 (darunavir): molecular dynamics simulation and binding free energy studies. *J Phys Chem B* **2012,** *116* (6), 1884-900.
92. Foulkes, J. E.; Prabu-Jeyabalan, M.; Cooper, D.; Henderson, G. J.; Harris, J.; Swanstrom, R.; Schiffer, C. A., Role of invariant Thr80 in human immunodeficiency virus type 1 protease structure, function, and viral infectivity. *J. Virol.* **2006,** *80* (14), 6906-16.
93. Freedberg, D. I.; Ishima, R.; Jacob, J.; Wang, Y. X.; Kustanovich, I.; Louis, J. M.; Torchia, D. A., Rapid structural fluctuations of the free HIV protease flaps in solution: relationship to crystal structures and comparison with predictions of dynamics calculations. *Protein Sci.* **2002,** *11* (2), 221-32.
94. Leonis, G.; Czyznikowska, Z.; Megariotis, G.; Reis, H.; Papadopoulos, M. G., Computational studies of darunavir into HIV-1 protease and DMPC bilayer: necessary conditions for effective binding and the role of the flaps. *J. Chem. Inf. Model.* **2012,** *52* (6), 1542-58.
95. Leonis, G.; Steinbrecher, T.; Papadopoulos, M. G., A contribution to the drug resistance mechanism of darunavir, amprenavir, indinavir, and saquinavir complexes with HIV-1 protease due to flap mutation I50V: a systematic MM-PBSA and thermodynamic integration study. *J. Chem. Inf. Model.* **2013,** *53* (8), 2141-53.
96. Ragland, D. A.; Whitfield, T. W.; Lee, S. K.; Swanstrom, R.; Zeldovich, K. B.; Kurt-Yilmaz, N.; Schiffer, C. A., Elucidating the Interdependence of Drug Resistance from Combinations of Mutations. *J Chem Theory Comput* **2017,** *13* (11), 5671-5682.
97. Hassan, S.; Srikakulam, S. K.; Chandramohan, Y.; Thangam, M.; Muthukumar, S.; Gayathri Devi, P. K.; Hanna, L. E., Exploring the conformational landscapes of HIV protease structural ensembles using principal component analysis. *Proteins* **2018,** *86* (9), 990-1000.
98. Palese, L. L., Conformations of the HIV-1 protease: A crystal structure data set analysis. *Biochim Biophys Acta Proteins Proteom* **2017,** *1865* (11 Pt A), 1416-1422.
99. King, N. M.; Melnick, L.; Prabu-Jeyabalan, M.; Nalivaika, E. A.; Yang, S. S.; Gao, Y.; Nie, X.; Zepp, C.; Heefner, D. L.; Schiffer, C. A., Lack of synergy for inhibitors targeting a multi-drug-resistant HIV-1 protease. *Protein Sci.* **2002,** *11* (2), 418-29.
100. Barrie, K. A.; Perez, E. E.; Lamers, S. L.; Farmerie, W. G.; Dunn, B. M.; Sleasman, J. W.; Goodenow, M. M., Natural variation in HIV-1 protease, Gag p7 and p6, and protease cleavage sites within gag/pol polyproteins: amino acid substitutions in the absence of protease inhibitors in mothers and children infected by human immunodeficiency virus type 1. *Virology* **1996,** *219* (2), 407-16.

101. Rose, J. R.; Salto, R.; Craik, C. S., Regulation of autoproteolysis of the HIV-1 and HIV-2 proteases with engineered amino acid substitutions. *J. Biol. Chem.* **1993,** *268* (16), 11939-45.
102. Hui, J. O.; Tomasselli, A. G.; Reardon, I. M.; Lull, J. M.; Brunner, D. P.; Tomich, C.-S. C.; Heinrikson, R. L., Large scale purification and refolding of HIV-1 protease fromEscherichia coli inclusion bodies. *J. Protein Chem.* **1993,** *12* (3), 323-327.
103. Liu, Y.; Kati, W.; Chen, C. M.; Tripathi, R.; Molla, A.; Kohlbrenner, W., Use of a fluorescence plate reader for measuring kinetic parameters with inner filter effect correction. *Anal. Biochem.* **1999,** *267* (2), 331-5.
104. Soumana, D. I.; Kurt Yilmaz, N.; Ali, A.; Prachanronarong, K. L.; Schiffer, C. A., Molecular and Dynamic Mechanism Underlying Drug Resistance in Genotype 3 Hepatitis C NS3/4A Protease. *J Am Chem Soc* **2016,** *138* (36), 11850-9.
105. Otwinowski, Z.; Minor, W., [20] Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **1997,** *276*, 307-326.
106. McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J., Phaser crystallographic software. *J. Appl. Crystallogr.* **2007,** *40* (Pt 4), 658-674.
107. Emsley, P.; Cowtan, K., Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **2004,** *60* (Pt 12 Pt 1), 2126-32.
108. Adams, P. D.; Afonine, P. V.; Bunkoczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L. W.; Kapral, G. J.; Grosse-Kunstleve, R. W.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R.; Read, R. J.; Richardson, D. C.; Richardson, J. S.; Terwilliger, T. C.; Zwart, P. H., PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **2010,** *66* (Pt 2), 213-21.
109. Moriarty, N. W.; Grosse-Kunstleve, R. W.; Adams, P. D., electronic Ligand Builder and Optimization Workbench (eLBOW): a tool for ligand coordinate and restraint generation. *Acta Crystallogr. D Biol. Crystallogr.* **2009,** *65* (Pt 10), 1074-80.
110. Brunger, A. T., Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* **1992,** *355* (6359), 472-5.
111. Davis, I. W.; Leaver-Fay, A.; Chen, V. B.; Block, J. N.; Kapral, G. J.; Wang, X.; Murray, L. W.; Arendall, W. B., 3rd; Snoeyink, J.; Richardson, J. S.; Richardson, D. C., MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* **2007,** *35* (Web Server issue), W375-83.
112. Jayaraman, S.; Shah, K., Comparative studies on inhibitors of HIV protease: a target for drug design. *In Silico Biol.* **2008,** *8* (5-6), 427-47.
113. Prabu-Jeyabalan, M.; Nalivaika, E. A.; Romano, K.; Schiffer, C. A., Mechanism of substrate recognition by drug-resistant human immunodeficiency virus type 1 protease variants revealed by a novel structural intermediate. *J. Virol.* **2006,** *80* (7), 3607-16.
114. Sastry, G. M.; Adzhigirey, M.; Day, T.; Annabhimoju, R.; Sherman, W., Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *J. Comput. Aided Mol. Des.* **2013,** *27* (3), 221-34.
115. Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J., Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* **1996,** *118* (45), 11225-11236.
116. Nalam, M. N.; Ali, A.; Altman, M. D.; Reddy, G. S.; Chellappan, S.; Kairys, V.; Ozen, A.; Cao, H.; Gilson, M. K.; Tidor, B.; Rana, T. M.; Schiffer, C. A., Evaluating the substrate-envelope

hypothesis: structural analysis of novel HIV-1 protease inhibitors designed to be robust against drug resistance. *J. Virol.* **2010,** *84* (10), 5368-78.
117. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V., Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011,** *12* (Oct), 2825-2830.
118. Galvani, A. P., The role of mutation accumulation in HIV progression. *Proc. Biol. Sci.* **2005,** *272* (1574), 1851-8.
119. Maldarelli, F.; Kearney, M.; Palmer, S.; Stephens, R.; Mican, J.; Polis, M. A.; Davey, R. T.; Kovacs, J.; Shao, W.; Rock-Kress, D.; Metcalf, J. A.; Rehm, C.; Greer, S. E.; Lucey, D. L.; Danley, K.; Alter, H.; Mellors, J. W.; Coffin, J. M., HIV populations are large and accumulate high genetic diversity in a nonlinear fashion. *J. Virol.* **2013,** *87* (18), 10313-23.
120. Patel, P. H.; Preston, B. D., Marked infidelity of human immunodeficiency virus type 1 reverse transcriptase at RNA and DNA template ends. *Proc. Natl. Acad. Sci. U. S. A.* **1994,** *91* (2), 549-53.
121. Abram, M. E.; Ferris, A. L.; Das, K.; Quinones, O.; Shao, W.; Tuske, S.; Alvord, W. G.; Arnold, E.; Hughes, S. H., Mutations in HIV-1 reverse transcriptase affect the errors made in a single cycle of viral replication. *J. Virol.* **2014,** *88* (13), 7589-601.
122. Smyth, R. P.; Schlub, T. E.; Grimm, A. J.; Waugh, C.; Ellenberg, P.; Chopra, A.; Mallal, S.; Cromer, D.; Mak, J.; Davenport, M. P., Identifying recombination hot spots in the HIV-1 genome. *J. Virol.* **2014,** *88* (5), 2891-902.
123. Yasuda, J. M.; Miller, C.; Currier, J. S.; Forthal, D. N.; Kemper, C. A.; Beall, G. N.; Tilles, J. G.; Capparelli, E. V.; McCutchan, J. A.; Haubrich, R. H.; California Collaborative Treatment, G., The correlation between plasma concentrations of protease inhibitors, medication adherence and virological outcome in HIV-infected patients. *Antivir Ther* **2004,** *9* (5), 753-61.
124. Wu, T. D.; Schiffer, C. A.; Gonzales, M. J.; Taylor, J.; Kantor, R.; Chou, S.; Israelski, D.; Zolopa, A. R.; Fessel, W. J.; Shafer, R. W., Mutation patterns and structural correlates in human immunodeficiency virus type 1 protease following different protease inhibitor treatments. *J Virol* **2003,** *77* (8), 4836-47.
125. Clavel, F.; Mammano, F., Role of Gag in HIV Resistance to Protease Inhibitors. *Viruses* **2010,** *2* (7), 1411-26.
126. Dam, E.; Quercia, R.; Glass, B.; Descamps, D.; Launay, O.; Duval, X.; Krausslich, H. G.; Hance, A. J.; Clavel, F.; Group, A. S., Gag mutations strongly contribute to HIV-1 resistance to protease inhibitors in highly drug-experienced patients besides compensating for fitness loss. *PLoS Pathog.* **2009,** *5* (3), e1000345.
127. Nalam, M. N.; Peeters, A.; Jonckers, T. H.; Dierynck, I.; Schiffer, C. A., Crystal structure of lysine sulfonamide inhibitor reveals the displacement of the conserved flap water molecule in human immunodeficiency virus type 1 protease. *J. Virol.* **2007,** *81* (17), 9512-8.
128. Lv, Z.; Chu, Y.; Wang, Y., HIV protease inhibitors: a review of molecular selectivity and toxicity. *HIV AIDS (Auckl.)* **2015,** *7*, 95-104.
129. Muzammil, S.; Ross, P.; Freire, E., A major role for a set of non-active site mutations in the development of HIV-1 protease drug resistance. *Biochemistry* **2003,** *42* (3), 631-8.
130. Ohtaka, H.; Schon, A.; Freire, E., Multidrug resistance to HIV-1 protease inhibition requires cooperative coupling between distal mutations. *Biochemistry* **2003,** *42* (46), 13659-66.

131. Clemente, J. C.; Moose, R. E.; Hemrajani, R.; Whitford, L. R.; Govindasamy, L.; Reutzel, R.; McKenna, R.; Agbandje-McKenna, M.; Goodenow, M. M.; Dunn, B. M., Comparing the accumulation of active- and nonactive-site mutations in the HIV-1 protease. *Biochemistry* **2004,** *43* (38), 12141-51.
132. Prachanronarong, K. L.; Canale, A. S.; Liu, P.; Somasundaran, M.; Hou, S.; Poh, Y.-P.; Han, T.; Zhu, Q.; Renzette, N.; Zeldovich, K. B.; Kowalik, T. F.; Kurt-Yilmaz, N.; Jensen, J. D.; Bolon, D. N. A.; Marasco, W. A.; Finberg, R. W.; Schiffer, C. A.; Wang, J. P., Mutations in Influenza A Virus Neuraminidase and Hemagglutinin Confer Resistance against a Broadly Neutralizing Hemagglutinin Stem Antibody. *Journal of Virology* **2019,** *93* (2), e01639-18.
133. Matthew, A. N.; Leidner, F.; Newton, A.; Petropoulos, C. J.; Huang, W.; Ali, A.; KurtYilmaz, N.; Schiffer, C. A., Molecular Mechanism of Resistance in a Clinically Significant Double-Mutant Variant of HCV NS3/4A Protease. *Structure* **2018,** *26* (10), 1360-1372 e5.
134. Lee, S. K.; Potempa, M.; Kolli, M.; Ozen, A.; Schiffer, C. A.; Swanstrom, R., 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* **2012,** *287* (16), 13279-90.
135. Deshmukh, L.; Louis, J. M.; Ghirlando, R.; Clore, G. M., Transient HIV-1 Gag-protease interactions revealed by paramagnetic NMR suggest origins of compensatory drug resistance mutations. *Proc Natl Acad Sci U S A* **2016,** *113* (44), 12456-12461.
136. King, N. M.; Prabu-Jeyabalan, M.; Nalivaika, E. A.; Wigerinck, P.; de Bethune, M. P.; Schiffer, C. A., Structural and thermodynamic basis for the binding of TMC114, a next-generation human immunodeficiency virus type 1 protease inhibitor. *J Virol* **2004,** *78* (21), 12012-21.
137. Horovitz, A., Double-mutant cycles: a powerful tool for analyzing protein structure and function. *Fold Des.* **1996,** *1* (6), R121-6.
138. Goldfarb, N. E.; Ohanessian, M.; Biswas, S.; McGee, T. D., Jr.; Mahon, B. P.; Ostrov, D. A.; Garcia, J.; Tang, Y.; McKenna, R.; Roitberg, A.; Dunn, B. M., Defective hydrophobic sliding mechanism and active site expansion in HIV-1 protease drug resistant variant Gly48Thr/Leu89Met: mechanisms for the loss of saquinavir binding potency. *Biochemistry* **2015,** *54* (2), 422-33.
139. Ozen, A.; Sherman, W.; Schiffer, C. A., Improving the Resistance Profile of Hepatitis C NS3/4A Inhibitors: Dynamic Substrate Envelope Guided Design. *J. Chem. Theory Comput.* **2013,** *9* (12), 5693-5705.
140. McClendon, C. L.; Hua, L.; Barreiro, A.; Jacobson, M. P., Comparing Conformational Ensembles Using the Kullback-Leibler Divergence Expansion. *J. Chem. Theory Comput.* **2012,** *8* (6), 2115-2126.
141. <Hydrophobic Sliding Structure 2007.pdf>.
142. Jacobson, M. P.; Friesner, R. A.; Xiang, Z.; Honig, B., On the role of the crystal environment in determining protein side-chain conformations. *J. Mol. Biol.* **2002,** *320* (3), 597-608.
143. Olsson, M. H.; Sondergaard, C. R.; Rostkowski, M.; Jensen, J. H., PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. *J. Chem. Theory Comput.* **2011,** *7* (2), 525-37.

144. Sondergaard, C. R.; Olsson, M. H.; Rostkowski, M.; Jensen, J. H., Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of pKa Values. *J. Chem. Theory Comput.* **2011,** *7* (7), 2284-95.
145. Banks, J. L.; Beard, H. S.; Cao, Y.; Cho, A. E.; Damm, W.; Farid, R.; Felts, A. K.; Halgren, T. A.; Mainz, D. T.; Maple, J. R.; Murphy, R.; Philipp, D. M.; Repasky, M. P.; Zhang, L. Y.; Berne, B. J.; Friesner, R. A.; Gallicchio, E.; Levy, R. M., Integrated Modeling Program, Applied Chemical Theory (IMPACT). *J. Comput. Chem.* **2005,** *26* (16), 1752-80.
146. Bowers, K. J.; Chow, D. E.; Xu, H.; Dror, R. O.; Eastwood, M. P.; Gregersen, B. A.; Klepeis, J. L.; Kolossvary, I.; Moraes, M. A.; Sacerdoti, F. D.; Salmon, J. K.; Shan, Y.; Shaw, D. E. In *Scalable Algorithms for Molecular Dynamics Simulations on Commodity Clusters*, SC '06: Proceedings of the 2006 ACM/IEEE Conference on Supercomputing, 11-17 Nov. 2006; 2006; pp 43-43.
147. Ozen, A.; Haliloglu, T.; Schiffer, C. A., Dynamics of preferential substrate recognition in HIV-1 protease: redefining the substrate envelope. *J. Mol. Biol.* **2011,** *410* (4), 726-44.
148. Ghosh, A. K.; Osswald, H. L.; Prato, G., Recent progress in the development of HIV-1 protease inhibitors for the treatment of HIV/AIDS. *J. Med. Chem.* **2016,** *59* (11), 5172–5208.
149. Ali, A.; Bandaranayake, R. M.; Cai, Y.; King, N. M.; Kolli, M.; Mittal, S.; Murzycki, J. F.; Nalam, M. N.; Nalivaika, E. A.; Ozen, A.; Prabu-Jeyabalan, M. M.; Thayer, K.; Schiffer, C. A., Molecular basis for drug resistance in HIV-1 protease. *Viruses* **2010,** *2* (11), 2509–2535.
150. Shen, L.; Peterson, S.; Sedaghat, A. R.; McMahon, M. A.; Callender, M.; Zhang, H.; Zhou, Y.; Pitt, E.; Anderson, K. S.; Acosta, E. P.; Siliciano, R. F., Dose-response curve slope sets class-specific limits on inhibitory potential of anti-HIV drugs. *Nat. Med.* **2008,** *14* (7), 762–766.
151. Sampah, M. E.; Shen, L.; Jilek, B. L.; Siliciano, R. F., Dose-response curve slope is a missing dimension in the analysis of HIV-1 drug resistance. *Proc. Natl. Acad. Sci. U. S. A.* **2011,** *108* (18), 7613–7618.
152. Cihlar, T.; Fordyce, M., Current status and prospects of HIV treatment. *Curr. Opin. Virol.* **2016,** *18*, 50–56.
153. El Bouzidi, K.; White, E.; Mbisa, J. L.; Sabin, C. A.; Phillips, A. N.; Mackie, N.; Pozniak, A. L.; Tostevin, A.; Pillay, D.; Dunn, D. T., HIV-1 drug resistance mutations emerging on darunavir therapy in PI-naive and -experienced patients in the UK. *J. Antimicrob. Chemother.* **2016,** *71* (12), 3487–3494.
154. Stockdale, A. J.; Saunders, M. J.; Boyd, M. A.; Bonnett, L. J.; Johnston, V.; Wandeler, G.; Schoffelen, A. F.; Ciaffi, L.; Stafford, K.; Collier, A. C.; Paton, N. I.; Geretti, A. M., Effectiveness of protease inhibitor/nucleos(t)ide reverse transcriptase inhibitor-based second-line antiretroviral therapy for the treatment of human immunodeficiency virus type 1 infection in sub-Saharan Africa: a systematic review and meta-analysis. *Clin. Infect. Dis.* **2018,** *66* (12), 1846–1857.
155. Shafer, R. W., Human immunodeficiency virus type 1 drug resistance mutations update. *J. Infect. Dis.* **2017,** *216* (suppl_9), S843–S846.
156. Wensing, A. M.; Calvez, V.; Gunthard, H. F.; Johnson, V. A.; Paredes, R.; Pillay, D.; Shafer, R. W.; Richman, D. D., 2017 Update of the drug resistance mutations in HIV-1. *Top. Antivir. Med.* **2017,** *24* (4), 132–133.
157. Gunthard, H. F.; Calvez, V.; Paredes, R.; Pillay, D.; Shafer, R. W.; Wensing, A. M.; Jacobsen, D. M.; Richman, D. D., Human immunodeficiency virus drug resistance: 2018 recommendations of the international antiviral society-USA panel. *Clin. Infect. Dis.* **2019,** *68* (2), 177–187.

158. Saag, M. S.; Benson, C. A.; Gandhi, R. T.; Hoy, J. F.; Landovitz, R. J.; Mugavero, M. J.; Sax, P. E.; Smith, D. M.; Thompson, M. A.; Buchbinder, S. P.; Del Rio, C.; Eron, J. J., Jr.; Fatkenheuer, G.; Gunthard, H. F.; Molina, J. M.; Jacobsen, D. M.; Volberding, P. A., Antiretroviral drugs for treatment and prevention of HIV infection in adults: 2018 recommendations of the international antiviral society-USA panel. *JAMA* **2018,** *320* (4), 379–396.
159. Ghosh, A. K.; Chapsal, B. D.; Weber, I. T.; Mitsuya, H., Design of HIV protease inhibitors targeting protein backbone: an effective strategy for combating drug resistance. *Acc. Chem. Res.* **2008,** *41* (1), 78–86.
160. Surleraux, D. L.; Tahri, A.; Verschueren, W. G.; Pille, G. M.; de Kock, H. A.; Jonckers, T. H.; Peeters, A.; De Meyer, S.; Azijn, H.; Pauwels, R.; de Bethune, M. P.; King, N. M.; Prabu-Jeyabalan, M.; Schiffer, C. A.; Wigerinck, P. B., Discovery and selection of TMC114, a next generation HIV-1 protease inhibitor. *J. Med. Chem.* **2005,** *48* (6), 1813–1822.
161. Tie, Y.; Boross, P. I.; Wang, Y. F.; Gaddis, L.; Hussain, A. K.; Leshchenko, S.; Ghosh, A. K.; Louis, J. M.; Harrison, R. W.; Weber, I. T., High resolution crystal structures of HIV-1 protease with a potent non-peptide inhibitor (UIC-94017) active against multi-drug-resistant clinical strains. *J. Mol. Biol.* **2004,** *338* (2), 341–352.
162. Ghosh, A. K.; Rao, K. V.; Nyalapatla, P. R.; Osswald, H. L.; Martyr, C. D.; Aoki, M.; Hayashi, H.; Agniswamy, J.; Wang, Y. F.; Bulut, H.; Das, D.; Weber, I. T.; Mitsuya, H., Design and development of highly potent HIV-1 protease inhibitors with a crown-like oxotricyclic core as the P2-ligand to combat multidrug-resistant HIV variants. *J. Med. Chem.* **2017,** *60* (10), 4267–4278.
163. Ghosh, A. K.; P, R. N.; Kovela, S.; Rao, K. V.; Brindisi, M.; Osswald, H. L.; Amano, M.; Aoki, M.; Agniswamy, J.; Wang, Y. F.; Weber, I. T.; Mitsuya, H., Design and synthesis of highly potent HIV-1 protease inhibitors containing tricyclic fused ring systems as novel P2 ligands: structure-activity studies, biological and X-ray structural analysis. *J. Med. Chem.* **2018,** *61* (10), 4561–4577.
164. Ghosh, A. K.; Rao, K. V.; Nyalapatla, P. R.; Kovela, S.; Brindisi, M.; Osswald, H. L.; Sekhara Reddy, B.; Agniswamy, J.; Wang, Y. F.; Aoki, M.; Hattori, S. I.; Weber, I. T.; Mitsuya, H., Design of Highly Potent, Dual-Acting and Central-Nervous-System-Penetrating HIV-1 Protease Inhibitors with Excellent Potency against Multidrug-Resistant HIV-1 Variants. *ChemMedChem* **2018,** *13* (8), 803-815.
165. Aoki, M.; Hayashi, H.; Rao, K. V.; Das, D.; Higashi-Kuwata, N.; Bulut, H.; Aoki-Ogata, H.; Takamatsu, Y.; Yedidi, R. S.; Davis, D. A.; Hattori, S. I.; Nishida, N.; Hasegawa, K.; Takamune, N.; Nyalapatla, P. R.; Osswald, H. L.; Jono, H.; Saito, H.; Yarchoan, R.; Misumi, S.; Ghosh, A. K.; Mitsuya, H., A novel central nervous system-penetrating protease inhibitor overcomes human immunodeficiency virus 1 resistance with unprecedented aM to pM potency. *Elife* **2017,** *6*, e28020.
166. King, N. M.; Prabu-Jeyabalan, M.; Nalivaika, E. A.; Schiffer, C. A., Combating susceptibility to drug resistance: lessons from HIV-1 protease. *Chem. Biol.* **2004,** *11* (10), 1333–1338.
167. Altman, M. D.; Ali, A.; Reddy, G. S.; Nalam, M. N.; Anjum, S. G.; Cao, H.; Chellappan, S.; Kairys, V.; Fernandes, M. X.; Gilson, M. K.; Schiffer, C. A.; Rana, T. M.; Tidor, B., HIV-1 protease inhibitors from inverse design in the substrate envelope exhibit subnanomolar binding to drug-resistant variants. *J. Am. Chem. Soc.* **2008,** *130* (19), 6099–6113.

168. Kurt Yilmaz, N.; Swanstrom, R.; Schiffer, C. A., Improving viral protease inhibitors to counter drug resistance. *Trends Microbiol.* **2016,** *24* (7), 547–557.
169. Nalam, M. N.; Ali, A.; Reddy, G. S.; Cao, H.; Anjum, S. G.; Altman, M. D.; Yilmaz, N. K.; Tidor, B.; Rana, T. M.; Schiffer, C. A., Substrate envelope-designed potent HIV-1 protease inhibitors to avoid drug resistance. *Chem. Biol.* **2013,** *20* (9), 1116–1124.
170. Paulsen, J. L.; Leidner, F.; Ragland, D. A.; Kurt Yilmaz, N.; Schiffer, C. A., Interdependence of inhibitor recognition in HIV-1 protease. *J. Chem. Theory. Comput.* **2017,** *13* (5), 2300–2309.
171. Surleraux, D. L.; de Kock, H. A.; Verschueren, W. G.; Pille, G. M.; Maes, L. J.; Peeters, A.; Vendeville, S.; De Meyer, S.; Azijn, H.; Pauwels, R.; de Bethune, M. P.; King, N. M.; Prabu-Jeyabalan, M.; Schiffer, C. A.; Wigerinck, P. B., Design of HIV-1 protease inhibitors active on multidrug-resistant virus. *J. Med. Chem.* **2005,** *48* (6), 1965-1973.
172. Amano, M.; Koh, Y.; Das, D.; Li, J.; Leschenko, S.; Wang, Y. F.; Boross, P. I.; Weber, I. T.; Ghosh, A. K.; Mitsuya, H., A novel bis-tetrahydrofuranylurethane-containing nonpeptidic protease inhibitor (PI), GRL-98065, is potent against multiple-PI-resistant human immunodeficiency virus in vitro. *Antimicrob. Agents Chemother.* **2007,** *51* (6), 2143–2155.
173. Windsor, I. W.; Palte, M. J.; Lukesh, J. C., 3rd; Gold, B.; Forest, K. T.; Raines, R. T., Sub-picomolar inhibition of HIV-1 protease with a boronic acid. *J. Am. Chem. Soc.* **2018,** *140* (43), 14015–14018.
174. Corey, E. J.; Bakshi, R. K.; Shibata, S., Highly enantioselective borane reduction of ketones catalyzed by chiral oxazaborolidines. Mechanism and synthetic implications. *J. Am. Chem. Soc.* **1987,** *109* (18), 5551–5553.
175. Sharpless, K. B.; Amberg, W.; Bennani, Y. L.; Crispino, G. A.; Hartung, J.; Jeong, K. S.; Kwong, H. L.; Morikawa, K.; Wang, Z. M., The osmium-catalyzed asymmetric dihydroxylation: a new ligand class and a process improvement. *J. Org. Chem.* **1992,** *57* (10), 2768–2771.
176. Szermerski, M.; Melesina, J.; Wichapong, K.; Loppenberg, M.; Jose, J.; Sippl, W.; Holl, R., Synthesis, biological evaluation and molecular docking studies of benzyloxyacetohydroxamic acids as LpxC inhibitors. *Bioorg. Med. Chem.* **2014,** *22* (3), 1016–1628.
177. Windsor, I. W.; Raines, R. T., Fluorogenic assay for inhibitors of HIV-1 protease with sub-picomolar affinity. *Sci. Rep.* **2015,** *5*, 11286.
178. Ghosh, A. K., *Aspartic Acid Proteases as Therapeutic Targets*. Wiley-VCH: Weinheim, Germany, 2010.
179. Varghese, V.; Mitsuya, Y.; Fessel, W. J.; Liu, T. F.; Melikian, G. L.; Katzenstein, D. A.; Schiffer, C. A.; Holmes, S. P.; Shafer, R. W., Prototypical recombinant multi-protease-inhibitor-resistant infectious molecular clones of human immunodeficiency virus type 1. *Antimicrob. Agents Chemother.* **2013,** *57* (9), 4290–4299.
180. Miller, J. F.; Andrews, C. W.; Brieger, M.; Furfine, E. S.; Hale, M. R.; Hanlon, M. H.; Hazen, R. J.; Kaldor, I.; McLean, E. W.; Reynolds, D.; Sammond, D. M.; Spaltenstein, A.; Tung, R.; Turner, E. M.; Xu, R. X.; Sherrill, R. G., Ultra-potent P1 modified arylsulfonamide HIV protease inhibitors: the discovery of GW0385. *Bioorg. Med. Chem. Lett.* **2006,** *16* (7), 1788–1794.
181. Wei, X.; Decker, J. M.; Liu, H.; Zhang, Z.; Arani, R. B.; Kilby, J. M.; Saag, M. S.; Wu, X.; Shaw, G. M.; Kappes, J. C., Emergence of resistant human immunodeficiency virus type 1 in patients receiving fusion inhibitor (T-20) monotherapy. *Antimicrob. Agents Chemother.* **2002,** *46* (6), 1896–1905.

182. Lee, S. K.; Harris, J.; Swanstrom, R., A strongly transdominant mutation in the human immunodeficiency virus type 1 gag gene defines an Achilles heel in the virus life cycle. *J. Virol.* **2009,** *83* (17), 8536–8543.
183. King, N. M.; Melnick, L.; Prabu-Jeyabalan, M.; Nalivaika, E. A.; Yang, S. S.; Gao, Y.; Nie, X.; Zepp, C.; Heefner, D. L.; Schiffer, C. A., Lack of synergy for inhibitors targeting a multi-drug-resistant HIV-1 protease. *Protein Sci.* **2002,** *11* (2), 418–429.
184. Ozen, A.; Lin, K. H.; Kurt Yilmaz, N.; Schiffer, C. A., Structural basis and distal effects of Gag substrate coevolution in drug resistance to HIV-1 protease. *Proc. Natl. Acad. Sci. U. S. A.* **2014,** *111* (45), 15993–15998.
185. Rose, J. R.; Salto, R.; Craik, C. S., Regulation of autoproteolysis of the HIV-1 and HIV-2 proteases with engineered amino acid substitutions. *J. Biol. Chem.* **1993,** *268* (16), 11939–11945.
186. Hui, J. O.; Tomasselli, A. G.; Reardon, I. M.; Lull, J. M.; Brunner, D. P.; Tomich, C. S.; Heinrikson, R. L., Large scale purification and refolding of HIV-1 protease from Escherichia coli inclusion bodies. *J. Protein Chem.* **1993,** *12* (3), 323–327.
187. Liu, Y.; Kati, W.; Chen, C. M.; Tripathi, R.; Molla, A.; Kohlbrenner, W., Use of a fluorescence plate reader for measuring kinetic parameters with inner filter effect correction. *Anal. Biochem.* **1999,** *267* (2), 331–335.
188. Lockbaum, G. J.; Leidner, F.; Rusere, L. N.; Henes, M.; Kosovrasti, K.; Nachum, G. S.; Nalivaika, E. A.; Ali, A.; Yilmaz, N. K.; Schiffer, C. A., Structural adaptation of darunavir analogues against primary mutations in HIV-1 protease. *ACS Infect. Dis.* **2019,** *5* (2), 316–325.
189. Soumana, D. I.; Kurt Yilmaz, N.; Ali, A.; Prachanronarong, K. L.; Schiffer, C. A., Molecular and dynamic mechanism underlying drug resistance in genotype 3 hepatitis C NS3/4A protease. *J. Am. Chem. Soc.* **2016,** *138* (36), 11850–11859.
190. Otwinowski, Z.; Minor, W., Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **1997,** *276*, 307–326.
191. McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J., Phaser crystallographic software. *J. Appl. Crystallogr.* **2007,** *40* (Pt 4), 658–674.
192. Emsley, P.; Cowtan, K., Coot: Model-building tools for molecular graphics. *Acta Crystallogr.* **2004,** *D60*, 2126–2132.
193. Adams, P. D.; Afonine, P. V.; Bunkoczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L.-W.; Kapral, G. J.; Grosse-Kunstleve, R. W.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R.; Read, R. J.; Richardson, D. C.; Richardson, J. S.; Terwilliger, T. C.; Zwart, P. H., PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **2010,** *66* (2), 213–221.
194. Davis, I. W.; Leaver-Fay, A.; Chen, V. B.; Block, J. N.; Kapral, G. J.; Wang, X.; Murray, L. W.; Arendall, W. B.; Snoeyink, J.; Richardson, J. S.; Richardson, D. C., MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucl. Acids Res.* **2007,** *35* (suppl 2), W375–W383.
195. *PyMOL: The PyMOL Molecular Graphics System, Version 2.1, Schrödinger, LLC.*
196. Schrödinger Schrödinger Release 2018-3; Schrödinger, LLC, New York, NY, United States: 2018.
197. Ozen, A.; Haliloglu, T.; Schiffer, C. A., Dynamics of preferential substrate recognition in HIV-1 protease: redefining the substrate envelope. *J. Mol. Biol.* **2011,** *410* (4), 726–744.

198. Ghosh, A. K.; Chapsal, B. D.; Weber, I. T.; Mitsuya, H., Design of HIV protease inhibitors targeting protein backbone: an effective strategy for combating drug resistance. *Acc Chem Res* **2008,** *41* (1), 78-86.
199. Hazen, R.; Harvey, R.; Ferris, R.; Craig, C.; Yates, P.; Griffin, P.; Miller, J.; Kaldor, I.; Ray, J.; Samano, V.; Furfine, E.; Spaltenstein, A.; Hale, M.; Tung, R.; St Clair, M.; Hanlon, M.; Boone, L., In vitro antiviral activity of the novel, tyrosyl-based human immunodeficiency virus (HIV) type 1 protease inhibitor brecanavir (GW640385) in combination with other antiretrovirals and against a panel of protease inhibitor-resistant HIV. *Antimicrob. Agents Chemother.* **2007,** *51* (9), 3147–3154.
200. Cihlar, T.; He, G. X.; Liu, X.; Chen, J. M.; Hatada, M.; Swaminathan, S.; McDermott, M. J.; Yang, Z. Y.; Mulato, A. S.; Chen, X.; Leavitt, S. A.; Stray, K. M.; Lee, W. A., Suppression of HIV-1 protease inhibitor resistance by phosphonate-mediated solvent anchoring. *J. Mol. Biol.* **2006,** *363* (3), 635–647.
201. Dandache, S.; Sevigny, G.; Yelle, J.; Stranix, B. R.; Parkin, N.; Schapiro, J. M.; Wainberg, M. A.; Wu, J. J., In vitro antiviral activity and cross-resistance profile of PL-100, a novel protease inhibitor of human immunodeficiency virus type 1. *Antimicrob. Agents Chemother.* **2007,** *51* (11), 4036–4043.
202. Gulnik, S. V.; Eissenstat, M., Approaches to the design of HIV protease inhibitors with improved resistance profiles. *Curr Opin HIV AIDS* **2008,** *3* (6), 633-41.
203. Aoki, M.; Hayashi, H.; Rao, K. V.; Das, D.; Higashi-Kuwata, N.; Bulut, H.; Aoki-Ogata, H.; Takamatsu, Y.; Yedidi, R. S.; Davis, D. A.; Hattori, S. I.; Nishida, N.; Hasegawa, K.; Takamune, N.; Nyalapatla, P. R.; Osswald, H. L.; Jono, H.; Saito, H.; Yarchoan, R.; Misumi, S.; Ghosh, A. K.; Mitsuya, H., A novel central nervous system-penetrating protease inhibitor overcomes human immunodeficiency virus 1 resistance with unprecedented aM to pM potency. *Elife* **2017,** *6*.
204. Ghosh, A. K.; Rao, K. V.; Nyalapatla, P. R.; Osswald, H. L.; Martyr, C. D.; Aoki, M.; Hayashi, H.; Agniswamy, J.; Wang, Y. F.; Bulut, H.; Das, D.; Weber, I. T.; Mitsuya, H., Design and Development of Highly Potent HIV-1 Protease Inhibitors with a Crown-Like Oxotricyclic Core as the P2-Ligand To Combat Multidrug-Resistant HIV Variants. *J. Med. Chem.* **2017,** *60* (10), 4267-4278.
205. Chen, X.; Kempf, D. J.; Li, L.; Sham, H. L.; Vasavanonda, S.; Wideburg, N. E.; Saldivar, A.; Marsh, K. C.; McDonald, E.; Norbeck, D. W., Synthesis and SAR studies of potent HIV protease inhibitors containing novel dimethylphenoxyl acetates as P2 ligands. *Bioorg Med Chem Lett* **2003,** *13* (21), 3657-60.
206. Ghosh, A. K.; Ramu Sridhar, P.; Kumaragurubaran, N.; Koh, Y.; Weber, I. T.; Mitsuya, H., Bis-tetrahydrofuran: a privileged ligand for darunavir and a new generation of hiv protease inhibitors that combat drug resistance. *ChemMedChem* **2006,** *1* (9), 939-50.
207. Cannizzaro, C. E.; Yang, Z.-Y.; Desai, M. C.; Mitchell, M. L.; Xu, L.; Swaminathan, S.; Chen, J. M. Antiviral protease inhibitors. PCT Int. App. WO 2008/011117, 2008.
208. Chen, X.; Li, L.; Kempf, D. J.; Sham, H.; Wideburg, N. E.; Saldivar, A.; Vasavanonda, S.; Marsh, K. C.; McDonald, E.; Norbeck, D. W., Evaluation of furofuran as a P2 ligand for symmetry-based HIV protease inhibitors. *Bioorganic & Medicinal Chemistry Letters* **1996,** *6* (23), 2847-2852.
209. Haight, A. R.; Stuk, T. L.; Allen, M. S.; Bhagavatula, L.; Fitzgerald, M.; Hannick, S. M.; Kerdesky, F. A. J.; Menzia, J. A.; Parekh, S. I.; Robbins, T. A.; Scarpetti, D.; Tien, J.-H. J., Reduction

of an Enaminone: Synthesis of the Diamino Alcohol Core of Ritonavir. *Organic Process Research & Development* **1999,** *3* (2), 94-100.
210. Engle, K. M.; Wang, D. H.; Yu, J. Q., Ligand-accelerated C-H activation reactions: evidence for a switch of mechanism. *J. Am. Chem. Soc.* **2010,** *132* (40), 14137–14151.
211. Scott, W. R.; Schiffer, C. A., Curling of flap tips in HIV-1 protease as a mechanism for substrate entry and tolerance of drug resistance. *Structure* **2000,** *8* (12), 1259-65.
212. Yu, Y.; Wang, J.; Chen, Z.; Wang, G.; Shao, Q.; Shi, J.; Zhu, W., Structural insights into HIV-1 protease flap opening processes and key intermediates. *RSC Adv.* **2017,** *7* (71), 45121–45128.
213. Rusere, L. N.; Lockbaum, G. J.; Lee, S. K.; Henes, M.; Kosovrasti, K.; Spielvogel, E.; Nalivaika, E. A.; Swanstrom, R.; Yilmaz, N. K.; Schiffer, C. A.; Ali, A., HIV-1 Protease Inhibitors Incorporating Stereochemically Defined P2' Ligands To Optimize Hydrogen Bonding in the Substrate Envelope. *J. Med. Chem.* **2019,** *62* (17), 8062-8079.
214. Otwinowski, Z.; Minor, W., Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol* **1997,** *276*, 307-26.
215. PyMOL *The PyMOL Molecular Graphics System, Version 2.3*, Schrödinger, LLC.
216. Schrödinger *Schrödinger Release 2019-2*, Schrödinger, LLC, New York, NY, United States: 2019.
217. Harder, E.; Damm, W.; Maple, J.; Wu, C.; Reboul, M.; Xiang, J. Y.; Wang, L.; Lupyan, D.; Dahlgren, M. K.; Knight, J. L.; Kaus, J. W.; Cerutti, D. S.; Krilov, G.; Jorgensen, W. L.; Abel, R.; Friesner, R. A., OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins. *J. Chem. Theory Comput.* **2016,** *12* (1), 281–296.
218. Nalam, M. N.; Ali, A.; Altman, M. D.; Reddy, G. S.; Chellappan, S.; Kairys, V.; Ozen, A.; Cao, H.; Gilson, M. K.; Tidor, B.; Rana, T. M.; Schiffer, C. A., Evaluating the Substrate-Envelope Hypothesis: Structural Analysis of Novel HIV-1 Protease Inhibitors Designed to be Robust against Drug Resistance. *J. Virol.* **2010,** *84* (10), 5368–5378.
219. Leidner, F.; Kurt Yilmaz, N.; Paulsen, J.; Muller, Y. A.; Schiffer, C. A., Hydration Structure and Dynamics of Inhibitor-Bound HIV-1 Protease. *J. Chem. Theory Comput.* **2018,** *14* (5), 2784–2796.
220. Henes, M.; Lockbaum, G. J.; Kosovrasti, K.; Leidner, F.; Nachum, G. S.; Nalivaika, E. A.; Lee, S. K.; Spielvogel, E.; Zhou, S.; Swanstrom, R.; Bolon, D. N. A.; Kurt Yilmaz, N.; Schiffer, C. A., Picomolar to Micromolar: Elucidating the Role of Distal Mutations in HIV-1 Protease in Conferring Drug Resistance. *ACS Chem. Biol.* **2019,** *14* (11), 2441–2452.
221. Jacobson, M. P.; Friesner, R. A.; Xiang, Z.; Honig, B., On the role of the crystal environment in determining protein side-chain conformations. *J. Mol. Biol.* **2002,** *320* (3), 597–608.
222. Olsson, M. H.; Sondergaard, C. R.; Rostkowski, M.; Jensen, J. H., PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. *J. Chem. Theory Comput.* **2011,** *7* (2), 525–537.
223. Banks, J. L.; Beard, H. S.; Cao, Y.; Cho, A. E.; Damm, W.; Farid, R.; Felts, A. K.; Halgren, T. A.; Mainz, D. T.; Maple, J. R.; Murphy, R.; Philipp, D. M.; Repasky, M. P.; Zhang, L. Y.; Berne, B. J.; Friesner, R. A.; Gallicchio, E.; Levy, R. M., Integrated Modeling Program, Applied Chemical Theory (IMPACT). *J. Comput. Chem.* **2005,** *26* (16), 1752–1780.

224. Proietti, F. A.; Carneiro-Proietti, A. B.; Catalan-Soares, B. C.; Murphy, E. L., Global epidemiology of HTLV-I infection and associated diseases. *Oncogene* **2005,** *24* (39), 6058-68.
225. Goncalves, D. U.; Proietti, F. A.; Ribas, J. G.; Araujo, M. G.; Pinheiro, S. R.; Guedes, A. C.; Carneiro-Proietti, A. B., Epidemiology, treatment, and prevention of human T-cell leukemia virus type 1-associated diseases. *Clin. Microbiol. Rev.* **2010,** *23* (3), 577-89.
226. Einsiedel, L.; Woodman, R. J.; Flynn, M.; Wilson, K.; Cassar, O.; Gessain, A., Human T-Lymphotropic Virus type 1 infection in an Indigenous Australian population: epidemiological insights from a hospital-based cohort study. *BMC Public Health* **2016,** *16*, 787.
227. De Mendoza, C.; Piron, M.; Gonzalez, R.; Jimenez, A.; Caballero, E.; Roc, L.; Benito, R.; Ramos, J. M.; Soriano, V.; Group, H. S. S., Clinical Presentation of Individuals With Human T-Cell Leukemia Virus Type-1 Infection in Spain. *Open Forum Infect Dis* **2019,** *6* (2), ofz036.
228. Dewan, M. Z.; Uchihara, J. N.; Terashima, K.; Honda, M.; Sata, T.; Ito, M.; Fujii, N.; Uozumi, K.; Tsukasaki, K.; Tomonaga, M.; Kubuki, Y.; Okayama, A.; Toi, M.; Mori, N.; Yamamoto, N., Efficient intervention of growth and infiltration of primary adult T-cell leukemia cells by an HIV protease inhibitor, ritonavir. *Blood* **2006,** *107* (2), 716-24.
229. Hurst, S. A.; Appelgren, K. E.; Kourtis, A. P., Prevention of mother-to-child transmission of HIV type 1: the role of neonatal and infant prophylaxis. *Expert Rev. Anti Infect. Ther.* **2015,** *13* (2), 169-81.
230. Wolfenden, R., Transition state analogues for enzyme catalysis. *Nature* **1969,** *223* (5207), 704-5.
231. Schramm, V. L., Enzymatic transition state theory and transition state analogue design. *J. Biol. Chem.* **2007,** *282* (39), 28297-300.
232. Girard, P. M.; Antinori, A.; Arribas, J. R.; Ripamonti, D.; Bicer, C.; Netzle-Sveine, B.; Hadacek, B.; Moecklinghoff, C., Week 96 efficacy and safety of darunavir/ritonavir monotherapy vs. darunavir/ritonavir with two nucleoside reverse transcriptase inhibitors in the PROTEA trial. *HIV Med.* **2017,** *18* (1), 5-12.
233. Kimura, T.; Nguyen, J. T.; Maegawa, H.; Nishiyama, K.; Arii, Y.; Matsui, Y.; Hayashi, Y.; Kiso, Y., Chipping at large, potent human T-cell leukemia virus type 1 protease inhibitors to uncover smaller, equipotent inhibitors. *Bioorg. Med. Chem. Lett.* **2007,** *17* (12), 3276-80.
234. Nguyen, J. T.; Zhang, M.; Kumada, H. O.; Itami, A.; Nishiyama, K.; Kimura, T.; Cheng, M.; Hayashi, Y.; Kiso, Y., Truncation and non-natural amino acid substitution studies on HTLV-I protease hexapeptidic inhibitors. *Bioorg. Med. Chem. Lett.* **2008,** *18* (1), 366-70.
235. Zhang, M.; Nguyen, J. T.; Kumada, H. O.; Kimura, T.; Cheng, M.; Hayashi, Y.; Kiso, Y., Synthesis and activity of tetrapeptidic HTLV-I protease inhibitors possessing different P3-cap moieties. *Bioorg. Med. Chem.* **2008,** *16* (10), 5795-802.
236. Zhang, M.; Nguyen, J. T.; Kumada, H. O.; Kimura, T.; Cheng, M.; Hayashi, Y.; Kiso, Y., Locking the two ends of tetrapeptidic HTLV-I protease inhibitors inside the enzyme. *Bioorg. Med. Chem.* **2008,** *16* (14), 6880-90.
237. Nguyen, J. T.; Kato, K.; Kumada, H. O.; Hidaka, K.; Kimura, T.; Kiso, Y., Maintaining potent HTLV-I protease inhibition without the P3-cap moiety in small tetrapeptidic inhibitors. *Bioorg. Med. Chem. Lett.* **2011,** *21* (6), 1832-7.
238. Satoh, T.; Li, M.; Nguyen, J. T.; Kiso, Y.; Gustchina, A.; Wlodawer, A., Crystal structures of inhibitor complexes of human T-cell leukemia virus (HTLV-1) protease. *J. Mol. Biol.* **2010,** *401* (4), 626-41.

239. Kuhnert, M.; Blum, A.; Steuber, H.; Diederich, W. E., Privileged Structures Meet Human T-Cell Leukemia Virus-1 (HTLV-1): C2-Symmetric 3,4-Disubstituted Pyrrolidines as Nonpeptidic HTLV-1 Protease Inhibitors. *J. Med. Chem.* **2015,** *58* (11), 4845-50.
240. Demir, A.; Oguariri, R. M.; Magis, A.; Ostrov, D. A.; Imamichi, T.; Dunn, B. M., Kinetic characterization of newly discovered inhibitors of various constructs of human T-cell leukemia virus-1 (HTLV-1) protease and their effect on HTLV-1-infected cells. *Antivir. Ther.* **2012,** *17* (5), 883-92.
241. Louis, J. M.; Oroszlan, S.; Tozser, J., Stabilization from autoproteolysis and kinetic characterization of the human T-cell leukemia virus type 1 proteinase. *J. Biol. Chem.* **1999,** *274* (10), 6660-6.
242. Callebaut, C.; Stray, K.; Tsai, L.; Williams, M.; Yang, Z. Y.; Cannizzaro, C.; Leavitt, S. A.; Liu, X.; Wang, K.; Murray, B. P.; Mulato, A.; Hatada, M.; Priskich, T.; Parkin, N.; Swaminathan, S.; Lee, W.; He, G. X.; Xu, L.; Cihlar, T., In vitro characterization of GS-8374, a novel phosphonate-containing inhibitor of HIV-1 protease with a favorable resistance profile. *Antimicrob. Agents Chemother.* **2011,** *55* (4), 1366-76.
243. Henes, M.; Lockbaum, G. J.; Kosovrasti, K.; Leidner, F.; Nachum, G. S.; Nalivaika, E. A.; Lee, S. K.; Spielvogel, E.; Zhou, S.; Swanstrom, R.; Bolon, D. N. A.; Kurt Yilmaz, N.; Schiffer, C. A., Picomolar to Micromolar: Elucidating the Role of Distal Mutations in HIV-1 Protease in Conferring Drug Resistance. *ACS Chem. Biol.* **2019**.
244. Miller, J. F.; Andrews, C. W.; Brieger, M.; Furfine, E. S.; Hale, M. R.; Hanlon, M. H.; Hazen, R. J.; Kaldor, I.; McLean, E. W.; Reynolds, D.; Sammond, D. M.; Spaltenstein, A.; Tung, R.; Turner, E. M.; Xu, R. X.; Sherrill, R. G., Ultra-potent P1 modified arylsulfonamide HIV protease inhibitors: the discovery of GW0385. *Bioorg. Med. Chem. Lett.* **2006,** *16* (7), 1788-94.
245. Hui, J. O.; Tomasselli, A. G.; Reardon, I. M.; Lull, J. M.; Brunner, D. P.; Tomich, C. S.; Heinrikson, R. L., Large scale purification and refolding of HIV-1 protease from Escherichia coli inclusion bodies. *J. Protein Chem.* **1993,** *12* (3), 323-7.
246. Henes, M.; Kosovrasti, K.; Lockbaum, G. J.; Leidner, F.; Nachum, G. S.; Nalivaika, E. A.; Bolon, D. N. A.; Kurt Yilmaz, N.; Schiffer, C. A.; Whitfield, T. W., Molecular Determinants of Epistasis in HIV-1 Protease: Elucidating the Interdependence of L89V and L90M Mutations in Resistance. *Biochemistry* **2019,** *58* (35), 3711-3726.
247. Pothineni, S. B.; Venugopalan, N.; Ogata, C. M.; Hilgart, M. C.; Stepanov, S.; Sanishvili, R.; Becker, M.; Winter, G.; Sauter, N. K.; Smith, J. L.; Fischetti, R. F., Tightly integrated single- and multi-crystal data collection strategy calculation and parallelized data processing in JBluIce beamline control system. *J. Appl. Crystallogr.* **2014,** *47* (Pt 6), 1992-1999.
248. Kabsch, W., Xds. *Acta Crystallogr. D Biol. Crystallogr.* **2010,** *66* (Pt 2), 125-32.
249. Evans, P., Scaling and assessment of data quality. *Acta Crystallogr. D Biol. Crystallogr.* **2006,** *62* (Pt 1), 72-82.
250. Evans, P. R., An introduction to data reduction: space-group determination, scaling and intensity statistics. *Acta Crystallogr. D Biol. Crystallogr.* **2011,** *67* (Pt 4), 282-92.
251. DeLano, W. L., Pymol: An open-source molecular graphics tool. *CCP4 Newsl. Protein Crystallogr.* **2002,** *40*, 82-92.
252. Powles, S. B.; Yu, Q., Evolution in action: plants resistant to herbicides. *Annu. Rev. Plant Biol.* **2010,** *61*, 317-47.

253. Munita, J. M.; Arias, C. A., Mechanisms of Antibiotic Resistance. *Microbiol Spectr* **2016,** *4* (2).
254. Queener, S. F.; Cody, V.; Pace, J.; Torkelson, P.; Gangjee, A., Trimethoprim resistance of dihydrofolate reductase variants from clinical isolates of Pneumocystis jirovecii. *Antimicrob. Agents Chemother.* **2013,** *57* (10), 4990–4998.
255. Roskoski, R., Jr., Properties of FDA-approved small molecule protein kinase inhibitors. *Pharmacol. Res.* **2019,** *144*, 19–50.
256. Redaelli, S.; Mologni, L.; Rostagno, R.; Piazza, R.; Magistroni, V.; Ceccon, M.; Viltadi, M.; Flynn, D.; Gambacorti-Passerini, C., Three novel patient-derived BCR/ABL mutants show different sensitivity to second and third generation tyrosine kinase inhibitors. *Am J Hematol* **2012,** *87* (11), E125-8.
257. Nan, X.; Xie, C.; Yu, X.; Liu, J., EGFR TKI as first-line treatment for patients with advanced EGFR mutation-positive non-small-cell lung cancer. *Oncotarget* **2017,** *8* (43), 75712–75726.
258. Morgillo, F.; Della Corte, C. M.; Fasano, M.; Ciardiello, F., Mechanisms of resistance to EGFR-targeted drugs: lung cancer. *ESMO Open* **2016,** *1* (3), e000060.
259. Wu, S. G.; Shih, J. Y., Management of acquired resistance to EGFR TKI-targeted therapy in advanced non-small cell lung cancer. *Mol Cancer* **2018,** *17* (1), 38.
260. Westover, D.; Zugazagoitia, J.; Cho, B. C.; Lovly, C. M.; Paz-Ares, L., Mechanisms of acquired resistance to first- and second-generation EGFR tyrosine kinase inhibitors. *Ann. Oncol.* **2018,** *29* (suppl_1), i10–i19.
261. Ricordel, C.; Friboulet, L.; Facchinetti, F.; Soria, J. C., Molecular mechanisms of acquired resistance to third-generation EGFR-TKIs in EGFR T790M-mutant lung cancer. *Ann. Oncol.* **2018,** *29* (suppl_1), i28–i37.