EXPLORING MOLECULAR MECHANISMS OF DRUG RESISTANCE IN HIV-1 PROTEASE THROUGH BIOCHEMICAL AND BIOPHYSICAL STUDIES

A Dissertation Presented

By

RAJINTHA MALALA BANDARANAYAKE

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

DOCTOR OF PHILOSOPHY

MAY 20th, 2010

BIOCHEMISTRY AND MOLECULAR PHARMACOLOGY

EXPLORING MOLECULAR MECHANISMS OF DRUG RESISTANCE IN HIV-1 PROTEASE THROUGH BIOCHEMICAL AND BIOPHYSICAL STUDIES

A Dissertation Presented
By

RAJINTHA MALALA BANDARANAYAKE

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

Celia A. Schiffer, Thesis Advisor

William E. Royer, Member of Committee

Scot A. Wolfe, Member of Committee

Mohan Somasundaran, Member of Committee

Joseph G. Sodroski, Member of Committee

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

Lawrence J. Stern, Chair of Committee

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

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

Biochemistry and Molecular Pharmacology
May 20th, 2010

DEDICATION

With love, I dedicate this thesis to
Ammi and Apoochchi who taught me the importance of education,
Akki, the greatest sister a brother could ever have and
my wife Thilinie,
who crossed oceans to be with me and whose love and support are never ending.

ACKNOWLEDGEMENTS

I have been very lucky to have received the loving support of many people in my path through life. I cannot thank them enough for helping me get to where I am today. However, in a little effort I would like to start my acknowledgements by first thanking my family.

First, I would like to thank my parents. Throughout my life they have given me unconditional love, support and encouragement. I am so thankful to them for teaching me the importance of getting a good education. They have sacrificed so much to give me the education options that led me to where I am today. Apoochchi, I miss you and wish you were here.

My wife, Thilinie. She has been very understanding and supportive of what I want to do in life. I truly appreciate the sacrifices she has made by giving up her life and work in Sri Lanka to be with me. I am so thankful that she is a part of my life and I am honored that I get to share my life with her.

My sister Sharmini and her husband, Nath. They have been so supportive in my choice to come to the United States for my studies. I would also like to thank Mr. and Mrs. Seneviratne Bandara for all their love and support and for permitting me to marry their only daughter even though it meant that they would have to be many miles away from her.

As a student, I have been blessed to have been influenced by a number of great mentors. Dr. Celia Schiffer is one such mentor. I am very fortunate to be a part of her laboratory. Celia has been an amazing mentor who has given me all the support, guidance

and resources to carry out my work. She has given me the independence to formulate my own ideas to carry out research but has always been there to guide me when I needed it. And I truly appreciate all her efforts in trying to make a better public speaker out of me!

I would like to thank the past and present members of the Schiffer Lab for all their support with my work and making the lab such a fun place to work in. I would specially like to thank Dr. Nancy King for mentoring me when I initially joined the lab. I would also like to thank Dr. Moses Prabu-Jeyabalan and Ellen Nalivaika and Dr. Madhavi Nalam for teaching many of the techniques that I have used to carry out my research.

I would like to thank my thesis committee, Drs. Larry Stern, Bill Royer, Scot Wolfe and Mohan Somasundaran for all their support and advice over the past five years. I would specially like to thank Dr. Bill Royer for all his advice on x-ray crystallography and for being a great resource to go to.

I would also like to thank Dr Art Felix who was my mentor at the Ramapo College of New Jersey. Dr. Felix provided me the opportunity to be involved in research as an undergrad. He was a truly amazing mentor who instilled in me the love for biochemistry and research. Through him, I learned to appreciate the importance of mentoring others. Someday, I aspire to be able to mentor students the same way Dr. Felix does.

There are many other friends and family who have helped me in many ways over the years. I am truly thankful for all their love and support.

ABSTRACT

The human immunodeficiency virus type-1 (HIV-1) is the leading cause of acquired immunodeficiency syndrome (AIDS) in the world. As there is no cure currently available to treat HIV-1 infections or AIDS, the major focus of drug development efforts has been to target viral replication in an effort to slow down the progression of the infection to AIDS. The aspartyl protease of HIV-1 is an important component in the viral replication cycle and thus, has been an important anti-HIV-1 drug target. Currently there are nine protease inhibitors (PIs) that are being used successfully as a part of highly active antiretroviral therapy (HAART). However, as is with all HIV-1 drug targets, the emergence of drug resistance substitutions within protease is a major obstacle in the use of PIs. Understanding how amino acid substitutions within protease confer drug resistance is key to develop new PIs that are not influenced by resistance mutations. Thus, the primary focus of my dissertation research was to understand the molecular basis for drug resistance caused by some of these resistance substitutions.

Until recently, the genetic diversity of the HIV-1 genome was not considered to be important in formulating treatment strategies. However, as the prevalence of HIV-1 continues, the variability of the HIV-1 genome has now been identified as an important factor in how the virus spreads as well as how fast the infection progresses to AIDS. Clinical studies have also revealed that the pathway to protease inhibitor resistance can vary between HIV-1 clades. Therefore, in studying the molecular basis of drug resistance in HIV-1 protease, I have also attempted to understand how genetic variability in HIV-1 protease contributes to PI resistance.

In Chapters II, III and Appendix 1, I have examined how clade specific amino acid variations within HIV-1 CRF01_AE and clade C protease affect enzyme structure and activity. Furthermore, I have examined how these sequence variations, which are predominantly outside the active site, contribute to inhibitor resistance in comparison to clade B protease. With the results presented in Chapter II, I was able to show that sequence variations within CRF01_AE protease resulted in structural changes within the protease that might influence enzyme activity. In Chapter III, I focused on how sequence variations in CRF01_AE influence protease activity and inhibitor binding in comparison to clade B protease. Enzyme kinetics data showed that the CRF01-AE had reduced catalytic turnover rates when compared to clade B protease. Binding data also indicated that CRF01_AE protease had an inherent weaker affinity for the PIs nelfinavir (NFV) and darunavir (DRV). In work described in Chapter III, I have also examined the different pathways to NFV resistance seen in CRF01_AE and clade B protease. Using x-ray crystallographic studies I have shown the molecular mechanism by which the two different pathways confer NFV resistance. Furthermore, I provide a rational for why different resistance pathways might emerge in the two clades. In Appendix I, I present results from a parallel study carried out on clade C protease.

In Chapter IV, I have examined the role of residue 50 in HIV-1 protease in modulating inhibitor binding. Patients failing amprevavir (APV) and DRV therapy often develop the I50V substitution while the I50L substitution is often observed in patients failing atazanavir (ATV) therapy. This indicates that by making subtle changes at residue 50 the protease is able to confer differential PI resistance. With binding data presented in this chapter I have shown that substitutions at residue 50 change the susceptibility profiles of APV, DRV and ATV. Furthermore, from analyses of protease-inhibitor complexes, I have described structural insights into how substitutions at residue 50 can modulate inhibitor binding.

This thesis presents results that reveal mechanistic insights into how a number of resistance substitutions within protease confer drug resistance. The results on non-B clade proteases demonstrate that clade specific sequence variations play a role in modulating enzyme activity and influence the pathway taken to confer PI resistance. Furthermore, the results provide structural insights into how amino acid substitutions outside the active site effectively alter inhibitor binding.

TABLE OF CONTENTS

Title Page i
Signature Page ii
Dedication iii
Acknowledgements iv
Abstract vi
Table of Contents ix
List of Tables xii
List of Figures xiv
List of Abbreviations xvi
Copyright Notice xviii
Preface xix
Chapter I 1

Introduction

HIV and the AIDS Pandemic 2
Biology of HIV 3
HIV-1 Protease 11
Drug Resistance 18
Summary 26

Chapter II 27

Structural Analysis of HIV-1 CRF01_AE Protease in Complex with the Substrate p1-p6.

Abstract 28
Introduction 29
Materials And Methods 32
Results 34
Conclusions 39
Aknowledgments 42

Chapter III **43**

The Effect of Clade Specific Sequence Polymorphisms on HIV-1 Protease Activity and Inhibitor Resistance Pathways.

Abstract 44
Introduction 45
Materials And Methods 46
Results 54
Discussion 65
Acknowledgments 73

Chapter IV **74**

Differential Resistance in HIV-1 Protease: The Role of Residue 50 in Pathways for Amprenavir/Darunavir and Atazanavir Resistance.

Abstract 75
Introduction 76
Materials And Methods 80
Results 86
Discussion 94
Acknowledgements 97

Chapter V **99**

Discussion

Drug Resistance In Non-B Clade Hiv-1 Proteases 101
The Role Of Residue 50 In Altering Inhibitor Susceptibility 105
Conclusions 107

Appendix I **108**

Structural and Biochemical Studies on HIV-1 Clade C Protease.

Introduction 109
Materials And Methods 110
Results 115
Discussion 117

Appendix II **122**

Entropy and Enthalpy Compensation in a Drug Resistant Variant of HIV-1 Protease.

Abstract 123
Introduction 124
Methods 128
Results 136
Discussion 158
Acknowledgements 160

Appendix III **161**

Binding Thermodynamics Profiles of Inhibitors that Bind within the Substrate Envelope.

Introduction 162
Materials And Methods 168
Results And Future Directions 168

References **173**

LIST OF TABLES

TABLE I-1	HIV-1 protease substrates within the Gag and Gag-Pro-Pol polyproteins and Nef.	15
TABLE I-2	Common drug resistance substitutions in HIV-1 protease.	23
TABLE I-3	Amino acid sequence variations between the major HIV-1 clades.	25
TABLE II-1	Crystallographic data and statistics for CRF01_AE in complex with the substrate p1-p6.	35
TABLE II-2	Substrate-protease hydrogen bonds.	40
TABLE III-1	Crystallographic statistics for CRF01_AE and clade B variants in complex with DRV.	55
TABLE III-2	Binding thermodynamic parameters for NFV, DRV, and APV binding to CRF01_AE and clade B variants.	62
TABLE III-3	Enzyme-kinetics parameters for clade B and CRF01_AE WT and NFV-resistant variants.	64
TABLE III-4	Vitality values for clade B and CRF01_AE WT and NFV-resistant variants.	66
TABLE IV-1	Crystallographic statistics for WT, I50V and I50L variants in complex with APV, DRV and ATV.	83
TABLE IV-2	Binding thermodynamic parameters for APV, DRV and ATV binding to WT, I50V and I50L protease.	85

TABLE AI-1 Binding thermodynamic parameters for NFV, DRV and APV binding to Clade C and clade B protease variants. 116

TABLE AII-1 Crystallographic statistics. 133

TABLE AII-2 Thermodynamics of the binding of inhibitors to WT, ACT and FLAP+ variants of HIV-1 protease. 137

TABLE AIII-1 Binding thermodynamic parameters for APV, KC47 and AF57. 169

TABLE AIII-2 Binding thermodynamic parameters DRV analogs binding to wild type protease. 170

LIST OF FIGURES

FIGURE I-1	**A.** HIV-1 genomic classification. **B.** Global distribution of HIV-1 clades.	5
FIGURE I-2	Structural organization of an HIV-1 virion and the HIV-1 viral replication cycle.	8
FIGURE I-3	Amino acid sequence and structure of HIV-1 clade B protease.	12
FIGURE I-4	HIV-1 protease substrate envelope.	16
FIGURE I-5	HIV-1 protease inhibitors.	19
FIGURE I-6	**A.** Drug resistance mutations effectively prevent inhibitors from binding but allow substrates to bind. **B-D.** The inhibitor envelope.	21
FIGURE II-1	**A.** Amino acid sequence alignment of CRF01_AE protease with clade B protease. **B.** CRF01_AE protease in complex with the p1-p6 substrate.	30
FIGURE II-2	Structural comparisons of CRF01_AE and clade B p1-p6 protease.	36
FIGURE III-1	Structural comparisons of CRF01_AE and clade B protease structures.	47
FIGURE III-2	Structural analyses of CRF01_AE and clade B protease flap hinge and core regions.	57
FIGURE III-3	Protease-inhibitor hydrogen bonding interactions.	59
FIGURE III-4	Hydrogen bond network involving residue 88.	70
FIGURE IV-1	**A.** Chemical structures of amprenavir, darunavir and atazanavir. **B.** Homodimeric HIV-1 protease with positions 50 and 71 indicated by red spheres in each monomer.	77
FIGURE IV-2	Structural comparisons of WT, I50V and I50L protease variants.	88

FIGURE IV-3 Van der Waals interactions of residue 50 with APV, DRV and ATV with the protease variants. 92

FIGURE AI-1 **A.** Amino acid sequence alignment of clade B and C protease. 111
B. Amino acid sequence differences in clade C map to positions outside the active site.

FIGURE AI-2 **A.** L90M is a non-active site substitution. 119
B. The Leu90 side chain packs against the floor region of the active site in the clade B protease in complex NFV (PDB code: 3EKX).
C. *In silico* modeling of methionine at residue 90 shows that the methionine side chain can interact with the active site backbone.

FIGURE AII-1 **A.** Chemical structures of inhibitors. 126
B. Overview of the mutation sites of FLAP+ and ACT mutants mapped on an HIV-1 protease dimer.

FIGURE AII-2 Differences in binding energetics between variants. 139

FIGURE AII-3 Distribution of root mean squared deviations (RMSD) in Cα coordinates between inhibitor. 143

FIGURE AII-4 The active site region of WT, ACT and FLAP+ variants. 146

FIGURE AII-5 Double difference plots illustrating the WT vs FLAP+ and ACT variants with APV, ATV, DRV and SQV. 148

FIGURE AII-6 Histogram representation of the total inhibitor-protease van der Waals interaction energies for WT, ACT and FLAP+ variants. 152

FIGURE AIII-1 Substrate and Inhibitor Envelopes of HIV-1 Protease. 163

FIGURE AIII-2 Chemical structures of APV, DRV and corresponding inhibitor analogs. 166

LIST OF ABBREVIATIONS

AIDS	Acquired Immunodeficiency Syndrome
APV	amprenavir
ATV	atazanavir
CA-p2	capsid-p2 substrate
CRF	circulating recombinant form
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
DRV	darunavir
FDA	United States Food and Drug Administration
HAART	Highly Active Anti-Retroviral Therapy
HIV	human immunodeficiency virus
IDV	indinavir
IN	integrase
ITC	isothermal titration calorimetry
LPV	lopinavir
MA-CA	matrix-capsid substrate
NFV	nelfinavir
Nef	Negative Regulatory Factor
p1-p6^{gag}	p1-p6^{gag} substrate
p2-NC	p2-nucleocapsid substrate
PR-RT	HIV-1 protease-reverse transcriptase substrate

PI	HIV protease inhibitor
RFU	relative fluorescence units
RH-IN	ribonuclease H-integrase substrate
RMSD	root mean square deviation
RNA	ribonucleic acid
RT	reverse transcriptase
RTI	reverse transcriptase inhibitor
RTV	ritonavir
RT-RH	reverse transcriptase-ribonuclease H substrate
SIV	simian immunodeficiency viruses
SQV	saquinavir
TFP-p6^{pol}	trans frame protein-p6^{pol} substrate
TPV	tipranavir
UNAIDS	Joint United Nations Programme on HIV and AIDS

Standard one or three letter code was used to abbreviate amino acids. Mutated residues are abbreviated using position number and single letter code.

COPYRIGHT NOTICE

The following chapters were previously published and is reproduced with permission:

Chapter number	Publisher	License number
CHAPTER II	American Society for Microbiology	2423440853759

The following figures were reproduced with permission:

Figure number	Publisher	License number
FIGURE I-2C	Macmillan Publishers Ltd.	2416761220260
FIGURE 1-4	Elsevier Ltd.	2410950230487
FIGURE I-5B-D	Elsevier Ltd.	2410950230487
FIGURE AIII-1	Elsevier Ltd.	2410950230487

The following figures were reproduced: No permission required.

Figure number	Source
FIGURE I-1B	Wikipedia (http://en.wikipedia.org/wiki/Subtypes_of_HIV)
FIGURE I-2A-B	Wikipedia (http://en.wikipedia.org/wiki/Structure_and_genome_of_HIV)
FIGURE I-3C	Wikipedia (http://en.wikipedia.org/wiki/Aspartate_protease)

PREFACE

CHAPTER 2

Chapter 2 has been previously published as:

Bandaranayake R.M, Prabu-Jeyabalan M, Kakizawa J, Sugiura W, Schiffer CA. "Structural analysis of human immunodeficiency virus type 1 CRF01_AE protease in complex with the substrate p1-p6." (2008) *J. Virol*. 82(13):6762-6.

Co-author contributions: Junko Kakizawa did protein over expression for crystallization trials. I performed crystallization trials, processed and refined the crystal structure, performed data analysis, and prepared the figures. Moses Prabu-Jeyabalan helped collect x-ray diffraction data and refine the structure. Celia Schiffer and I wrote the manuscript.

CHAPTER 3

Chapter 3 is a submitted manuscript in press:

Bandaranayake R.M., Kolli M., King N.M., Nalivaika E., Héroux A., Kakizawa J., Sugiura W. and Schiffer C.A. "The Effect of Clade Specific Sequence Polymorphisms on HIV-1 Protease Activity and Inhibitor Resistance Pathways." (in press) *J. Virol.*

Co-author contributions: Nancy King performed isothermal titration calorimetry experiments (ITC) on clade WT-B protease. Madhavi Kolli solved the crystal structure of $DRV_{B\text{-}D30N/N88D}$ and performed ITC experiments on B-D30N/N88D protease. Ellen Nalivaika helped generate the WT-AE protease gene. Annie Héroux collected x-ray data on $DRV_{AE\text{-}N88S}$ crystals. I performed all experiments that generated all other data outlined in this chapter. Furthermore, I performed all the data analyses and prepared the figures. Celia Schiffer and I wrote the manuscript.

CHAPTER 4

Data presented in this chapter was a collaborative effort of members in the Schiffer Lab. The results are being prepared as a manuscript for publication. ITC experiments on WT protease was performed by Nancy King. Nancy King, Moses Prabu-Jeyabalan and Madhavi Nalam solved the WT and I50L crystal structures. Seema Mittal performed ITC experiments on I50V protease and crystallized and refined the APV_{I50V} complex and crystallized the DRV_{I50V} complex. I performed ITC experiments on I50L protease, crystallized and solved the ATV_{I50V} structure and refined the DRV_{I50V} structure. I did experiments that generated all the other data and performed all the data analysis. I wrote the manuscript and prepared the figures.

APPENDIX II

Appendix II is a manuscript being revised for submission:

King N.M., Prabu-Jeyabalan M., Bandaranayake R.M., Nalam M.N., Özen A., Haliloglu T. and Schiffer C.A. “Extreme Entropy-Enthalpy Compensation in a Drug Resistant Variant of HIV-1 Protease.”

My contributions to this manuscript were that I crystallized and solved the ATV_{FLAP+} structure and performed structural refinement of the APV_{FLAP+} and DRV_{FLAP+} complexes. I also helped perform hydrogen bond analyses on structures presented in this manuscript. Nancy King, Moses Prabu-Jeyabalan and Celia Schiffer wrote the manuscript.

CHAPTER I

Introduction

Since first identified in 1981, the Acquired Immunodeficiency Syndrome (AIDS) has become a serious public health concern and has resulted in major worldwide socio-economic consequences. While the causative agent for AIDS was identified as the human immunodeficiency virus (HIV), there is no cure available for the treatment of HIV infection and AIDS at present (Barre-Sinoussi et al., 1983; Popovic et al., 1984). However, over a period spanning nearly three decades, a vast body of knowledge has been gathered on the biology of HIV and has lead to the discovery of a number of therapeutics and treatment procedures which have been successful in slowing the progression of HIV infections to AIDS. However, until a proper cure for HIV infections is available, continued study of the virus and the disease is essential to formulate proper treatment strategies and thereby improve the life expectancy of infected individuals.

HIV and the AIDS pandemic

Over a span of nearly thirty years, HIV infection and AIDS have reached pandemic proportions with over 33 million people currently living with HIV (UNAIDS, 2009). With the lack of a cure for HIV/AIDS over 2 million AIDS related deaths are reported annually making HIV/AIDS a major global health concern (UNAIDS, 2009). Currently available highly active anti-retroviral therapy (HAART) has proved to be effective in improving the life expectancy and quality of life of individuals infected with HIV. With the increased accessibility to anti-retrovirals across the world, the number of new HIV infections have stabilized over the past few years (UNAIDS, 2009).

While HIV/AIDS statistics appear to have stabilized across the globe, the effects of the disease continue to be more prominent in resource poor settings of the world with

Sub-Saharan Africa and South and South-East Asia being the regions of highest HIV/AIDS prevalence in the world (Asamoah-Odei, Garcia Calleja, and Boerma, 2004; UNAIDS, 2009). Given that these regions are resource poor, setting up proper infrastructure to provide infected individuals access to affordable and continuous treatment is essential in order control the disease. Moreover, global efforts to educate people on disease prevention are vital to reduce the number of new HIV infections. Furthermore, a continued global effort is required to study and monitor the disease in order to develop better treatment strategies until a cure for HIV infections becomes available.

Biology of HIV

HIV was identified to be the causative agent for AIDS in the early 1980s (Barre-Sinoussi et al., 1983; Popovic et al., 1984). It is now widely accepted that HIV has its origins in the simian immunodeficiency virus (SIV) found in west and central Africa and crossed species at some point in time (Sharp, Robertson, and Hahn, 1995). To date, two types of HIV have been identified: HIV-1 and HIV-2. Evidence has been presented to show that HIV-1 is closely related to SIV_{CPZ} that infects chimpanzees while SIV_{SM}, which infects sooty mangabey monkeys, is thought to be the closest relative of HIV-2 (Gao et al., 1999; Hirsch et al., 1989; Huet et al., 1990).

The HIV-1 genome is categorized into groups and clades that include subtypes, sub-subtypes and circulating recombinant forms (CRFs) based on phylogenetic analyses of *envelope*, *gag*, and *pol* regions of the retroviral genome (Los Alamos Laboratory, 2010; Robertson et al., 2000). Through the course of HIV-1 infection in humans, three distinct lineages have been identified. Group M (main or major) is the predominant form throughout the world while Group O (outlier) and Group N (non-M-non-O) are rare forms identified only in western-central Africa (Figure I-1A) (Kuiken et al., 2000; Robertson et al., 2000). Recently, a new group termed as group P, was described with the virus thought to be closely related to SIV_{GOR} that infects gorillas (Plantier et al., 2009). To date, nine genetically distinct clades (A, B, C, D, F, G, H, J and K) and 43 CRFs have been identified within Group M (Los Alamos Laboratory, 2010). Clade B, which accounts for about 12% of HIV-1 infections globally, has been the focus of the majority of HIV-1 studies over the years. However, the majority of global HIV-1 infections are a result of non-B clade viruses with clade C being the most prevalent form (Figure I-1B) (Osmanov et al., 2002). Clinical observations in patient populations with different clades have indicated that the HIV-1 clade plays a role in viral transmission, progression to AIDS as well as viral susceptibility to anti-retrovirals (Ariyoshi et al., 2003b; Grossman et al., 2004b; Holguin et al., 2006; Hu et al., 1999; Kanki et al., 1999a; Kiwanuka et al., 2008).

HIV is a lentivirus belonging to the *Retroviridae* family. A mature HIV virion carries two copies of a single stranded positive-sense RNA genome that contains at least ten viral genes: *gag, pol, env, vif, vpr, tat, rev, vpu, nef* and *tev* (Kuiken et al., 2008). Of these genes, the *gag, pol* and *env* genes code for viral structural proteins and viral

FIGURE I-1

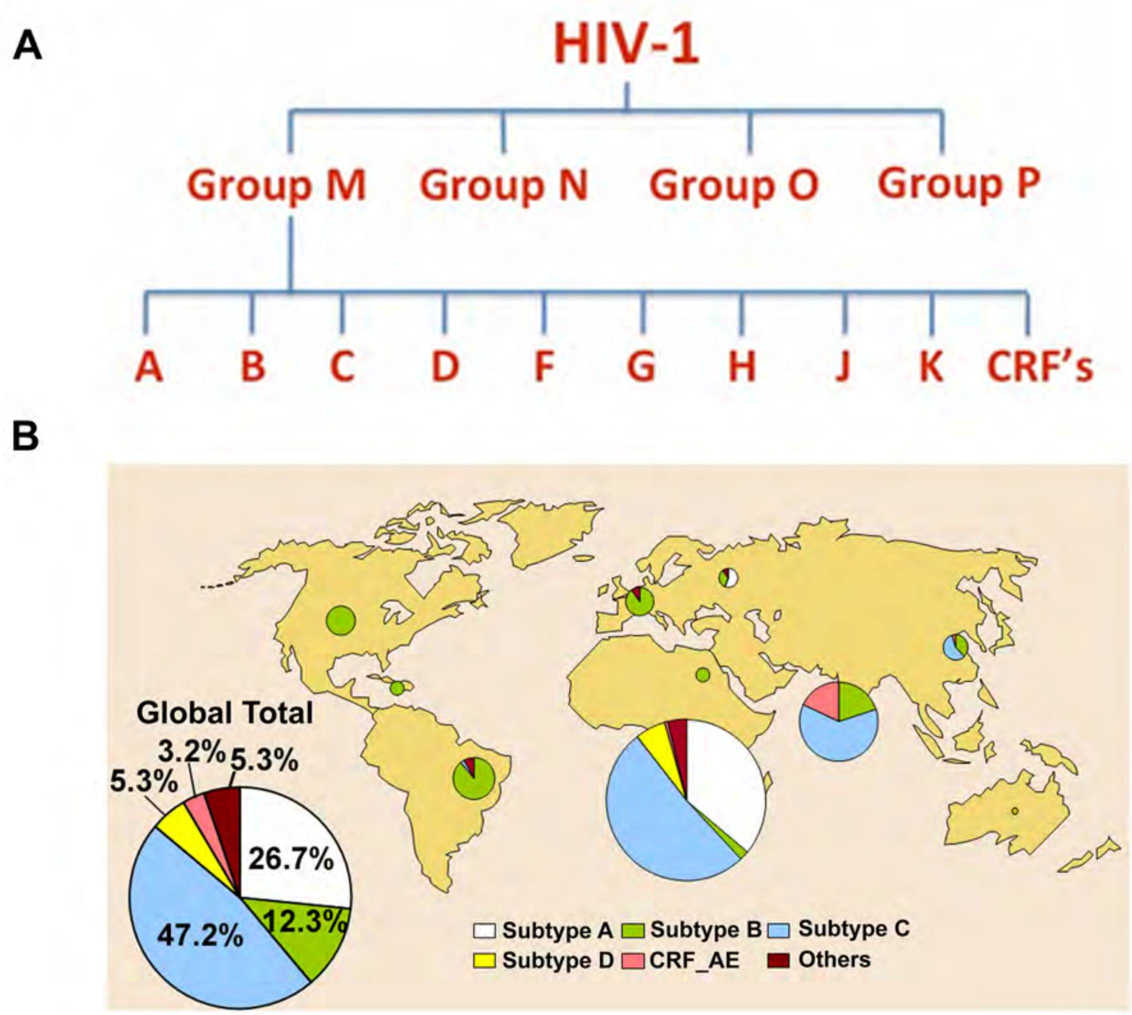

A
HIV-1
Group M
Group N
Group O
Group P
A
B
C
D
F
G
H
J
K
CRF's
B
Global Total
3.2%
5.3%
5.3%
26.7%
12.3%
47.2%
Subtype A
Subtype B
Subtype C
Subtype D
CRF_AE
Others

FIGURE I-1 A. HIV-1 genomic classification (Los Alamos Laboratory, 2010; Robertson et al., 2000). **B.** Global distribution of the most prominent HIV-1 clades.

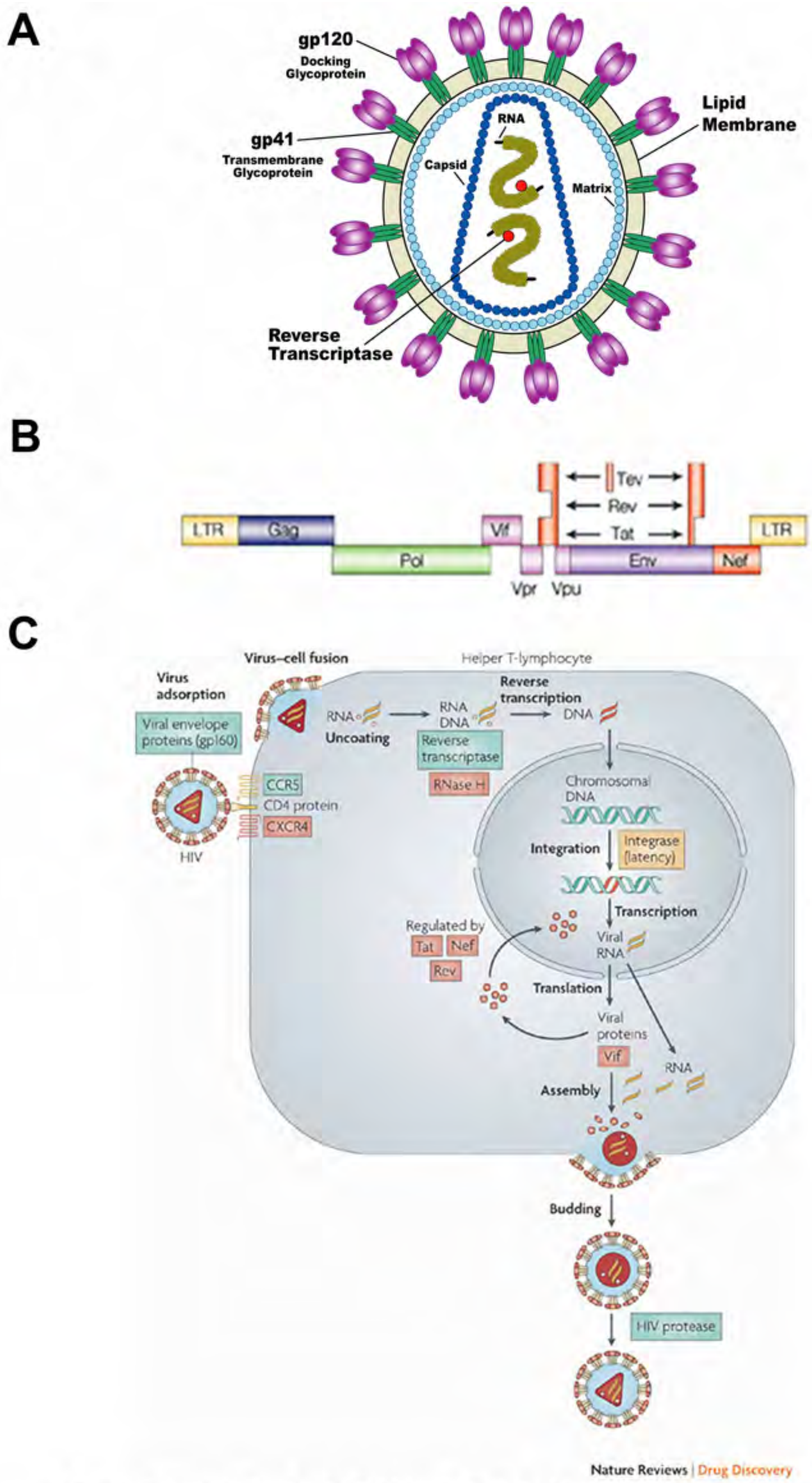
.

enzymes while the other genes code for accessory viral proteins (Kuiken et al., 2008; Subbramanian and Cohen, 1994; Wang et al., 2000). The virus primarily infects CD4+ T-lymphocytes, macrophages and dendritic cells (Alimonti, Ball, and Fowke, 2003; Chan and Kim, 1998; Knight, Macatonia, and Patterson, 1990; Koenig et al., 1986; Orenstein, 2001). The infection is initiated by the virus fusing with the host cell and is followed by the release of viral components into the host cell cytoplasm (Figure I-2). The initial fusing of the virus with the host cell is mediated through the viral surface glycoprotein gp120, which interacts with a CD4 molecule and a co-receptor molecule, CCR5 or CXCR4, on the host cell surface (Chan and Kim, 1998; Dimitrov, 1997; Wyatt and Sodroski, 1998). However, a number of recent studies have observed that HIV particles can enter cells through receptor mediated and non-receptor mediated endocytic pathways (Fackler and Peterlin, 2000; Miyauchi et al., 2009; Pauza, 1991).

Once the viral components enter the cell the viral reverse transcriptase (RT) reverse transcribes the positive-sense RNA genome into its complementary DNA sequence. The HIV RT is extremely error prone and with the lack of a proof reading mechanism, accumulates a high number of mutations in the newly synthesized DNA strand (Bebenek et al., 1993; Preston, Poiesz, and Loeb, 1988; Roberts, Bebenek, and Kunkel, 1988; Takeuchi, Nagumo, and Hoshino, 1988). Thus, the error prone reverse transcription of the viral genome results in the high degree of genomic variability observed in HIV.

The reverse transcribed viral genome is then translocated into the cell nucleus where it is integrated with the host genome by the viral enzyme integrase (IN). The

FIGURE I-2


A
gp120
Docking Glycoprotein
gp41
Transmembrane Glycoprotein
RNA
Capsid
Matrix
Lipid Membrane
Reverse Transcriptase
B
LTR
Gag
Pol
Vif
Vpr
Vpu
Tev
Rev
Tat
Env
Nef
LTR
C
Virus–cell fusion
Helper T-lymphocyte
Virus adsorption
Viral envelope proteins (gp160)
CCR5
CD4 protein
CXCR4
HIV
RNA
Uncoating
RNA
DNA
Reverse transcriptase
RNase H
Reverse transcription
DNA
Chromosomal DNA
Integration
Integrase (latency)
Transcription
Regulated by
Tat
Nef
Rev
Viral RNA
Translation
Viral proteins
Vif
RNA
Assembly
Budding
HIV protease
Nature Reviews | Drug Discovery

Reprinted by permission from Macmillan Publishers Ltd: Nature Medicine (Flexner, 2007). ©2007 www.nature.com/nm/.

Figure I-2. A. Basic structural organization of an HIV-1 virion. **B.** Components of the HIV-1 viral genome. **C.** HIV-1 replication cycle (Flexner, 2007).

integrated viral genome is transcribed into mRNA that is then translocated in to the cytosol. The mRNA transcripts are translated into Gag and Gag-Pro-Pol polyproteins, Env and other accessory proteins. These viral proteins, along with copies of the transcribed viral genome, are assembled at the host cell membrane and get packaged into immature virions that then bud out from the cell membrane (Navia and McKeever, 1990; Wang et al., 2000). While complete details of when the viral maturation process begins is not clear, there is evidence to show that once the immature virion has budded off, the viral protease (PR) cleaves the Gag and Gag-Pro-Pol polyproteins to release the structural components that then form the mature virus particle (Navia and McKeever, 1990).

A number of host factors, such as TRIM5α and APOBEC3G, that are capable of counteracting HIV infections have been identified (Lama and Planelles, 2007; Sheehy et al., 2002; Zheng, Lovsin, and Peterlin, 2005). However, action of these host factors are either insufficient to suppress infections in humans or are counteracted by HIV viral proteins. Thus, targeting the HIV replication cycle has been a key focus in anti-HIV drug development to slow down the progression of the infection to AIDS.

The HIV-1 reverse transcriptase and protease are two viral proteins to be successfully targeted by inhibitors very early in the HIV infection and AIDS pandemic. At present, there are thirteen reverse transcriptase inhibitors (RTIs) and nine protease inhibitors (PIs) that are being used successfully to prevent viral replication. Furthermore, in recent years, work on the fusion and integration stages of the viral replication cycle has resulted in two fusion and entry inhibitors and an integrase inhibitor that are now used in combination with RTIs and PIs. All currently available anti-HIV inhibitors are used as a part of highly active antiretroviral therapy (HAART) where a combination of two or more inhibitors from different inhibitor classes is used to treat patients. HAART has proved to be a potent method to prevent viral replication as well as to suppress the emergence of drug resistant viral strains (Bonfanti, Capetti, and Rizzardini, 1999; Shafer and Vuitton, 1999).

HIV-1 protease

The aspartyl protease of HIV-1 is an essential component in the replication cycle of the virus (Debouck and Metcalf, 1990; Kohl et al., 1988). HIV-1 protease is a homo-dimeric aspartyl protease with 99 amino acids in each monomer (Figure I-3A-B). The catalytic active site is formed at the dimeric interface with aspartic acid residues at position twenty-five in each monomer catalyzing the cleavage reaction (Berg, Tymoczko, and Stryer, 2007; Liu, Muller-Plathe, and van Gunsteren, 1996). While the complete details of the mechanisms by which HIV-1 protease cleaves substrates is not very clear, a general acid-base catalysis mechanism has been proposed based on a number of structural and enzyme kinetics studies (Brik and Wong, 2003; Hyland, Tomaszek, and Meek, 1991; Hyland et al., 1991; James et al., 1992; Parris et al., 1992; Pearl, 1985; Silva et al., 1996; Suguna et al., 1987; Veerapandian et al., 1992). The proposed mechanism suggests that only one aspartic acid residue is protonated within the active site while the other is deprotonated (Adachi et al., 2009). The mechanism also involves a water molecule that is coordinated within the active site by the catalytic aspartates. The charged aspartic acid residue is thought deprotonate the water molecule to activate it. The activated water

FIGURE I-3

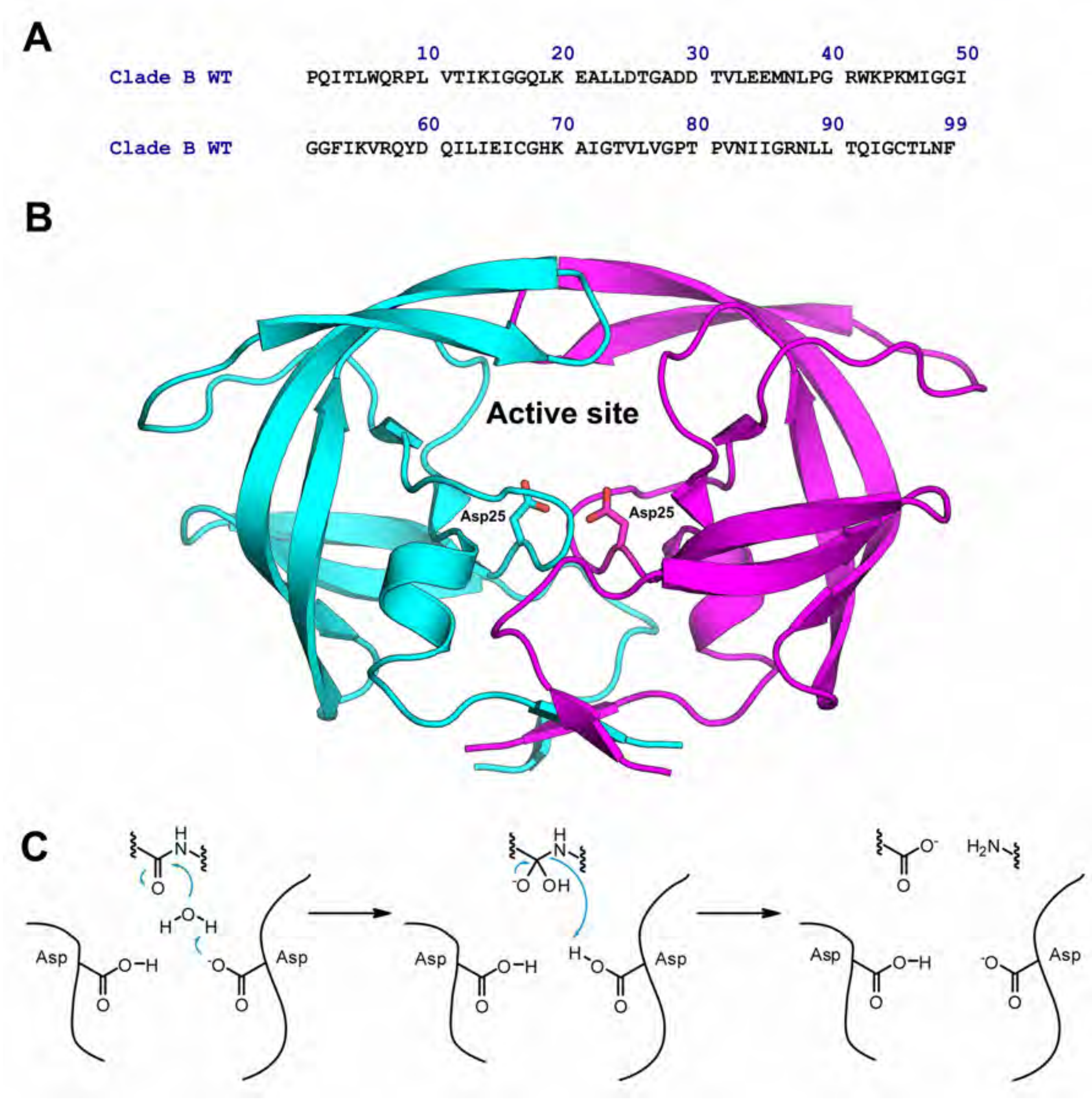

A
10 20 30 40 50
Clade B WT PQITLWQRPL VTIKIGGQLK EALLDTGADD TVLEEMNLPG RWKPKMIGGI
60 70 80 90 99
Clade B WT GGFIKVRQYD QILIEICGHK AIGTVLVGPT PVNIIGRNLL TQIGCTLNF
B
Active site
Asp25
Asp25
C
Asp
Asp
Asp
Asp
Asp
Asp

FIG. I-3. HIV-1 protease is homodimeric with 99 amino acid residues in each monomer. **A.** Amino acid sequence from HIV-1 clade B protease. **B.** Structure of HIV-1 protease. Each monomer in the dimeric complex is colored in cyan and magenta. The active site is indicated at the dimer interface and the side chains of the catalytic aspartic acid residues are also displayed. **C.** Cleavage mechanism proposed for HIV-1 protease.

molecule then disrupts the scissile bond through a nucleophilic attack on the carbonyl group (Figure I-3C).

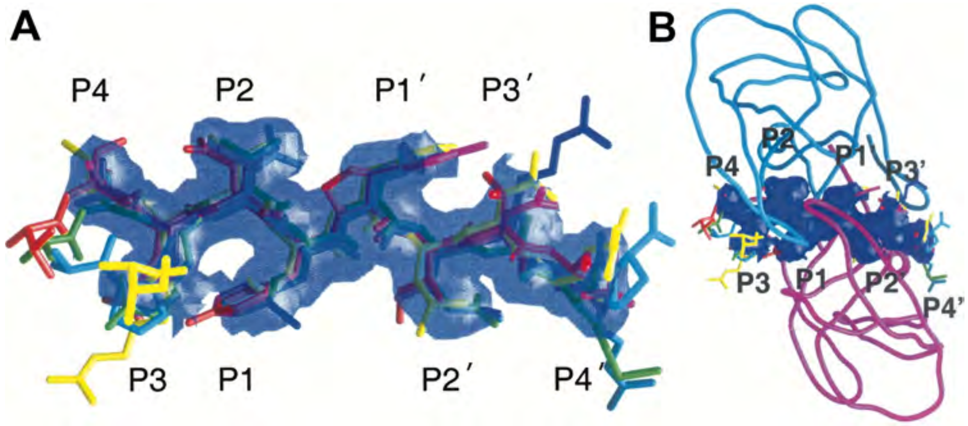
The HIV-1 protease cleaves at eleven asymmetric and non-homologous substrate cleavage sites within the Gag and Gag-Pro-Pol polyproteins and the Nef precursor protein (Table I-1) (de Oliveira et al., 2003). In addition to these cleavage sites, at least three auto-proteolysis sites are recognized within the protease (Rose, Salto, and Craik, 1993a). The active site of the protease has distinct pockets, S4 to S1 and S1’ to S4’, which accommodate substrate side chains P5 to P1 (towards the N-terminus from cleavage site) and P1’ to P5’ (towards the C-terminus from cleavage site) (Wlodawer and Erickson, 1993). Analysis of the substrate cleavage sites has revealed that the protease does not recognize a specific amino acid sequence (Debouck, 1992). Structural analyses of HIV-1 protease in complex with decamer substrate peptides have revealed the presence of a conserved shape taken by the substrates, termed the “substrate envelope” (Figure I-4A and 4B) (King et al., 2004a; Prabu-Jeyabalan, Nalivaika, and Schiffer, 2002). Structural data has further revealed that substrate recognition is mainly through conserved hydrogen bond interactions between the protease molecule and backbone atoms of the substrates (Prabu-Jeyabalan, Nalivaika, and Schiffer, 2002).

Given the crucial role played by HIV-1 protease in the viral replication cycle, it has been an important drug target in the effort to slow down the progression of HIV infections to AIDS (Debouck, 1992). Currently, there are nine FDA approved protease inhibitors (PIs), amprenavir (APV), atazanavir (ATV), darunavir (DRV), indinavir (IDV), lopinavir (LPV), nelfinavir (NFV), ritonavir (RTV), saquinavir (SQV) and tipranavir

TABLE I-1. HIV-1 protease substrates within the Gag and Gag-Pro-Pol polyproteins and Nef. The cleavage sites are indicated by the red lines.

	P5	**P4**	**P3**	**P2**	**P1**	**P1'**	**P2'**	**P3'**	**P4'**	**P5'**
MA/CA (p17/p24)	V	S	Q	N	Y	P	I	V	Q	N
CA/p2 (p24/p2)	K	A	R	V	L	A	E	A	M	S
p2/NC (p2/p7)	S	T	A	I	M	M	Q	K	G	N
NC/p1 (p7/p1)	E	R	Q	A	N	F	L	G	K	I
NC/TFP	E	R	Q	A	N	F	L	R	E	N
p1/p6^{gag}	R	P	G	N	F	L	Q	S	R	P
TFP/p6^{pol}	E	N	L	A	F	Q	Q	G	E	A
p6^{pol}/PR	T	S	F	S	F	P	Q	I	T	C
PR/RT	C	T	L	N	F	P	I	S	P	I
RT-RH(RT/p66)	G	A	E	T	F	Y	V	D	G	A
RH/IN (p66/IN)	I	R	K	V	L	F	L	D	G	I
Nef	P	D	C	A	W	L	E	A	Q	E

FIGURE I-4


A
P4
P2
P1′
P3′
P3
P1
P2′
P4′
B
P2
P4
P1′
P3′
P3
P1
P2′
P4′

Reprinted from Chemistry & Biology, Vol. 11, *"Combating Susceptibility to Drug Resistance: Lessons from HIV-1 Protease"*, Pages 1333-8, © 2004, with permission from Elsevier.

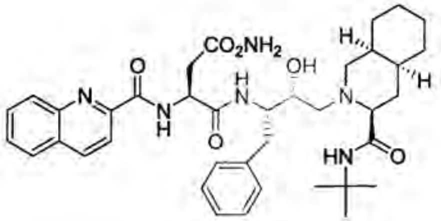
FIGURE I-4. A The substrate envelope calculated with GRASP (Nicholls, Sharp, and Honig, 1991a) from the overlapping van der Waals volume of four or more substrate peptides. The colors of the substrate peptides are red, matrix-capsid; green, capsid-p2; blue, p2-nucleocapsid; cyan, p1-p6; magenta, reverse-transcriptase-ribonucleaseH; and yellow, ribonucleaseH-integrase. **B.** The substrate envelope as it fits within the active site of HIV-1 protease. The α-carbon trace is of the CA-p2 substrate peptide complex (Prabu-Jeyabalan, Nalivaika, and Schiffer, 2000b).

(TPV), that are being used successfully to treat HIV infections as a part of HAART (Figure I-5) (De Meyer et al., 2005; Dorsey et al., 1994; Kaldor et al., 1997; Kempf et al., 1995; Kim et al., 1995; Roberts et al., 1990; Robinson et al., 2000; Sham et al., 1998; Turner et al., 1998). Except for TPV, all the PIs are peptidomimetics. These inhibitors are competitive inhibitors that bind within the active site of the protease and effectively prevent substrates from binding. The PIs act by mimicking a transition state of the substrates and thus, are able to maintain high affinity to the protease (De Clercq, 2009). Furthermore, the scissile bond of these peptidomimetics has been engineered in a way that it can no longer be cleaved by simple hydrolysis.

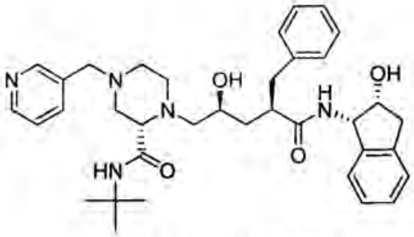
Drug resistance

Drug resistance is a major drawback in using all currently available anti-HIV drugs. Due to the error prone nature of HIV-1 reverse transcriptase, mutations can be introduced into the reverse transcribed viral genome. These mutations can manifest as drug resistant amino acid substitutions in all HIV drug targets. Thus, amino acid substitutions that result from mutations in the protease gene can weaken the protease-inhibitor interactions. Drug resistance mutations function by causing a change in molecular recognition (Figure I-6A). Wild type HIV-1 protease can bind both substrates and inhibitors. However, drug resistance mutations that accumulate within the protease prevent the effective binding of inhibitors but still allow substrates to bind the protease and be cleaved. Drug resistance substitutions develop against all currently available PIs (Table I-2).

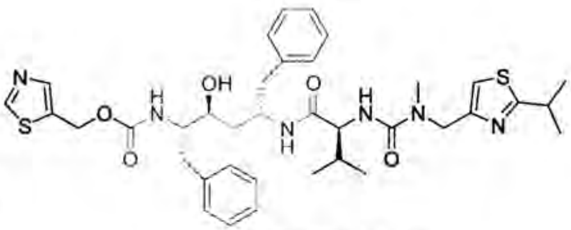
FIGURE I-5


CO2NH2
OH
H
N
N
H
N
H
H
HN
O
O
O

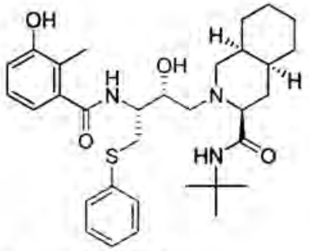
Saquinavir (SQV)


N
N
N
OH
H
N
OH
HN
O
O

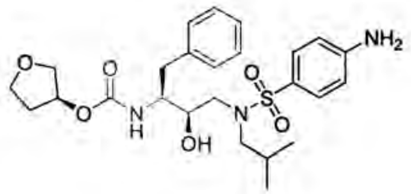
Indinavir (IDV)


N
S
O
H
N
OH
O
N
H
O
H
N
N
S
N
O

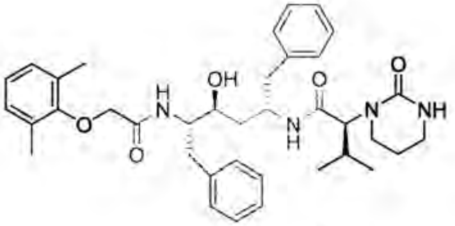
Ritonavir (RTV)


OH
H
N
OH
N
H
H
O
S
HN
O

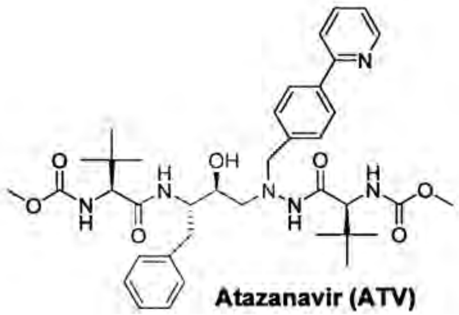
Nelfinavir (NFV)


O
O
O
N
H
OH
N
S
O
O
NH2

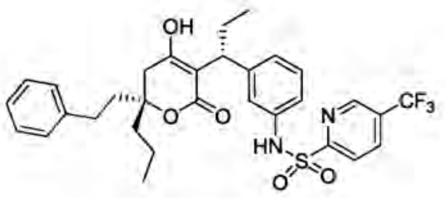
Amprenavir (APV)


O
O
H
N
OH
N
H
O
N
O
NH

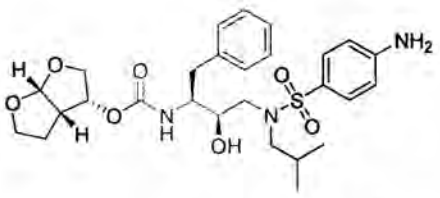
Lopinavir (LPV)


O
O
N
H
O
H
N
OH
N
N
H
O
H
N
O
O
N

Atazanavir (ATV)


OH
O
O
N
CF3
HN
S
O
O

Tipranavir (TPV)

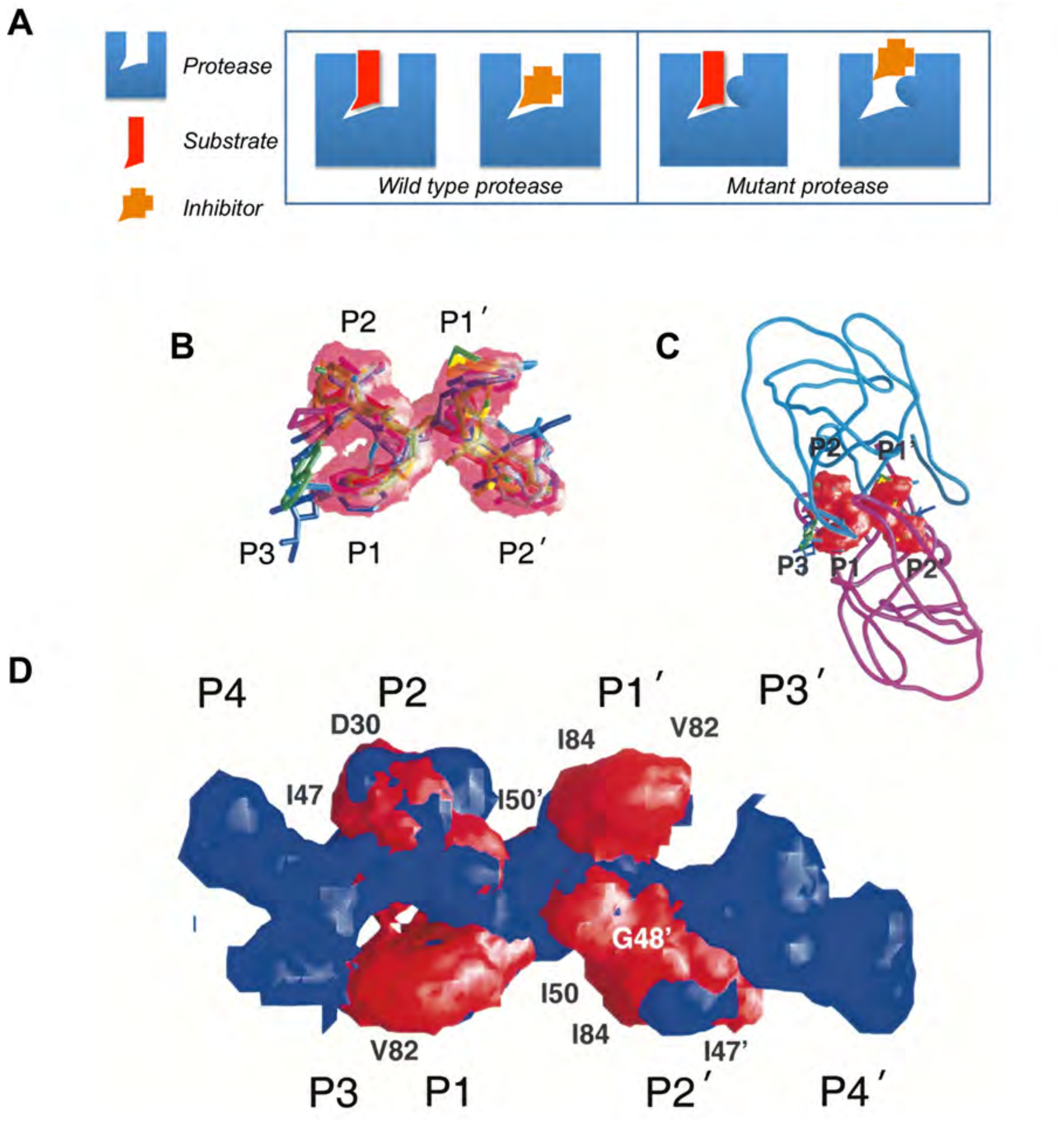

H
O
O
H
O
O
N
H
OH
N
S
O
O
NH2

Darunavir (DRV)

FIGURE I-5

FDA approved HIV-1 protease inhibitors that are used in HAART.

FIGURE I-6


A
Protease
Substrate
Inhibitor
Wild type protease
Mutant protease
B
P2
P1′
P3
P1
P2′
C
P2
P1′
P3
P1
P2′
D
P4
P2
P1′
P3′
D30
I84
V82
I47
I50’
G48’
I50
I84
I47’
V82
P3
P1
P2′
P4′

Panels 6B-5D are reprinted from Chemistry & Biology, Vol. 11, *“Combating Susceptibility to Drug Resistance: Lessons from HIV-1 Protease”*, Pages 1333-8, © 2004, with permission from Elsevier.

FIGURE I-6 A. Drug resistance is a change in molecular recognition. Drug resistance mutations effectively prevent inhibitors from binding but allow substrates to bind. **B.** The inhibitor envelope calculated from overlapping van der Waals volume of five or more of eight inhibitor complexes. The colors of the inhibitors are yellow, nelfinavir (NFV); gray, saquinavir (SQV); cyan, indinavir (IDV); light blue, ritonavir (RTV); green, amprenavir (APV); magenta, lopinavir (LPV); blue, atazanavir (ATV); and red, darunavir (DRV). **C.** The inhibitor envelope as it fits within the active site of HIV-1 protease. **D.** Superposition of the substrate envelope (blue) with the inhibitor envelope (red). Residues that contact the inhibitors where the inhibitors protrude beyond the substrate envelope and confer drug resistance when they mutate are labeled.

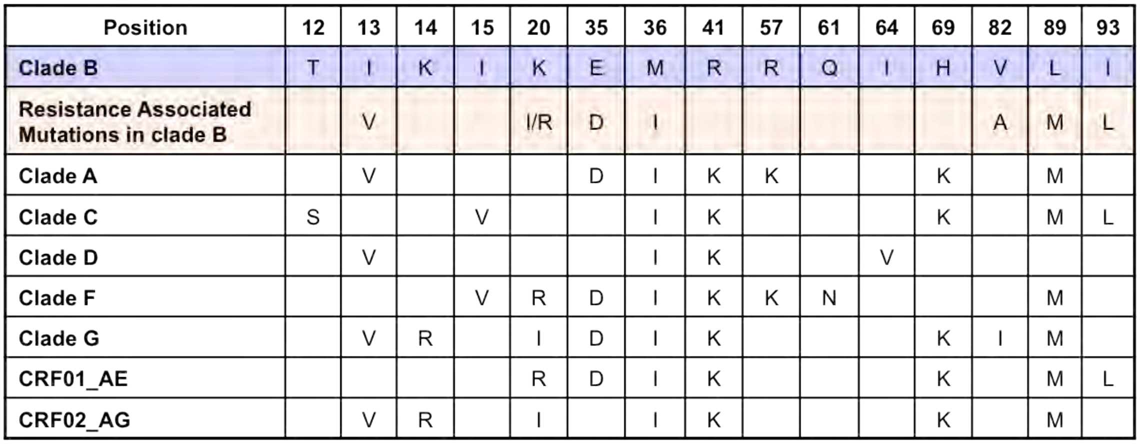
TABLE I-2. Common drug resistance substitutions (primary drug resistance substitutions) that are observed in patients failing currently available FDA approved inhibitors. Active site amino acid substitutions are indicated in bold font.

Inhibitor	**Common drug resistance mutations**
Amprenavir	**V32I, I47V, I50V, I54M/L, I84V,** L76V
Atazanavir	**I50L**, N88S
Darunavir	**V32I, I47V/A, I50V, I54M/L,** L76V, **I84V/A/C**
Indinavir	**V82A/T/F/S/M, M46I/L, I54V/T/A, I84V,** L90M
Lopinavir	**V82A/T/F/S, I54V/L/M/A/T/S, M46I/L, I50V, I47A/V, V32I,** L76V
Nelfinavir	**D30N/A**± N88D, **M46I/L**, L90M, N88S
Saquinavir	**G48V/M**, L90M, **I84V/A/C**
Tipranavir	**V82L/T, V32I, I47V, I54V/A/M, I84V**

The obvious drug resistance mutations occur within the active site and directly disrupt interactions between the protease and the inhibitor. Similar to the substrate envelope, an inhibitor envelope can be defined using the consensus volume taken up by PIs when bound within the active site (Figure I-5B and C). The majority of drug resistance mutations occur at positions within the protease where the inhibitor envelope protrudes outside the substrate envelope (Figure I-5D) (King et al., 2004a). However, a number of drug resistance substitutions occur outside the active site. The mechanism by which these substitutions modulate inhibitor binding is poorly understood. Furthermore, drug resistance substitutions are usually accompanied by secondary resistance substitutions and are mostly outside the active site. While several of these secondary substitutions enhance protease stability and activity, others enhance the effect of primary resistance substitutions. The extent to which these secondary mutations contribute to drug resistance is not clearly understood.

Drug resistance in HIV-1 protease has been extensively studied in clade B viral strains. The amino acid sequences of non-B clade proteases differ by about 10% when compared to the clade B sequence (Table I-3). The effect of these clade specific sequence variations on protease activity and drug resistance has not been looked at in detail. Several clade specific amino acid changes have been associated with protease inhibitor resistance in clade B protease (Table I-3). Given that there is clinical evidence to indicate that factors like virus transmission and disease progression are affected by the clade, it is important to fill the void in our understanding of how clade specific sequence variations contribute to inhibitor resistance.

TABLE I-3. Amino acid sequence variations between the major HIV-1 clades. The clade B amino acid residues at positions that are variable between clades are shown in the row shaded in blue. Amino acid substitutions that are associated with protease inhibitor resistance in clade B are indicated in the row shaded in pink.



Position	12	13	14	15	20	35	36	41	57	61	64	69	82	89	93
Clade B	T	I	K	I	K	E	M	R	R	Q	I	H	V	L	I
Resistance Associated Mutations in clade B		V			I/R	D	I						A	M	L
Clade A		V				D	I	K	K			K		M	
Clade C	S			V			I	K				K		M	L
Clade D		V					I	K			V				
Clade F				V	R	D	I	K	K	N				M	
Clade G		V	R		I	D	I	K				K	I	M	
CRF01_AE					R	D	I	K				K		M	L
CRF02_AG		V	R		I		I	K				K		M	

Summary

Emergence of drug resistance continues to be a major problem in using PIs. Inhibitor resistance can be defined as a change in molecular recognition where the protease no longer binds PIs efficiently but maintains its ability to bind and cleave substrates. The work presented in the following chapters is an attempt to understand the molecular mechanisms by which amino acid substitutions within HIV-1 protease confer inhibitor resistance. A combination of biochemical and biophysical techniques were used to explore how amino acid substitutions within the active site as well as those outside the active site affect inhibitor binding. Furthermore, these studies attempted to investigate the role of clade specific amino acid variations on inhibitor resistance. With limited biochemical and biophysical data available on non-B clade proteases, the findings presented in the following chapters add to our understanding of how clade specific sequence variations affect protease structure, function and inhibitor resistance. Overall, being able to understand how drug resistance occurs at a molecular level can provide valuable insights that can help in the design of robust PIs that are not susceptible to resistance mutations and aid in the formulation of better treatment strategies to slow down the progression of HIV infections to AIDS.

CHAPTER II

Structural Analysis of HIV-1 CRF01_AE Protease in Complex with the Substrate p1-p6

ABSTRACT

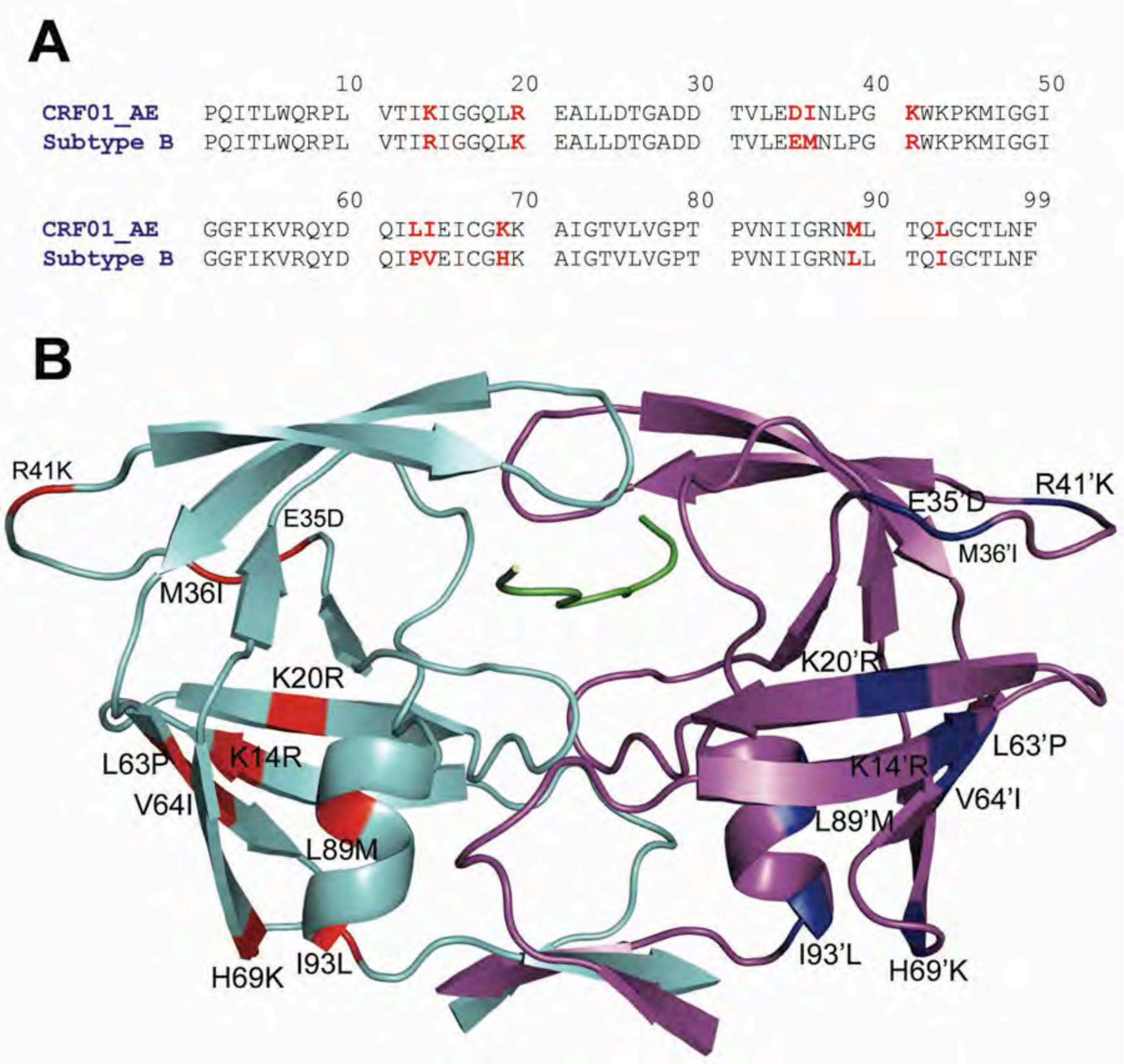
The effect of amino acid variability between human immunodeficiency virus type-1 (HIV-1) clades on structure and the emergence of resistance mutations in HIV-1 protease has become an area of significant interest in recent years. We determined the first crystal structure of HIV-1 CRF01_AE protease in complex with the p1-p6 substrate determined to a resolution of 2.8 Å. Hydrogen bonding between the flap hinge and protease core regions show significant structural rearrangements in CRF01_AE protease when compared to the clade B protease structure.

INTRODUCTION

Based on its genomic diversity, the HIV-1 has been classified into three groups, M (major), N (non-major) and O (other/outlier) (Robertson et al., 2000). Group M has been further defined into nine clades (clades A-D, F-H, J and K) and a number of sub-clades and circulating recombinant forms (CRFs). HIV-1 protease is one of the major proteins targeted for anti-HIV drug development. The *pol* gene which codes for protease differs by 10-15 % between clades (Gonzales, Machekano, and Shafer, 2001) and sequence diversity within HIV-1 clades has been an important area of study in recent years due to its possible role in altering resistance pathways within the protease (Ariyoshi et al., 2003a; Kantor et al., 2005). In particular HIV-1 CRF01_AE acquires Nelfinavir resistance via an alternative mutational pathway (Ariyoshi et al., 2003a), making the detailed study of non-B proteases strongly warranted.

Structural studies on clade B protease have lead to the successful development of a number of protease inhibitors (PIs). However, the majority of HIV-1 infections in the world result from non-clade B variants and there is limited evidence that non-clade B variants respond differently to currently available PIs (Clemente et al., 2006; Velazquez-Campoy, Vega, and Freire, 2002). Although a large number of clade B protease structures have been solved over the years, to date, very little structural information is available on non-B HIV proteases. The first non-B clade protease structures for clade F were recently published by Sanches *et al*. (Sanches et al., 2007b) and the crystallization of clade C protease-inhibitor complexes has been reported by Corman *et al.* (Coman et al., 2007). We present here, the crystal structure of an inactive HIV-1 CRF01_AE

FIGURE II-1



```
                      10          20          30          40          50
CRF01_AE   PQITLWQRPL  VTIKIGGQLR  EALLDTGADD  TVLEDINLPG  KWKPKMIGGI
Subtype B  PQITLWQRPL  VTIRIGGQLK  EALLDTGADD  TVLEEMNLPG  RWKPKMIGGI

                      60          70          80          90          99
CRF01_AE   GGFIKVRQYD  QILIEICGKK  AIGTVLVGPT  PVNIIGRNML  TQLGCTLNF
Subtype B  GGFIKVRQYD  QIPVEICGHK  AIGTVLVGPT  PVNIIGRNLL  TQIGCTLNF
```

A
B
R41K
E35D
M36I
K20R
L63P
K14R
V64I
L89M
H69K
I93L
R41'K
E35'D
M36'I
K20'R
L63'P
K14'R
V64'I
L89'M
I93'L
H69'K

FIGURE II-1. A. Amino acid sequence alignment of CRF01_AE protease with clade B protease. Positions where sequences differ are indicated in red. **B.** CRF01_AE protease in complex with the p1-p6 substrate (green). Amino acid changes in monomer A (cyan) are indicated in red and changes in monomer B (magenta) are indicated in blue.

protease variant (D25N) in complex with a decameric peptide corresponding to the p1-p6 cleavage site within the Gag and Gag-Pro-Pol polyprotein. CRF01_AE was one of the first CRFs to be identified and is now the predominant HIV-1 variant in Southeast Asia (McCutchan, 2006). The protease was derived from a Japanese patient isolate and has ten amino acid substitutions (R14K, K20R, E35D, M36I, R41K, P63L, V64I, H69K, L89M and I93L) compared to clade B (Figure II-1A-B).

MATERIALS AND METHODS

Crystallization and structure determination.

The CRF01_AE protease was expressed and purified as previously described (Prabu-Jeyabalan et al., 2004a). The protein was concentrated to 1.8 mg ml^{-1} using a 10 kDa molecular weight limit Amicon Ultra-15 centrifugal filter device. The decameric p1-p6 peptide (Arg-Pro-Gly-Asn-Phe-Leu-Gln-Ser-Arg-Pro, Quality Controlled Biochemicals Inc. (QBC), Hopkinton, MA) was solubilized in Dimethyl sulfoxide (DMSO) and was equilibrated with the protein with a 5 fold molar excess for 1 hour on ice. Crystals were grown over a reservoir solution consisting of 126mM phosphate buffer at pH 6.2, 63mM sodium citrate and ammonium sulfate in the range of 18-33% (Silva et al., 1996). A 2:1 volume ratio of substrate-protein solution and reservoir solution were combined to set up hanging drops with a final volume of 6 μL. The crystals were grown at ambient temperature.

Crystallographic data was collected under cryo conditions using a R-AXIS IV image plate mounted on a Rigaku rotating anode X-ray generator. The data were reduced and scaled using the programs DENZO and SCALEPACK respectively (Otwinowski and Minor, 1997). The structure solution was obtained by using a HIV-1 protease inhibitor complex (PDB code: 1T3R) with the program AMoRe (Navaza, 1994). Structure determination and refinement was carried out using programs within the CCP4 suite as previously described (Prabu-Jeyabalan et al., 2006b). Model building was carried out followed by real space refinement with the molecular graphics program COOT (Emsley and Cowtan, 2004). Refinement of initial models was done without the p1-p6 substrate and the peptide was built into the F_o-F_c density within the active site as the refinement progressed. A truncated p1-p6 peptide lacking ArgP5 and ProP4 was modeled into the active site as the *2Fo-Fc* and *Fo-Fc* maps indicated weak and discontinuous electron density at the N-terminus of the peptide. The ArgP4' of the p1-p6 peptide was modeled in as alanine since the electron density was not well defined to model in the arginine side chain. The stereochemical parameters of the final model were checked using PROCHECK (Laskowski et al., 1993). The CRF01_AE protease in complex with p1-p6 was determined to a resolution of 2.8 Å (Table II-1).

Protease structure comparison.

The inactive clade B protease in complex with p1-p6 (PDB code: 1KJF) was used for structural comparisons. The terminal regions (residues 1-9 and 86-99) from both monomers were used to superimpose the clade B structure onto the CRF01_AE complex. The superimposition was performed in a way that the orientation of the substrate peptide between the two structures was preserved. A double difference plot was generated to visualize structural differences between the two complexes. Distances between all the Cα atoms within the dimer were calculated for each complex and then the difference of the difference between the two dimers was plotted as a contour plot as previously described (Prabu-Jeyabalan et al., 2006b). The presence of significant contour peaks within the plot indicates regions that differ between the two structures.

RESULTS

Based on the Cα superimposition, the CRF01_AE and clade B structures display a high level of structural similarity to each other with a root mean square deviation (RMSD) of 0.37 Å (Figure II-2A). However, peaks within the double difference plot

TABLE II-1. Crystallographic data and statistics for CRF01_AE in complex with the substrate p1-p6

Parameter	**Value**
Resolution (Å)	2.8
Temperature (K)	100
Space group	$P6_1$
Cell dimensions	
a=b (Å)	62.1
c (Å)	82.1
Z^a	6
R_{merge} (%)	9.5
Completeness (%)	99.4
Total number of relections	45,495
Number of unique reflections	4,523
I/σ^I	8.4
RMSD[b] in:	
Bond lengths (Å)	0.008
Bond angles (Å)	1.2
R_{factor} (%)	19.9
R_{free} (%)	25.8

[a] Number of dimeric molecules in the unit cell.

[b] Root mean square deviation

FIGURE II-2

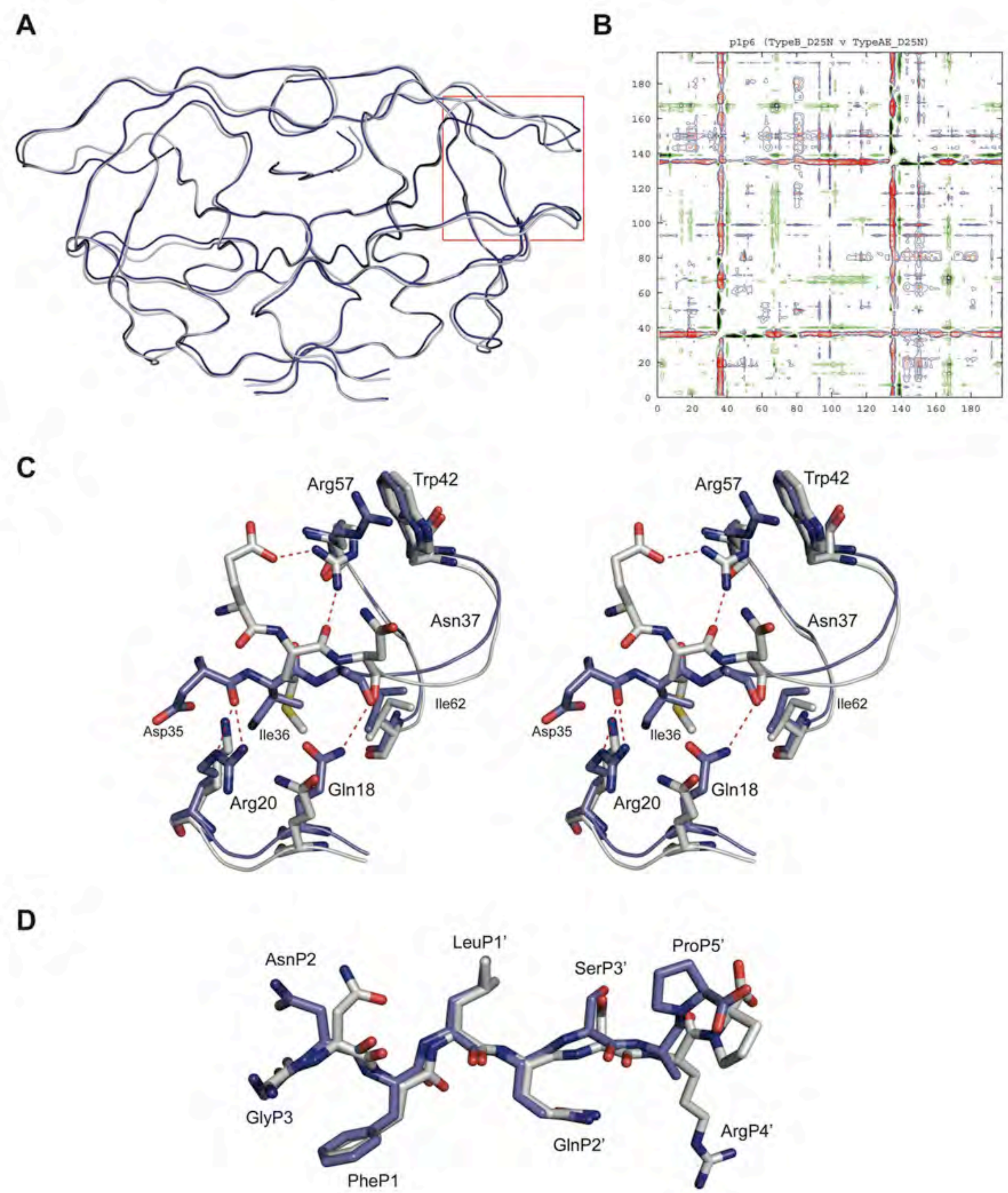

A
B
p1p6 (TypeB_D25N v TypeAE_D25N)
0
20
40
60
80
100
120
140
160
180
C
Arg57
Trp42
Asn37
Ile62
Asp35
Ile36
Gln18
Arg20
Arg57
Trp42
Asn37
Ile62
Asp35
Ile36
Gln18
Arg20
D
AsnP2
LeuP1'
SerP3'
ProP5'
GlyP3
PheP1
GlnP2'
ArgP4'

FIGURE II-2 A. Ribbon diagram superposition of CRF01_AE (blue) and clade B (gray) p1p6 structures. The flap hinge region is indicated by the red box. **B.** Double difference plot comparing the CRF01_AE and clade B p1p6 structures. Contours in the plot represent the degree of how close or far residues are between the structures being compared. Black indicates a distance of <-0.6 Å, red -0.59 Å and -0.3 Å, blue 0.3 Å and 0.59 Å and green >0.6 Å. **C.** Stereo view of the rearrangement of the flap hinge region of CRF01_AE structure (blue) compared to the clade B structure (gray). The Asn37 side chain in the CRF01_AE structure is disordered and has been modeled in as alanine. **D.** p1-p6 substrate conformation (blue – CRF01_AE, gray – clade B)

showed that the CRF01_AE complex had significant structural rearrangement at the flap hinge region (residues 33-39) and near the protease core region (residues 16-22) (Figure II-2B). These structural differences are present in both monomers of the complex. Closer examination of the Ile36 showed that its shorter side chain is stabilized through van der Waals interactions with the side chains of Asn18, Leu38 and Arg20 allowing the flap hinge region to pack closer to the core region when compared to longer M36 side chain in the clade B structure (Fig II-2C). The collapse of the flap hinge towards the core is further enhanced by the formation of a hydrogen bond between the carbonyl oxygen of Asp35 with NE or NH2 of the Arg20 side chain. This interaction causes the Asp35 side chain to flip inward towards the core region. In comparison, the longer Glu35 side chain in clade B is flipped out into the solvent allowing its OE2 oxygen to form a hydrogen bond with the side chain NH1 of Arg57. The positioning of the Arg57 side chain also allows the NH2 hydrogen to form a hydrogen bond with the carbonyl oxygen of Met36. The Arg57 side chain of the CRF01_AE structure is not involved in making any interactions with the flap hinge and packs against Trp42. These observations suggest that the flap hinge region of the CRF01_AE protease is likely to have reduced flexibility as a result of its tighter packing against the protease core region when compared to that of clade B.

Substrate conformation.

The p1-p6 peptide is bound within the active site in an extended conformation with the Phe-Leu cleavage site at P1 and P1’ positions oriented between the “catalytic” Asn25 residues. AsnP2 in the CRF01_AE structure adopts a conformation different to that of the clade B structure (Figure 2II-D). OG of SerP3’ also adopts a different orientation to that seen in the clade B structure. ProP5’ is rotated by 180° which causes the C-terminus of the peptide to kink towards the P4’ position whereas in the clade B complex the p1-p6 peptide adopts an extended conformation at the C-terminus. Despite changes in peptide conformation, the protease-substrate hydrogen bonding patterns show a high degree of similarity between the two structures with thirteen substrate-protease hydrogen bonds conserved between the CRF01_AE and clade B structures (Table II-2). However, the CRF01_AE structure makes four additional substrate-protease hydrogen bonds that are not seen in clade B structure. The AsnP2 side chain conformation allows OD1 to form a hydrogen bond with either Asp29 N or Asp30 N while ND2 forms a hydrogen bond with Asp30 OD2. Furthermore, the AsnP2 side chain confirmation allows it to make significant van der Waals interactions with Asp29 and Asp30. The hydrogen bond formed between SerP3’ OG and Arg8 NH1 is a result of SerP3’ OG adopting a different orientation to that seen in the clade B structure. Thus, when compared to the clade B structure the p1-p6 substrate appears to form better interactions with the CRF01_AE active site.

CONCLUSIONS

The structure described in this study is the first CRF01_AE protease structure as well as the first non-B HIV-1 protease-substrate complex structure to be reported to date. Arg20, Asp35, Ile36, Lys69, Met89 and Leu93 seen in the structure have been implicated

TABLE II-2 Substrate-protease hydrogen bonds

Substrate Atom	**Protease Atom**	**Distance (Å)**[a]	
		CRF01_AE	**Clade B**
GlyP3 N	Gly48 O	3.5	3.0
GlyP3 O	Asp29 N	3.0	2.8
AsnP2 OD1	Asp29 N	**2.8**	-
AsnP2 OD1	Asp30 N	**2.8**	-
AsnP2 ND2	Asp30 OD2	**3.1**	-
PheP1 N	Gly27 O	2.6	2.9
PheP1 O	Asn25 ND2	2.8	2.7
GlnP2’ N	Gly27’ O	2.9	3.0
GlnP2’ OE1	Asp29’ N	3.1	2.9
GlnP2’ OE1	Asp30’ N	2.7	2.7
GlnP2’ NE2	Asp30’ OD2	2.6	3.0
GlnP2’ O	Asp29’ N	2.8	3.1
Gln P2’ OE1	Asp30’ OD2	3.4	3.3
SerP3’ OG	Asp29’ OD2	2.8	3.4
SerP3’ OG	Arg8 NH1	**3.1**	-
SerP3’ N	Gly48’ O	3.1	3.0
SerP3’ O	Gly48’ N	3.2	2.8
ArgP4’ NE	Asp30’ OD1	NA[b]	3.2
ArgP4’ NE	Asp30’ OD2	NA[b]	3.0
ArgP4’ NH2	Gln58’ OE1	NA[b]	3.4

[a] Distances highlighted in bold are hydrogen bonds that are observed only in the CRF01_AE structure.

[b] Distances are not available as the ArgP4’ side chain is disordered.

as resistance associated mutations in clade B protease (Becker-Pergola et al., 2000; Julg and Goebel, 2005; Kantor and Katzenstein, 2003). While no significant structural changes were observed at Lys69, Met89 and Leu93, the substitutions Arg20, Asp35 and Ile36 in the CRF01_AE protease result in significant structural rearrangements of the flap hinge and core regions when compared to the clade B structure. We have observed a similar structural rearrangement in other CRF01_AE protease-inhibitor complexes (Chapter 3), which might be an indication that the interactions observed are unique to CRF01_AE protease.

Movement of the flaps is essential for substrate binding and the flap hinge and core regions play a key role in flap dynamics (Rose, Craik, and Stroud, 1998a; Scott and Schiffer, 2000; Todd and Freire, 1999a). The close packing observed between these regions in the CRF01_AE protease structure is likely to restrict flexibility and thereby affect flap dynamics. The protease molecule itself undergoes large conformational changes in order to facilitate substrate binding and product release following substrate cleavage (Foulkes-Murzycki, Scott, and Schiffer, 2007). Thus, reduced flexibility resulting from the packing of the flap hinge and core regions may have an effect on protease activity as well. Previously reported enzyme kinetics data on a CRF01_AE variant indicate that the active site specificity and catalytic efficiency are slightly lower for the CRF01_AE protease when compared to the clade B protease (Clemente et al., 2006). Furthermore, polymorphisms occurring within these regions are thought to affect binding affinities for PIs (Velazquez-Campoy et al., 2001). Therefore, the structural changes observed might influence how CRF01_AE protease interacts with PIs and may thereby alter levels of resistance to currently available inhibitors when compared to clade B protease.

AKNOWLEDGMENTS

I wish to thank Drs. Moses Prabu-Jeyabalan, Madhavi Nalam and Balaji Bhyravbhatla for helping me with the structural refinement. This work was supported by the National Institutes of Health grant 2R01-GM064347-06.

CHAPTER III

The Effect of Clade Specific Sequence Polymorphisms on HIV-1 Protease Activity and Inhibitor Resistance Pathways

ABSTRACT

The majority of HIV-1 infections around the world result from non-B clade HIV-1 strains. The CRF01_AE (AE) strain is seen principally in Southeast Asia. AE protease differs by ~10% in amino-acid sequence from clade B protease and carries several naturally occurring polymorphisms that are associated with drug resistance in clade B. AE protease has been observed to develop resistance through a non-active site N88S mutation in response to nelfinavir (NFV) therapy, whereas clade B protease develops both the active site mutation D30N and the non-active site mutation N88D. Structural and biochemical studies were carried out on wild type and NFV resistant clade B and AE protease variants. The relationship between clade specific sequence variations and pathways to inhibitor resistance was also assessed. AE protease has a lower catalytic turnover rate when compared to clade B, and also has weaker affinity for both NFV and darunavir (DRV). This weaker affinity may lead to the non-active site N88S variant in AE, which exhibits significantly decreased affinity for both NFV and DRV. The D30N/N88D in clade B resulted in a significant loss of affinity for NFV and to a lesser extent for DRV. Comparison of crystal structures of AE protease shows significant structural rearrangement in the flap hinge region compared with clade B and suggests insights into the altered pathways to NFV resistance. In combination, our studies show that sequence polymorphisms within clades can alter protease activity and inhibitor binding, and are capable of altering the pathway to inhibitor resistance.

INTRODUCTION

HIV-1 is classified into nine major clades (A, B, C, D, F, G, H, J, and K) and 43 circulating recombinant forms (CRFs) based on viral genomic diversity (Plantier et al., 2009; Robertson et al., 2000). The majority of HIV-1 infections across the globe result from non-B clade HIV-1 variants; clade B only accounts for ~12% of infections (Hemelaar et al., 2006). However, the development of currently available anti-HIV therapies has been based on the virology of clade B variants. In recent years, several studies have shown that there are clear differences between clades when it comes to viral transmission and the progression to AIDS, which raises questions about the effectiveness of the currently available anti-HIV therapies against the other clades and CRFs (Hu et al., 1999; Kaleebu et al., 2002b; Kanki et al., 1999b; Spira et al., 2003a).

HIV-1 protease has been an important drug target in the global effort to curb the progression to AIDS. However, the accumulation of drug-resistant mutations in the protease has been a major drawback in using HIV-1 protease inhibitors. The effects of mutations associated with drug resistance in HIV-1 clade B protease have been studied extensively over the years. For the most part, resistance-mutation patterns are very similar in HIV-1 clade B and non-B clade proteases (Kantor et al., 2005). However, several altered resistance pathways have been observed in non-B clade proteases compared with those of clade B protease (Ariyoshi et al., 2003a; Gomes et al., 2002; Grossman et al., 2004a; Nunez et al., 2002). Limited data are available on how sequence polymorphisms, some of which are associated with drug resistance in clade B protease, might influence the pathway to drug resistance in non-B clade proteases. Furthermore, very little is understood about how sequence polymorphisms in non-B clade proteases affect protease function and inhibitor binding.

HIV-1 CRF01_AE (AE) was the first CRF to be observed in patient populations and is seen principally in Southeast Asia (Carr et al., 1996; Gao et al., 1996; Murphy et al., 1993). AE protease differs by ~10% in amino-acid sequence from that of clade B protease (Figure III-1A). Interestingly, AE protease develops a different resistance pathway from that of clade B protease to confer resistance to the protease inhibitor nelfinavir (NFV) (Ariyoshi et al., 2003a). In patients infected with AE the protease predominantly acquires N88S in response to NFV therapy, whereas in patients with clade B the protease acquires D30N/ N88D.

In the present study, biochemical and biophysical methods were used to determine the effect of sequence polymorphisms in AE protease on enzyme activity and inhibitor binding. Through determination of crystal structures and analysis of changes in hydrogen bonding patterns, a structural rationalization is described for the two different pathways observed in clade B and AE proteases to attain resistance to NFV.

MATERIALS AND METHODS

Protease-gene construction

The clade B wild-type (B-WT) protease gene was generated as previously described (Prabu-Jeyabalan et al., 2004a) with the Q7K substitution introduced to prevent autoproteolysis (Rose, Salto, and Craik, 1993a). The AE wild type (AE-WT) protease gene was synthesized in fragments (Integrated DNA Technologies, Coralville, IA), with

FIGURE III-1

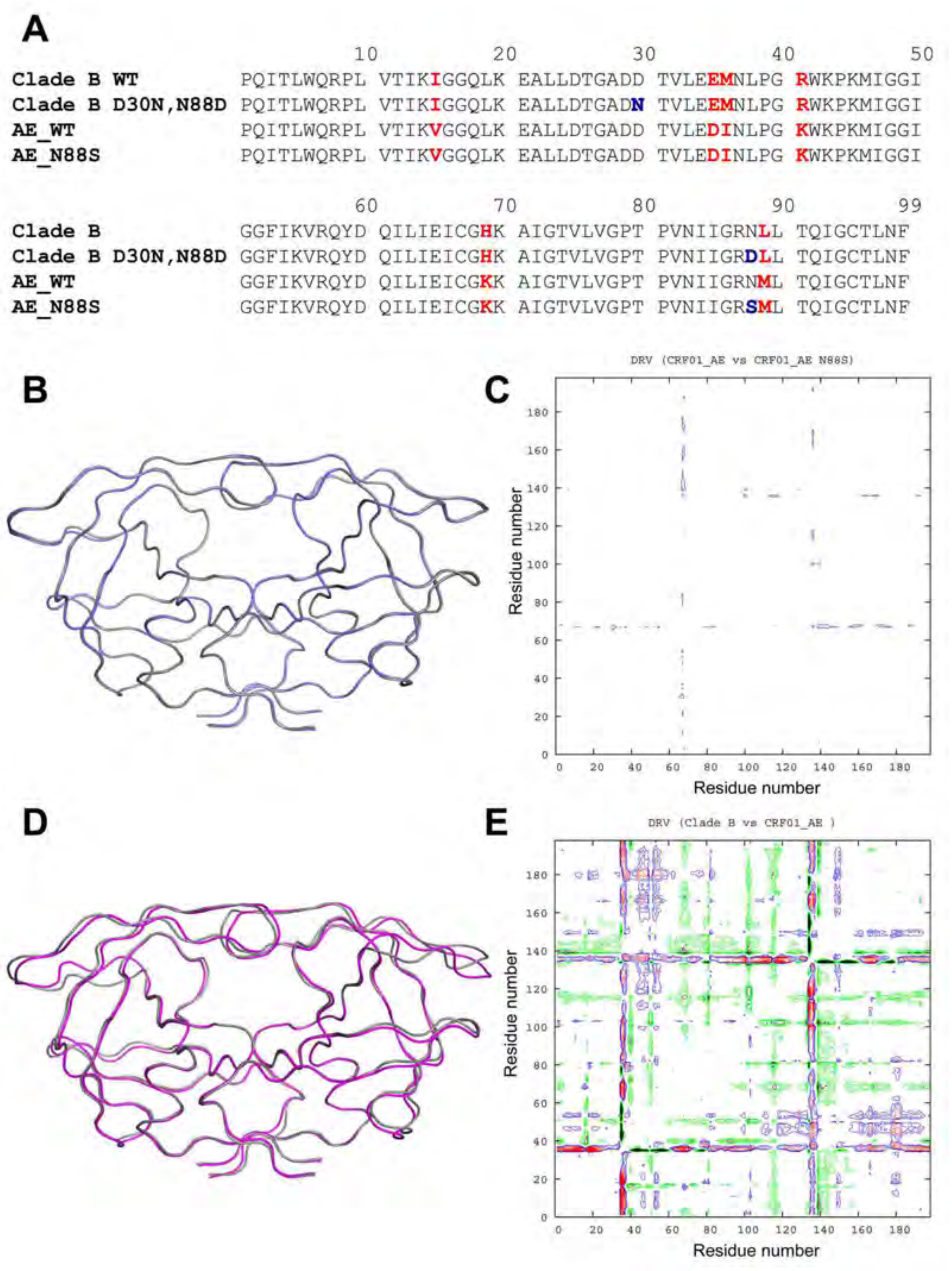

A
10 20 30 40 50
Clade B WT PQITLWQRPL VTIKIGGQLK EALLDTGADD TVLEEMNLPG RWKPKMIGGI
Clade B D30N,N88D PQITLWQRPL VTIKIGGQLK EALLDTGADN TVLEEMNLPG RWKPKMIGGI
AE_WT PQITLWQRPL VTIKVGGQLK EALLDTGADD TVLEDINLPG KWKPKMIGGI
AE_N88S PQITLWQRPL VTIKVGGQLK EALLDTGADD TVLEDINLPG KWKPKMIGGI
60 70 80 90 99
Clade B GGFIKVRQYD QILIEICGHK AIGTVLVGPT PVNIIGRNLL TQIGCTLNF
Clade B D30N,N88D GGFIKVRQYD QILIEICGHK AIGTVLVGPT PVNIIGRDLL TQIGCTLNF
AE_WT GGFIKVRQYD QILIEICGKK AIGTVLVGPT PVNIIGRNML TQIGCTLNF
AE_N88S GGFIKVRQYD QILIEICGKK AIGTVLVGPT PVNIIGRSML TQIGCTLNF
B
C
DRV (CRF01_AE vs CRF01_AE N88S)
Residue number
0
20
40
60
80
100
120
140
160
180
Residue number
D
E
DRV (Clade B vs CRF01_AE)
Residue number
0
20
40
60
80
100
120
140
160
180
Residue number

FIGURE III-1. (A) Amino acid sequence alignments of B-WT and AE-WT and NFV resistant mutants. Residue positions that differ between clade B and AE are indicated in red. NFV resistance mutations are indicated in blue. **(B)** Ribbon diagram superposition of $DRV_{AE\text{-}WT}$ (blue) and $DRV_{AE\text{-}N88S}$ (gray). **(C)** Double-difference plot comparing $DRV_{AE\text{-}WT}$ and $DRV_{AE\text{-}N88S}$. The color contours indicate distance differences of < -1.0 Å (black), -1.0 to -0.5 Å (green), 0.5 to 1.0 Å (blue) and > 1.0 Å (red). **(D)** Ribbon diagram superposition of clade $DRV_{AE\text{-}WT}$ (magenta) and clade $DRV_{B\text{-}WT}$ (gray). **(E)** Double-difference plot comparing $DRV_{AE\text{-}WT}$ and $DRV_{B\text{-}WT}$.

codons optimized for expression in *Escherichia coli*. The fragments were ligated to form the complete gene, which was then inserted into the pET11a expression vector (Novagen/EMD Chemicals, Gibbstown, NJ). The protease sequence was confirmed by DNA sequencing. The NFV-resistance mutations, N88S in AE (AE-N88S) and D30N/N88D in clade B (B-D30N/N88D) were generated by site-directed mutagenesis using the Stratagene QuikChange site-directed mutagenesis kit (Agilent Technologies, La Jolla, CA). Mutagenesis was confirmed by sequencing.

Protease-gene construction

The clade B wild-type (B-WT) protease gene was generated as previously described (Prabu-Jeyabalan et al., 2004a) with the Q7K substitution introduced to prevent autoproteolysis (Rose, Salto, and Craik, 1993a). The AE wild type (AE-WT) protease gene was synthesized in fragments (Integrated DNA Technologies, Coralville, IA), with codons optimized for expression in *Escherichia coli*. The fragments were ligated to form the complete gene, which was then inserted into the pET11a expression vector (Novagen/EMD Chemicals, Gibbstown, NJ). The protease sequence was confirmed by DNA sequencing. The NFV-resistance mutations, N88S in AE (AE-N88S) and D30N/N88D in clade B (B-D30N/N88D) were generated by site-directed mutagenesis using the Stratagene QuikChange site-directed mutagenesis kit (Agilent Technologies, La Jolla, CA). Mutagenesis was confirmed by sequencing.

Protein expression and purification

The clade B and AE variants were subcloned into the heat-inducible pXC35 expression vector (American Type Culture Collection (ATCC), Manassas, VA) and transformed into *E. coli* TAP-106 cells. Protein overexpression, purification and refolding was carried out, as previously described (King et al., 2002). Protein used for crystallographic studies was further purified with a Pharmacia Superdex 75 fast-performance liquid-chromatography column (GE Healthcare, Chalfont St. Giles, U.K.) equilibrated with refolding buffer (50mM sodium acetate pH 5.5, 10% glycerol, 5% ethylene glycol, and 5mM dithiothreitol).

Crystallization and structure determination

Protease solutions between 1.0 and 2.0 mg mL^{-1} were equilibrated with fivefold molar excess of NFV, darunavir (DRV) and amprenavir (APV) for 1 hour on ice. Crystals were grown over a reservoir solution consisting of 126-mM phosphate buffer (pH 6.2), 63-mM sodium citrate and 18% to 23% ammonium sulfate by the hanging-drop vapor-diffusion method. X-ray diffraction data for AE-WT were collected on BioCARS beamline 14-BMC at the Advanced Photon Source (Argonne National Laboratory, Argonne, IL) at a wavelength tuned to 0.9 Å with a Quantum 315 CCD X-ray detector (Area Detector Systems Corporation, Poway, CA). Diffraction data for AE-N88S were collected on beamline X29A at the National Synchrotron Light Source (Brookhaven National Laboratory, Upton, NY) at a wavelength tuned to 1.08 Å with a Quantum 315 CCD X-ray detector (Area Detector Systems Corporation). Data for the B-D30N/N88D variant was collected in-house on an R-Axis IV imaging-plate system (Rigaku Corporation, Tokyo, Japan) mounted on a rotating-anode X-ray source (Rigaku Corporation). All data were collected under cryo-cooled conditions.

The data were indexed, integrated and scaled using HKL/HKL-2000 software (HKL Research, Charlottesville, VA) (Otwinowski and Minor, 1997). Structure determination was done through molecular replacement with the program AMoRe (Navaza, 1994) and refinement were carried out using the CCP4 program suite (Collaborative Computational Project Number 4, 1994), as previously described (Prabu-Jeyabalan et al., 2006b). The tensor (T), libration (L), and screw (S) parameter files used in TLS refinement were generated using the TLS Motion Determination server (Painter and Merritt, 2006). Model building was carried out, followed by real-space refinement, with COOT molecular-graphics software (Emsley and Cowtan, 2004). Structure comparisons were made by superposing the structures using the Cα atoms of the terminal regions (residues 1 to 9 and 86 to 99) from both monomers. In the case of the AE complexes, which have multiple orientations for the inhibitor, each orientation was refined with an occupancy of 0.5. Only the orientation common with the clade B structures was used for analysis. Structures were visualized using PyMol molecular-graphics software (DeLano, 2002).

Double-difference plots were generated for AE and clade B protease structures to graphically visualize structural differences between the clades, as previously described (Prabu-Jeyabalan et al., 2006b). Briefly, distances between all Cα atoms within the dimer were calculated for each complex. A distance-difference matrix was then computed for each atom for a given pair of complexes. The distance-difference matrix was then plotted as a contour plot using gnuplot plotting software (Williams and Kelley, 1998b).

Nomenclature

The following nomenclature will be followed to refer to each crystal structure: $\text{Inhibitor}_{\text{protease variant}}$. Thus, DRV in complex with AE-WT, clade B-WT, AE-N88S and AE-D30N/N88D protein will be described as $\text{DRV}_{\text{AE-WT}}$, $\text{DRV}_{\text{B-WT}}$, $\text{DRV}_{\text{AE-N88S}}$, and $\text{DRV}_{\text{B-D30N/N88D.}}$ Prime notation is used to distinguish the two monomers in the protease dimer. For example, residue 30 from the first monomer would be referred to as Asp30 and the same residue from the second monomer would be referred to as Asp30'.

Isothermal titration calorimetry (ITC)

Binding affinities and thermodynamic parameters of inhibitor binding to clade B and AE variants were determined by ITC on a VP isothermal titration calorimeter (MicroCal, LLC, Northampton, MA). The buffer used for all protease and inhibitor solutions consisted of 10 mM sodium acetate (pH 5.0), 2% dimethyl sulfoxide, and 2 mM *tris*[2-carboxyethyl]phosphine. Binding affinities for all protease variants were obtained by competitive-displacement titration using acetyl-pepstatin as the weaker binder. A solution of 30–45 μM protease was titrated with 10-μL injections of 200 μM acetyl-pepstatin to saturation. The pepstatin was then displaced by titrating 36 8-μL injections of 200 μM amprenavir (APV) or NFV or 41 7-μL injections of 40 μM DRV. Heats of dilution were subtracted from the corresponding heats of reaction to obtain the heat resulting solely from the binding of the ligand to the enzyme. Data were processed and analyzed with the ITC data analysis module (Microcal) for Origin 7 data analysis and graphing software (OriginLab, Northampton, MA). Final results represent the average of at least two measurements.

Measurement of protease activity

Protease activity was assayed by following each variant's ability to hydrolyze the fluorogenic substrate HiLyte Fluor 488-Lys-Ala-Arg-Val-Leu-Ala-Glu-Ala-Met-Ser-Lys (QXL-520) (AnaSpec Inc., Fremont, CA) which corresponds to the HIV-1 Ca-p2 substrate. The assay was carried out in a 96-well plate, and the enzymatic reaction was initiated by adding 20 μL of a solution of 100–250 nM protease to 80 μL of substrate solution. The buffer used in all reactions consisted of 10 mM sodium acetate (pH 5.0), 2% dimethyl sulfoxide, and 2 mM *tris*[2-carboxyethyl]phosphine. Final concentrations in each experiment were 0–40 μM substrate and 20–50 nM protease. Accurate concentrations of properly folded active protease were determined by carrying out ITC experiments for each variant with acetyl-pepstatin as described in the previous section. Fluorogenic response to protease cleavage was monitored at 23°C using a Victor^3 microplate reader (PerkinElmer, Waltham, MA) by exciting the donor molecule at 485 nM and recording emitted light at 535 nM. Data points were acquired every 30 seconds. The data points in relative fluorescence units (RFU) were converted into concentrations using standard calibration curves generated for HiLyte Fluor 488 at each substrate concentration. In addition to converting RFUs to concentrations, generating calibration curves at each substrate concentration allowed us to correct for the inner filter effect (Copeland, 1996). Rates of each enzymatic reaction were determined from the linear portion of the data and were fitted against substrate concentrations to determine K_M and k_{cat} (catalytic-turnover rate) values using VisualEnyzmics enzyme-kinetics software

(SoftZymics, Princeton, NJ). Final results for each variant represent the average of at least two experiments.

In order to determine the biochemical fitness of a particular variant in the presence of a given inhibitor, vitality values were calculated using Equation 1, based on the vitality function previously described (Gulnik et al., 1995; Velazquez-Campoy et al., 2003).

$$Vitality = \frac{\left[K_d \times \left(k_{cat} / K_M\right)\right]_{mutant}}{\left[K_d \times (k_{cat} / K_M)\right]_{clade\ B\ WT}} \qquad \text{(Equation 1)}$$

The calculated vitality value for B-WT for a particular inhibitor would be 1.0 and vitality values greater than 1.0 would indicate that a given variant had a selective advantage over the same inhibitor while values lower than 1.0 would indicate that the variant did not have selective advantage.

RESULTS

Crystal structures

The AE WT and NFV-resistant clade B and AE variants were co-crystallized with NFV, DRV, and APV to reveal the structural basis for the altered NFV-resistance pathways. In addition the effects of background polymorphisms in AE-WT on inhibitor binding compared with clade B-WT were discerned. Crystals of AE protease in complex with NFV and APV did not diffract to high resolution, and therefore structural comparisons were carried out on AE and clade B protease in complex with DRV. The structure of $DRV_{B\text{-}WT}$ was solved previously in the lab and was used for structural comparisons (PDB code: 1T3R). Both DRV$_{\text{AE-WT}}$ and DRV$_{\text{AE-N88S}}$ crystallized with DRV bound in two orientations in the active site and therefore, the inhibitor was refined with an occupancy factor of 0.5. Crystallographic data and refinement statistics for DRV$_{\text{AE-WT}}$, DRV$_{\text{B-WT}}$, DRV$_{\text{AE-N88S}}$, and DRV$_{\text{B-D30N/N88D}}$ are given in Table III-1.

Structural comparisons were carried out for AE and clade B DRV complexes by pairwise structural superposition and double-difference plots (Figure III-1B-E). The DRV$_{\text{AE-WT}}$ and DRV$_{\text{AE-N88S}}$ complexes were structurally similar (Figure III-1B-C). Although the DRV$_{\text{AE-WT}}$ and DRV$_{\text{B-WT}}$ superimposed on each other very well (RMSD of 0.21 Å), there were clear and significant differences between the variants in the main chain at the flap-hinge region (residues 33–39) and 21the protease-core region (residues 16–22) (Figure III-1D and Figure III-2A-D). These differences were 22further evident by the presence of significant peaks in the double-difference plot (Figure III-1E). The Ile36 side chain in DRV$_{\text{AE-WT}}$ packs well against the core region through favorable van der Waals interactions and is shorter than the Met36 in DRV$_{\text{B-WT}}$ (Figure III-2C). In addition the shorter Asp35 in DRV$_{\text{AE-WT}}$ further enhances the packing by being flipped inward against the core, while in DRV$_{\text{B-WT}}$, the longer Glu35 is flipped outward into the solvent and forms a salt bridge with Arg57 (Figure III-2D). The packing of the flap-hinge and core regions in DRV$_{\text{AE-WT}}$ is further stabilized by a hydrogen bond between the carbonyl oxygen of Asp35 and Lys20 NZ that is not present in DRV$_{\text{B-WT}}$.

The Asp30’ side chain of DRV$_{\text{AE-WT}}$ forms a direct hydrogen bond with the N1 atom of DRV (Figure III-3C). In contrast, the Asp30’ side chain of DRV$_{\text{B-WT}}$ does not directly hydrogen bond with DRV but indirectly interacts with the N1 atom of DRV

TABLE III-1 Crystallographic statistics for CRF01_AE and clade B variants in complex with DRV.

	$DRV_{B\text{-}WT}$[a]	$DRV_{B\text{-}D30N/N88D}$	$DRV_{AE\text{-}WT}$	$DRV_{AE\text{-}N88S}$
Inhibitor	DRV	DRV	DRV	DRV
Resolution (Å)	1.2	2.15	1.96	1.76
Space group	$P\ 2_1\ 2_1\ 2_1$	$P\ 2_1\ 2_1\ 2_1$	$P\ 6_1$	$P\ 6_1$
Z[b]	4	4	6	6
Cell dimensions: a (Å)	54.9	50.9	62.2	61.9
b (Å)	57.8	57.7		
c (Å)	62.0	61.6	82.7	82.1
Total no. of reflections	302,022	108,838	89,284	79,445
No. of unique reflections	55,056	10,326	12,493	17,277
R_{symm} (%)	3.8	6.7	5.5	4.6
Completeness (%)	95.5	99.6	93.9	97.4
I/σ^I	25.0	9.6	11.2	19.6
R_{work} (%)	14.1	18.1	20.0	19.6
R_{free} (%)	17.9	23.6	25.9	23.9
RMSD[c]: Bond length (Å)	0.004	0.009	0.009	0.007
Bond angle	1.5	1.9	1.5	1.7
PDB code	1T3R	3LZV	3LZS	3LZU

[a] King *et al.*(King et al., 2004d) Surleraux *et al.*(Surleraux et al., 2005b)
[b] Number of dimeric molecules in the unit cell
[c] Root mean square deviation

FIGURE III-2

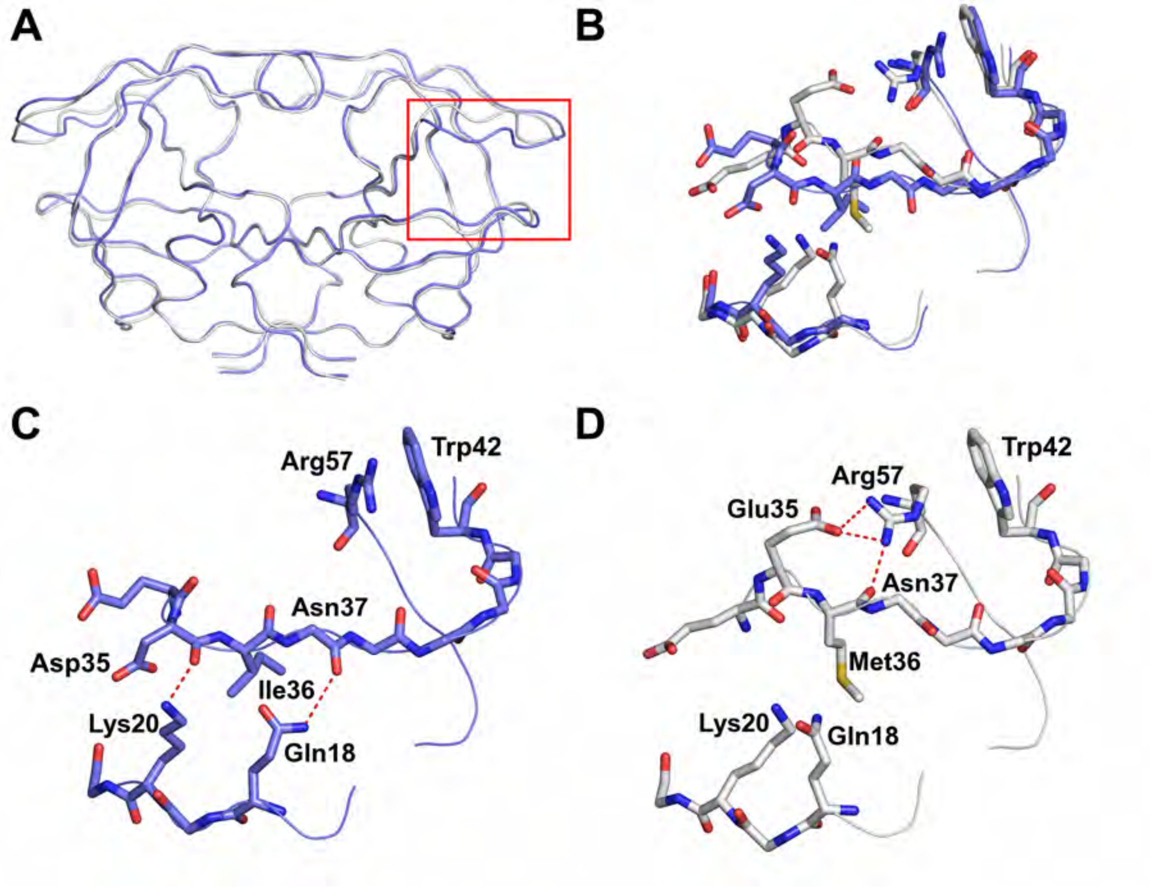

A
B
C
Arg57
Trp42
Asn37
Asp35
Ile36
Lys20
Gln18
D
Trp42
Arg57
Glu35
Asn37
Met36
Lys20
Gln18

FIGURE III-2. (A) Ribbon diagram superposition of $DRV_{AE\text{-}WT}$ (blue) and $DRV_{B\text{-}WT}$ (gray). The red box indicates the region of the protease molecule highlighted in panels B-D. **(B)** Structural rearrangement of the flap hinge and core regions between $DRV_{AE\text{-}WT}$ (blue) and $DRV_{B\text{-}WT}$ (gray). **(C)** Flap hinge and core regions of $DRV_{AE\text{-}WT}$. **(D)** Flap hinge and core regions of $DRV_{B\text{-}WT}$ protease. Hydrogen bond interactions are indicated in red dashed lines.

FIGURE III-3

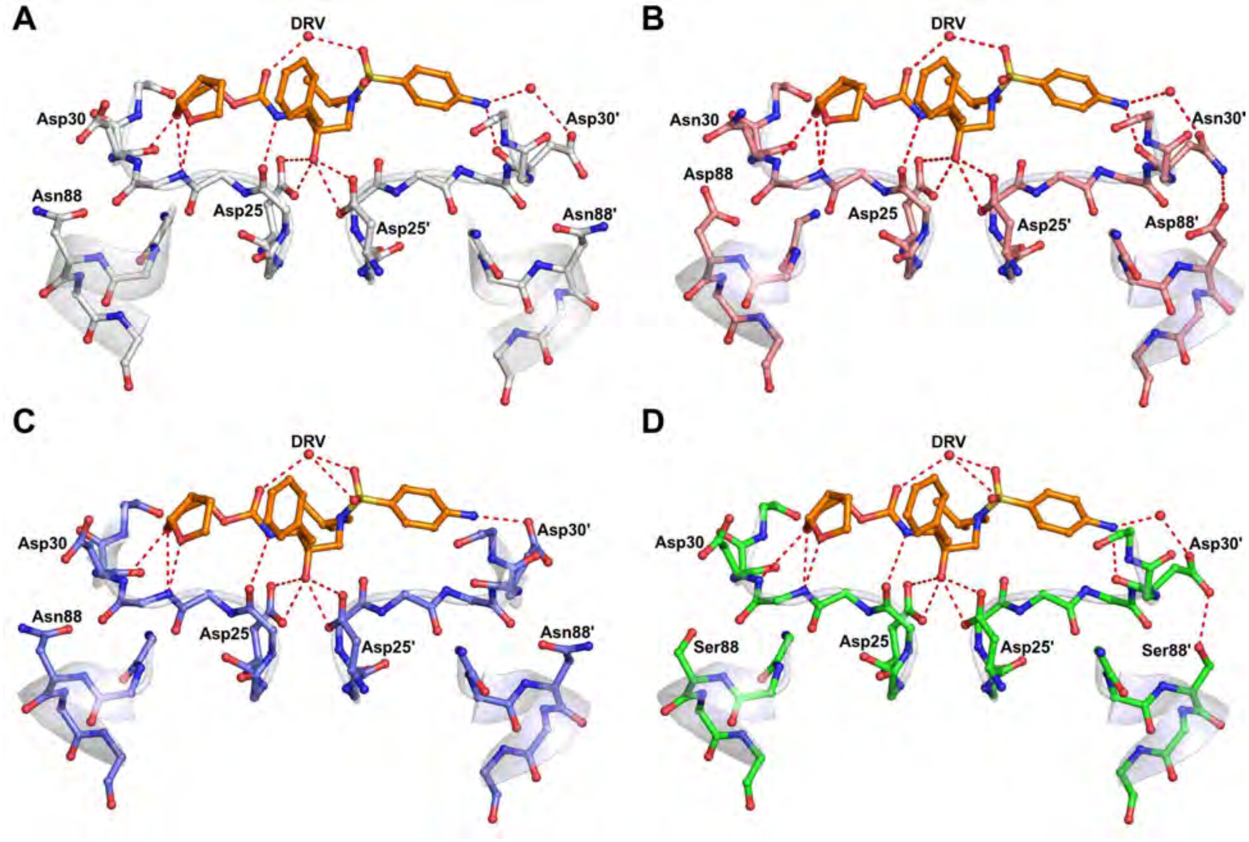

A
DRV
Asp30
Asp30'
Asn88
Asp25
Asp25'
Asn88'
B
DRV
Asn30
Asn30'
Asp88
Asp25
Asp25'
Asp88'
C
DRV
Asp30
Asp30'
Asn88
Asp25
Asp25'
Asn88'
D
DRV
Asp30
Asp30'
Ser88
Asp25
Asp25'
Ser88'

FIGURE III-3. Protease-inhibitor hydrogen bonding interactions. DRV is shown in orange and hydrogen bond interactions are indicated in red dashed lines. **(A)** $\mathrm{DRV_{B\text{-}WT}}$ (gray). **(B)** $\mathrm{DRV_{B\text{-}D30N/N88D}}$ (salmon). **(C)** $\mathrm{DRV_{AE\text{-}WT}}$ (blue). **(D)** $\mathrm{DRV_{AE\text{-}N88S}}$ (green).

through a water molecule mediated hydrogen bond network (Figure III-3A). Residue 30 of both NFV resistant variants also interacts with the N1 atom of DRV through water molecule mediated hydrogen bonding (Figure III-3B and 3D). However, in addition to this interaction, Asn30 of $DRV_{B\text{-}D30N/N88D}$ and Asp30 of $DRV_{AE\text{-}N88S}$ are oriented away from the active site enabling them to form hydrogen bonds with Asp88 and Ser88 respectively. In both cases, the NFV resistance mutations stabilize residue 30 away from the active site via hydrogen bonding.

Binding thermodynamics

To determine the effects of background sequence polymorphisms and NFV-resistance mutations on inhibitor binding, the binding thermodynamic parameters of NFV, DRV, and APV binding to WT and resistant AE and clade B variants were determined by isothermal titration calorimetry. The AE-WT protease had a 6.9-fold weaker affinity for NFV and 2.7-fold weaker affinity for DRV compared with the affinities of B-WT protease (Table III-2). This indicates that the AE-WT protease has an inherent weaker affinity for NFV and DRV.

No significant differences in the enthalpy of NFV binding were observed among any of the variants. Although the binding of DRV to all protease variants was enthalpically favorable, the enthalpic contribution was reduced with the AE variants (–10.1 kcal mol^{-1} for AE-WT and –5.1 kcal mol^{-1} AE-N88S) compared with the clade B variants (–12.1 kcal mol^{-1} for B-WT and –12.5 kcal mol^{-1} for B-D30N/N88D) (Table III-2). As expected the NFV resistant variants showed a significant reduction in binding affinity for NFV when compared to the wild type variants. With the AE-N88S variant the

TABLE III-2 Binding thermodynamic parameters for NFV, DRV, and APV binding to CRF01_AE and clade B variants.

Protease Variant	K_a (M^{-1})	K_d (nM)	K_d ratio	ΔH (kcal mol^{-1})	ΔΔH	–TΔS (kcal mol^{-1})	Δ(–TΔS)	ΔG (kcal mol^{-1})	ΔΔG
NFV									
Clade B WT	$(2.6 \pm 0.5) \times 10^{9}$	0.39 ± 0.07	1.0	4.4 ± 0.1	-	–17.0	-	–12.6 ± 0.1	-
Clade B D30N, N88D	$(1.2 \pm 0.4) \times 10^{8}$	8.1 ± 2.8	20.7 ± 8.1	6.7 ± 0.3	2.3	–17.5	0.5	–10.8 ± 0.2	1.8
CRF01_AE WT	$(3.7 \pm 1.0) \times 10^{8}$	2.7 ± 0.7	6.9 ± 2.1	5.0 ± 0.3	0.6	–16.6	0.9	–11.5 ± 0.2	1.1
CRF01_AE N88S	$(5.8 \pm 1.2) \times 10^{7}$	17.2 ± 3.5	44.1 ± 12.0	6.2 ± 0.7	1.8	–16.6	0.9	–10.4 ± 0.1	2.2
DRV									
Clade B WT	$(2.2 \pm 1.1) \times 10^{11}$	0.004 ± 0.002	1.0	–12.1 ± 0.9	-	–3.1	-	–15.2 ± 0.3	-
Clade B D30N, N88D	$(3.7 \pm 0.7) \times 10^{10}$	0.026 ± 0.005	6.5 ± 3.5	–12.5 ± 0.4	–0.4	–1.6	1.5	–14.2 ± 0.1	1.0
CRF01_AE WT	$(9.1 \pm 0.3) \times 10^{10}$	0.0109 ± 0.0003	2.7 ± 1.4	–10.1 ± 0.5	2.0	–4.6	–1.5	–14.7 ± 0.02	0.5
CRF01_AE N88S	$(1.1 \pm 0.8) \times 10^{10}$	0.087 ± 0.062	21.8 ± 18.9	–5.1 ± 3.6	7.0	–8.4	–5.3	–13.5 ± 0.4	1.7
APV									
Clade B WT	$(2.6 \pm 1.3) \times 10^{9}$	0.39 ± 0.20	1.0	–7.3 ± 0.9	-	–5.3	-	–12.6 ± 0.3	-
Clade B D30N, N88D	$(1.2 \pm 0.2) \times 10^{10}$	0.08 ± 0.01	0.2 ± 0.1	–10.2 ± 1.5	–2.9	–3.3	2.0	–13.5 ± 0.09	–0.9
CRF01_AE WT	$(3.1 \pm 0.2) \times 10^{9}$	0.32 ± 0.02	0.8 ± 0.4	–5.5 ± 0.3	–1.8	–7.3	–2.0	–12.70 ± 0.03	–0.1
CRF01_AE N88S	$(1.3 \pm 0.9) \times 10^{10}$	0.08 ± 0.06	0.2 ± 0.2	–5.0 ± 3.6	2.3	–8.6	–3.3	–13.6 ± 0.4	–1.0

affinity for NFV was reduced by 44.1-fold (K_d = 17.2 nM) and was far more significant than the D30N, N88D mutations in clade B protease, which reduced the affinity for NFV by 20.7-fold (K_d = 8.1 nM). Similarly the AE-N88S variant had a 21.8-fold weaker affinity (K_d = 0.087 nM) for DRV compared to a 6.5-fold weaker affinity (K_d = 0.026 nM) in B-D30N/N88D variant. Thus, the single N88S substitution in the AE protease has a profound effect on the binding of NFV and DRV.

In contrast to NFV and DRV, clade-specific sequence differences and NFV-resistance mutations had only minimal effect on the affinities for APV of both AE and clade B protease. Despite this, there were some differences in energy parameters. The binding of APV to the clade B variants appeared to be more enthalpically favorablecompared with that of the AE variants. This was compensated for by an increase in the entropic component to the binding energy for the AE proteases.

Protease activity and vitality

The enzyme-kinetic parameters determined for each clade B and AE variant with the Ca-p2 fluorogenic substrate analog are summarized in Table III-3. The K_M value for B-D30N/N88D protease (35.9 μM) was 2.1-fold greater than that of B-WT protease (16.7 μM). However, the K_M values for the AE protease variants (17.5 μM for AE-WT and 19.0 μM for AE-N88S) were similar to that of B-WT protease. The turnover rate for B-D30N/N88D protease (k_{cat} = 0.13 sec^{-1}) was significantly lower than that of B-WT (k_{cat} = 1.79 sec^{-1}) protease. Turnover rates for AE-WT (k_{cat} = 0.7 sec^{-1}) and AE-N88S (k_{cat} = 0.2 sec^{-1}) were 2.5 and 8.5-fold lower, respectively, than that of clade B-WT. The k_{cat}/K_M values, or catalytic efficiency, for B-D30N/N88D and AE variants were lower than that

TABLE III-3. Enzyme-kinetics parameters for clade B and CRF01_AE WT and NFV-resistant variants.

	B WT	B D30N, N88D	AE WT	AE N88S
K_M (μM)	16.7 ± 6.0	35.9 ± 0.1	17.5 ± 4.0	19.0 ± 0.8
k_{cat} (sec^{-1})	1.79 ± 0.28	0.13 ± 0.09	0.70 ± 0.08	0.20 ± 0.02
k_{cat}/K_M ($\text{sec}^{-1}\mu\text{M}^{-1}$)	0.11 ± 0.04	0.004 ± 0.002	0.04 ± 0.01	0.010 ± 0.001

of B-WT protease. Therefore, the reduction in catalytic efficiency of the B-D30N/N88D protease compared with that of B-WT protease resulted from the combined effects of the K_M and k_{cat} values. However, for the AE variants the smaller turnover rates alone were responsible for the reduced catalytic efficiencies. Overall, these results indicate that the polymorphic sequence differences in AE protease can alter the activity profile of the enzyme when compared to the clade B protease.

Vitality values were calculated to determine if the protease variants had a selectie advantage over NFV, DRV and APV, AE-WT and AE-N88S protease had calculated vitality values of 2.5 and 4.0 respectively when compared with B-WT for NFV (Table III-4). However, vitality values for DRV were not significantly different from that of B-WT protease. Vitality values for APV were significantly lower in all variants when compared to that of B-WT protease. These results indicate that AE-WT may have a selective advantage over NFV when compared to B-WT but the AE variants may not have a significant selective advantage against DRV or APV

DISCUSSION

Although the majority HIV-1 patients are infected with non-B forms of the virus, molecular studies have been predominantly carried out on clade B variants. The AE protease has several polymorphisms that are associated with inhibitor resistance in clade B. AE also shows altered patterns of drug resistance to NFV. We have performed detailed studies to determine the effects of sequence polymorphisms on enzyme structure,

TABLE III-4 Vitality values for clade B-D30N, N88D, AE-WT and AE-N88S variants.

Inhibitor	B D30N, N88D	AE WT	AE N88S
NFV	0.8 ± 0.6	2.5 ± 1.4	4.0 ± 1.9
DRV	0.2 ± 0.2	0.9 ± 0.7	2.0 ± 2.0
APV	0.007 ± 0.005	0.3 ± 0.2	0.02 ± 0.02

activity and inhibitor binding. These analyses lead to a structural rationalization for the altered pathways for drug resistance. AE-WT protease has an inherently weaker affinity for NFV and DRV than B-WT as is evident from the thermodynamic data (Table III-2). The weaker affinity observed for NFV is consistent with previously published data for another AE protease variant (Clemente et al., 2006) as well as for clade A protease (Valzaquez-Campoy et al., 2001), which is closely related. The inherent weaker affinity for NFV likely allows the AE protease to gain resistance to NFV through a single non-active site substitution, N88S. The clade B protease, by contrast, which has a relatively stronger affinity for NFV, requires a combination of an active site (D30N) and non-active site mutation (N88D) to gain NFV resistance. The ability of the AE-N88S protease to maintain affinity for substrates is evident from our enzyme-kinetics data (Table III-4), in which the K_M value for AE-N88S was comparable to that of AE-WT and B-WT protease. The K_M value for clade B-D30N/N88D, on the other hand, was significantly worse than that of the B-WT, likely reflecting the effect of the altered active site.

As an active site residue, Asp30 plays a key role in substrate recognition by interacting with substrates through side-chain–mediated hydrogen bonds with the MA-CA, CA-p2, p1-p6 and p2-NC cleavage sites (Prabu-Jeyabalan, Nalivaika, and Schiffer, 2002). Therefore, as is evident from our enzyme kinetics data, the D30N/N88D mutations in clade B will likely affect substrate binding and processing. Several studies have observed substrate co-evolution in instances where the protease mutates active site residues in order to confer inhibitor resistance (Kolli, Lastere, and Schiffer, 2006; Kolli et al., 2009). However, since the AE-N88S protease variant has no active site mutations, the enzyme retains the ability to effectively recognize substrates while conferring NFV resistance. Therefore, the presence of the N88S substitution in AE protease is unlikely to induce co-evolution of the viral substrates in order maintain effective enzymatic activity.

Despite having K_M values that were comparable to that of B-WT protease, both AE-WT and AE-N88S had significantly lower catalytic-turnover rates (k_{cat}) than that of the B-WT protease (Table III-3). As a result, the catalytic efficiency of the AE variants is lower than that of the B-WT protease. The lower turnover rates of the AE variants could be a direct result of the reduced flexibility of the flap-hinge (residues 33–39) and core regions (residues 16–22) of the protein. Molecular-dynamics studies have revealed that hydrophobic sliding of the core region facilitates substrate binding through the opening of the active site (Foulkes-Murzycki, Scott, and Schiffer, 2007). The unique hydrogen bonds observed between the flap hinge and the core in the AE variants alter movement of the core, thus impacting the ability of the active site to open up for substrate binding and product release. Based on our enzyme kinetics data, this altered flexibility of the flap hinges in the AE variants has little effect on substrate binding but rather affects the catalytic step of the reaction by slowing down product release.

The higher vitality value observed for AE-WT with NFV provides supporting evidence for the reduction in the efficacy of NFV against the AE protease compared with that of clade B (Table III-4). This result is consistent with previous vitality calculations for the clade A protease (Valzaquez-Campoy et al., 2001). In addition, these results further highlight the idea that background polymorphic sequence variations in the AE protease can affect the potency of NFV. The sub-optimal efficacy of NFV against the

AE-WT protease likely permits a non-active site variant, AE-N88S, to emerge over variants with active site mutations to effectively confer resistance to NFV.

The impact on other inhibitors, however, is complex. APV and DRV are chemically very closely related compounds, and similar susceptibility and resistance patterns have been observed for these two inhibitors (Parkin et al., 2007). However, this pattern is not evident for this series of resistant variants. Both the N88S in the AE and D30N, N88D in the clade B proteases result in hypersusceptibility to APV. Similar results have been observed also for a B-N88S protease variant (Masquelier et al., 2008; Ziermann et al., 2000). In contrast, these same substitutions in the protease give rise to even greater resistance to DRV. However, since DRV presents a greater genetic barrier to resistance than APV (Poveda et al., 2007), the *in vivo* implications of weaker affinity for DRV in the AE variants are likely negligible. Indeed our calculated vitality values indicate that DRV maintains its potency against the AE variants despite having a weaker affinity for AE-WT and AE-N88S relative to clade B protease.

A close look at the $\text{NFV}_{\text{B-WT}}$ protease complex reveals an important interaction between the Asp30 residue side chain and the inhibitor bound in the active site. (PDB code: 3EKX, Figure III-4B). One of the side chain oxygen atoms of Asp30 forms a direct hydrogen bond with the O38 atom of NFV. Our crystal structures of the NFV resistant variants show that N88S in AE and N88D in clade B have the ability to interact with residue 30 and orient it away from the active site (Figure III-3B and III-3D) and thereby disrupt the interaction between residue 30 and the inhibitor. These structural observations are similar to interpretations made in previous molecular-dynamics studies involving

FIGURE III-4

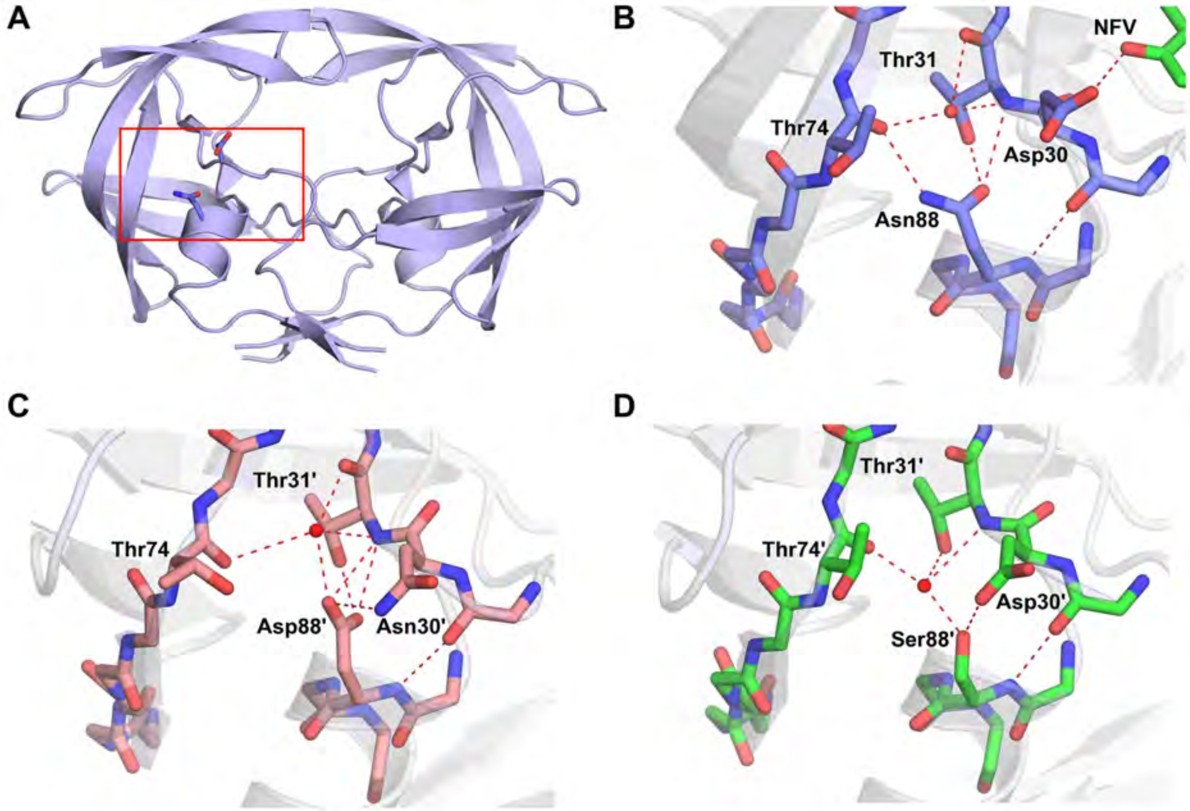

A
B
Thr31
NFV
Thr74
Asp30
Asn88
C
Thr31'
Thr74
Asp88'
Asn30'
D
Thr31'
Thr74'
Asp30'
Ser88'

FIGURE III-4 Hydrogen bond network involving residue 88 **(A)** Asn88 bridges the terminal helix with Asp30 from the active site and Thr74 from one of the outer beta-strands. The red box indicates the region of the protease molecule highlighted in panels B-D. **(B)** Asn88 in $NFV_{B\text{-}WT}$ (PDB code: 3EKX). **(C)** Asp88 in $DRV_{B\text{-}D30N/N88D}$. **(D)** Ser88 in $DRV_{N88S\text{-}WT}$. Hydrogen bond interactions are indicated in red dashed lines.

NFV-protease complexes (Ode et al., 2007; Ode et al., 2005). Thus, NFV resistance is likely caused in large part due to the loss of this interaction in the NFV resistant variants.

Overall, mutations that emerge in response to inhibitor therapy need to have a minimal impact on protease structure and activity to maintain the enzyme's function. The D30N substitution, which is associated with NFV resistance, is one of the few drug resistant mutations that involve a change in charge. The additional substitution of N88D likely helps preserve the net charge on the protein. In AE, resistance to NFV occurs indirectly with the N88S mutation. Likewise, the sole NFV resistant alteration N88S in the AE protease, does not change the overall electrostatics. Thus, in both clade B and AE, NFV resistance is attained with no change to the net charge of the enzyme. In the wild type variants Asn88 is one of the few internal hydrogen bonding side chains in the core of the protease monomer. The side chain of Asn88 has a key role in the protease structure bridging the terminal helix with residues 30 and 31 coming from the active site to the backbone of Thr 74 in the center of one of the outer beta-strands (Figure III-4A-B). With the substitutions of Asp in clade B and Ser in AE for Asn at position 88 in the NFV resistant protease variants, the hydrogen bonding network is preserved through the coordination of some key water molecules in the core of the protease monomer (Figure III-4B-D). Thus mutations confer resistance to NFV through a series of interdependent changes that preserve the structural and electrostatic properties of HIV-1 protease.

In conclusion protease activity and the response to protease inhibitors can be affected by clade-specific sequence differences. Our findings likely extend beyond HIV-1 protease and to other drug targets within HIV and underscore the need to consider clade-specific polymorphisms when developing new drugs and formulating treatment plans. Furthermore, drug-resistance pathways observed in the context of clade B viruses cannot be assumed to hold true for other HIV-1 clades.

ACKNOWLEDGMENTS

This work was supported by the National Institutes of Health grant (P01-GM66524) and Tibotec Inc. to CAS. Additionally this study was supported by a Grant-in-Aid for AIDS research from the Ministry of Health, Labor, and Welfare of Japan (H19-AIDS-007) to WS. I gratefully acknowledge the Mail-In Data Collection Program of the National Synchrotron Light Source, Brookhaven National Laboratory, for collecting X-ray data at the X29A beamline for which financial support comes principally from the Offices of Biological and Environmental Research and of Basic Energy Sciences of the US Department of Energy, and from the National Center for Research Resources of the National Institutes of Health. Use of the Advanced Photon Source for x-ray data collection was supported by the U.S. Department of Energy, Basic Energy Sciences, Office of Science, under Contract No. DE-AC02-06CH11357. Use of the BioCARS Sector 14 was supported by the National Institutes of Health, National Center for Research Resources, under grant number RR007707. The protease inhibitors used in this study were obtained through the NIH AIDS Research and Reference Reagent Program, Division of AIDS, NIAID, NIH. I also wish to thank Drs. William Royer, Moses Prabu-Jayabalan and Madhavi Nalam for helpful discussion and Christina Ng and Brendan Hilbert for assistance with data collection.

CHAPTER IV

Differential Resistance in HIV-1 Protease: The Role of Residue 50 in Pathways for Amprenavir/Darunavir and Atazanavir Resistance.

ABSTRACT

The emergence of inhibitor resistance mutations is a major drawback in the use protease inhibitors in the treatment of HIV-1 infections. Residue 50 in HIV-1 protease is an active site residue that is often substituted in order to confer resistance to amprenavir (APV), darunavir (DRV) and atazanavir (ATV). The I50V substitution is often associated with APV and DRV resistance while the I50L substitution is observed in patients failing ATV therapy. Structural and binding thermodynamics studies were carried out to explain how APV, DRV and ATV susceptibility are influenced by mutations at residue 50 in HIV-1 protease. Binding affinity data indicates that the I50V substitution can reduce affinity for APV and DRV. Interestingly, the I50V substitution made the protease more susceptible to ATV. The I50L variant had reduced affinity for ATV and DRV but was more susceptible to APV binding. Structural data indicate that perturbations resulting from substitutions at residue 50 affect how APV, DRV and ATV interact with protease. Thus, the varied inhibitor susceptibilities observed between the protease variants are likely a direct consequence of the structural differences resulting from the substitutions at residue 50.

INTRODUCTION

With no cure available for the treatment of HIV infections at present, slowing down the progression of the infection to AIDS has been a major focus in anti-HIV therapy development. The aspartyl protease of HIV-1 has been an important drug target and presently there are nine FDA approved protease inhibitors (PIs) that are an important part of highly active anti retroviral therapy (HAART) (Debouck, 1992). However, the selection of HIV-1 variants with inhibitor resistant mutations in the protease gene impairs the ability of PIs to effectively block protease activity (Condra et al., 1995; Emini et al., 1996; Roberts, 1995).

The HIV-1 protease is a homodimeric protein with 99 amino acid residues in each monomer with the dimeric interface forming the active site. At least ten non-homologous and asymmetric substrate sites within the Gag and Gag-Pro-Pol polyproteins are cleaved by the HIV-1 protease to release viral enzymes and structural proteins essential for the formation of mature virions (Chou, 1996b; Henderson et al., 1988; Pettit et al., 1998; Sadler and Stein, 2002). PIs have been designed to effectively inhibit the binding and subsequent cleavage of the HIV-1 substrates and thereby prevent mature virion formation. Amino acid substitutions that accumulate within the protease can reduce the ability of a PI to effectively bind within the active site. This leads to PI resistance where the protease retains the ability to bind and cleave viral substrates regardless of the presence of the inhibitor.

Amprenavir (APV), darunavir (DRV) and atazanavir (ATV) are three potent FDA approved PIs that bind to HIV-1 protease with high affinity (Figure IV-1A) (King et al.,

FIGURE IV-1

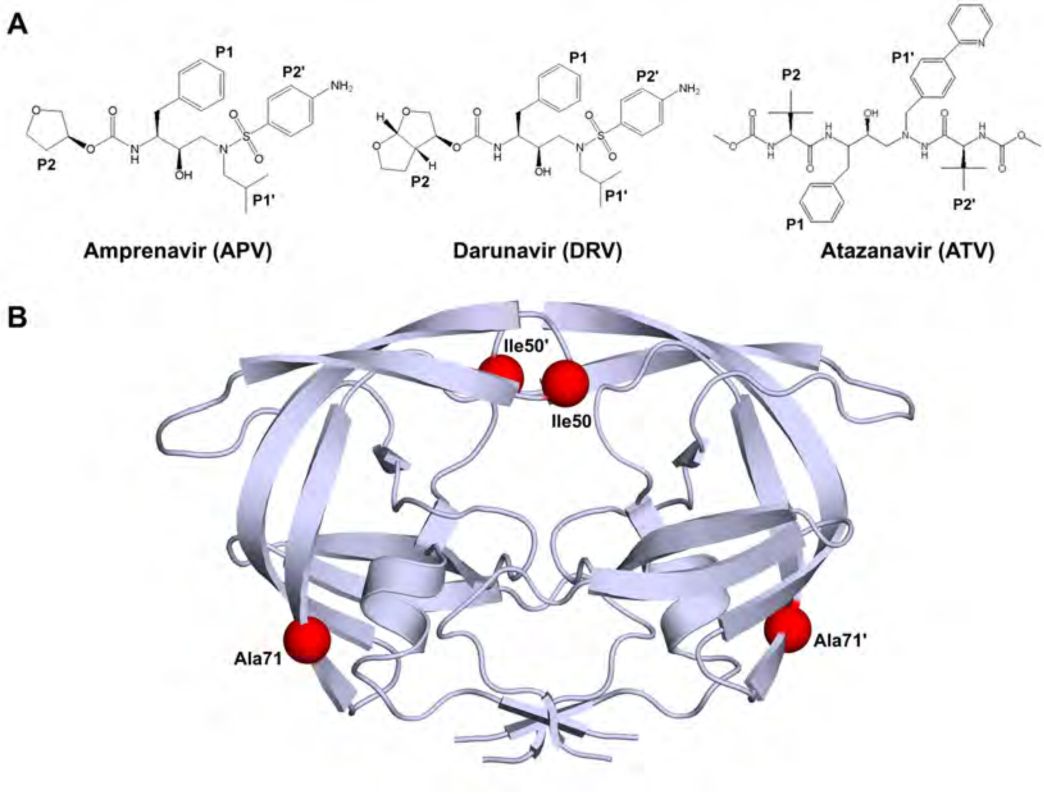

A
P1
P2'
NH2
P2
P1'
OH
Amprenavir (APV)
Darunavir (DRV)
Atazanavir (ATV)
B
Ile50'
Ile50
Ala71
Ala71'

FIGURE IV-1 (A) Chemical structures of amprenavir, darunavir and atazanavir. Chemical groups that correspond to substrate residues when bound within the active site are indicated as P1, P2, P1’ and P2’. **(B)** Homodimeric HIV-1 protease with positions 50 and 71 indicated by red spheres in each monomer.

2004d; Robinson et al., 2000; Sadler and Stein, 2002). However, substitutions at a number of amino acid positions within the protease have been observed to reduce the efficacy of APV, DRV or ATV. From these observations, substitutions accumulating at the active site residue position 50 (Figure IV-1B) are commonly associated with resistance for APV, DRV and ATV.

Ile to Val substitution at residue 50 has been identified as the signature APV and DRV resistance mutation in patients failing APV and DRV therapy (Partaledis et al., 1995b; Tisdale et al., 1999; Van Marck et al., 2007; Vermeiren et al., 2007). In patients undergoing ATV therapy, Ile to Leu at residue 50 was observed to accumulate in patients failing therapy (Colonno et al., 2004a; Colonno et al., 2004b). However, patients with the I50L substitution in protease have been observed to respond significantly better to other PIs and suggest that I50L makes the protease hypersusceptible to other PIs (Colonno et al., 2004b). The substitutions at residue 50 are often observed together with a secondary A71V substitution (Figure IV-1B) that is outside the active site (Colonno et al., 2004b). The A71V substitution has been shown compensate the loss of activity resulting from primary drug resistance substitutions (Nijhuis et al., 1999). Overall, changes at residue 50 appear to have the ability modulate differential resistance for APV, DRV and ATV.

In the present study, structural and biophysical methods were used to determine the impact of substitutions at residue 50 on APV, DRV and ATV susceptibility. Binding thermodynamics data show that the I50V and I50L substitutions indeed affect protease susceptibility to APV, DRV and ATV. Through the determination of crystal structures and analysis of interactions between inhibitors and protease, structural rationalizations are described for the molecular basis for the varied susceptibility observed in the residue 50 variants.

MATERIALS AND METHODS

Protease-gene construction

The wild-type protease (WT) gene was generated as previously described (Prabu-Jeyabalan et al., 2004a) with the Q7K substitution introduced to prevent autoproteolysis (Rose, Salto, and Craik, 1993b). The APV and DRV resistant variant (I50V) and the ATV resistant variant (I50L) protease genes were generated by introducing either I50V and A71V or I50L and A71V to the wild-type gene by site-directed mutagenesis using the Stratagene QuikChange site-directed mutagenesis kit (Agilent Technologies, La Jolla, CA). Mutagenesis was confirmed by sequencing.

Protein expression and purification

Each variant was subcloned into the heat-inducible pXC35 expression vector (American Type Culture Collection (ATCC), Manassas, VA) and transformed into *E. coli* TAP-106 cells. Protein overexpression, purification and refolding was carried out, as previously described (King et al., 2002). Protein used for crystallographic studies was further purified with a Pharmacia Superdex 75 fast-performance liquid-chromatography column (GE Healthcare, Chalfont St. Giles, U.K.) equilibrated with refolding buffer (50mM sodium acetate pH 5.5, 10% glycerol, 5% ethylene glycol, and 5mM dithiothreitol).

Isothermal titration calorimetry (ITC)

Binding affinities and thermodynamic parameters of APV, DRV and ATV binding to the WT, I50V and I50L protease variants were determined by ITC on a VP isothermal titration calorimeter (MicroCal, LLC, Northampton, MA). The buffer used for all protease and inhibitor solutions consisted of 10 mM sodium acetate (pH 5.0), 2% dimethyl sulfoxide, and 2 mM *tris*[2-carboxyethyl]phosphine. Binding affinities for all protease variants were obtained by competitive-displacement titration using acetyl-pepstatin as the weaker binder. A solution of 6–40 μM protease was titrated with 200-400 μM acetyl-pepstatin to saturation. The pepstatin was then displaced by titrating 200-330 μM of APV, DRV or ATV. Heats of dilution were subtracted from the corresponding heats of reaction to obtain the heat resulting solely from the binding of the ligand to the enzyme. Data were processed and analyzed with the ITC data analysis module (Microcal) for Origin 7 data analysis and graphing software (OriginLab, Northampton, MA). Final results represent the average of at least two measurements.

Crystallization and structure determination

Protease solutions of I50V and I50L at concentrations between 1.0 and 2.0 mg mL^{-1} were equilibrated with three- to five-fold molar excess of APV, DRV and ATV for 1 hour on ice. Crystals were grown over a reservoir solution consisting of 126-mM phosphate buffer (pH 6.2), 63-mM sodium citrate and 18% to 40% ammonium sulfate by the hanging-drop vapor-diffusion method. X-ray diffraction data for I50V in complex with DRV were collected on BioCARS beamline 14-BMC at the Advanced Photon Source (Argonne National Laboratory, Argonne, IL) at a wavelength tuned to 0.9 Å with a Quantum 315 CCD X-ray detector (Area Detector Systems Corporation, Poway, CA). Diffraction data for I50V in complex with APV and ATV were collected on BioCARS beamline 14-IDB at the Advanced Photon Source (Argonne National Laboratory) at a wavelength tuned to 1.03 Å with a MarCCD 165 X-ray detector (Rayonix, LLC, Evanston, IL). Data for all other complexes were collected in-house on an R-Axis IV imaging-plate system (Rigaku Corporation, Tokyo, Japan) mounted on a rotating-anode X-ray source (Rigaku Corporation) at a wavelength of 1.54 Å. All data were collected under cryo-cooled conditions.

The data were indexed, integrated and scaled using HKL/HKL-2000 software (HKL Research, Charlottesville, VA) (Otwinowski and Minor, 1997). Structure determination was done by molecular replacement with the program AMoRe (Navaza, 1994) and refinement were carried out using the CCP4 program suite (Collaborative Computational Project Number 4, 1994), as previously described (Prabu-Jeyabalan et al., 2006b). The tensor (T), libration (L), and screw (S) parameter files used in TLS refinement of the I50V complexes were generated using the TLS Motion Determination server (Painter and Merritt, 2006). Model building was carried out, followed by real-space refinement, with either the O molecular graphics software (Jones, Bergdoll, and Kjeldgaard, 1990) or the COOT molecular-graphics software (Emsley and Cowtan, 2004). Structure comparisons were made by superposing structures using the Cα atoms of the terminal regions (residues 1 to 9 and 86 to 99) from both monomers. Structures were visualized using PyMol molecular-graphics software (DeLano, 2002).

TABLE IV-1. Crystallographic statistics for WT, I50V and I50L variants in complex with APV, DRV and ATV.

	WT			I50V			I50L		
Inhibitor	APV	DRV[a]	ATV	APV	DRV	ATV	APV	DRV	ATV
Resolution (Å)	1.8	1.2	1.8	1.6	1.9	1.6	2.2	2.1	2.1
Space group	$P\,2_12_12_1$	$P\,2_12_12_1$	$P\,2_12_12_1$	$P2_1$	$P2_1$	$P2_1$	$P\,2_12_12_1$	$P\,2_12_12_1$	$P2_1$
Cell dimensions: a (Å)	50.7	54.9	50.9	50.6	50.62	51.2	50.9	51.4	51.2
b (Å)	57.9	57.8	58.1	63.3	63.34	58.4	58.1	57.9	59.5
c (Å)	61.7	62.0	62.0	58.6	58.64	62.1	61.2	61.4	59.9
β (°)	-	-	-	96.6	97.3	95.8	-	-	82.2
Z[c]	4	4	4	4	4	4	4	4	4
Total no. of reflections	49,469	302,022	114,163	148,290	118,410	146,999	32,037	33,870	56,701
Unique no. of reflections	14,987	50,056	16,905	36283	26,674	42,520	9,311	10,762	20,006
R_{merge} (%)	2.9	3.8	7.1	7.0	7.0	4.9	8.6	6.2	4.7
Completeness (%)	79.7	95.5	95.7	98.0	99.6	96.2	95.7	96.6	95.8
I/σ^I	19.0	25.0	11.3	8.7	8.4	11.6	6.4	7.9	9.8
R_{work} (%)	19.6	14.1	17.7	16.9	16.3	17.3	19.6	18.0	19.0
R_{free} (%)	22.6	17.9	20.9	20.5	21.2	21.5	25.8	24.0	24.8
RMSD[b]:Bond length (Å)	0.007	0.004	0.008	0.008	0.010	0.009	0.008	0.004	0.008
Bond angle (°)	1.4	1.5	1.2	1.4	1.4	1.4	1.4	1.2	1.4
PDB code	3EKV	1T3R	3EKY				3EM3	3EM6	3EM4

[a] Adapted from (King *et al*., 2004)

[b] Root mean square deviation.

[c] Number of dimeric molecules in a unit cell.

Crystallographic and refinement statistics are given in Table IV-1. The structures of WT protease in complex with APV, DRV and ATV were previously solved in the lab.

Double-difference plots were generated for each wild-type-inhibitor and mutant-inhibitor protease structures to graphically visualize structural differences between the wild type and mutant inhibitor complexes, as previously described (Prabu-Jeyabalan et al., 2006b). Briefly, distances between all Cα atoms within the dimer were calculated for each complex. A distance-difference matrix was then computed for each atom for a given pair of complexes. The distance-difference matrix was then plotted as a contour plot using gnuplot plotting software (Williams and Kelly, 1998).

The inhibitor-protease van der Waals contacts were estimated for each structure using a simplified Lennard-Jones potential as previously described (King et al., submitted).

Nomenclature

The following nomenclature will be followed to refer to each crystal structure: $\text{Inhibitor}_{\text{protease variant}}$. For example, APV_{WT}, APV_{I50V} and APV_{I50V} refer to the WT, I50V and I50L variants in complex with APV. Prime notation is used to distinguish the two monomers in the protease dimer. For example, residue 30 from the first monomer would be referred to as Asp30 and the same residue from the second monomer would be referred to as Asp30’.

TABLE IV-2. Binding thermodynamic parameters for APV, DRV and ATV binding to WT, I50V and I50L protease.

Protease Variant	K_a (M$^{-1)}$	K_d (nM)	K_d ratio	ΔH (kcal mol^{-1})	ΔΔH	−TΔS (kcal mol^{-1})	Δ(−TΔS)	ΔG (kcal mol^{-1})	ΔΔG
APV									
WT	$(2.6 \pm 1.3) \times 10^9$	0.39 ± 0.20	1.0	−7.3 ± 0.9	-	−5.3	-	−12.6	-
I50V	$(1.1 \pm 0.03) \times 10^9$	0.88 ± 0.02	2.3 ± 1.8	−11.8 ± 0.3	−4.5	−0.3	5.0	−12.1	0.5
I50L	$(7.3 \pm 1.2) \times 10^9$	0.14 ± 0.02	0.4 ± 0.2	−9.5 ± 0.2	−2.2	−3.8	1.5	−13.2	−0.6
DRV									
WT	$(2.2 \pm 1.1) \times 10^{11}$	0.004 ± 0.002	1.0	−12.1 ± 0.8	-	−3.1	-	−15.2	-
I50V	$(1.5 \pm 0.5) \times 10^{10}$	0.067 ± 0.021	16.8 ± 9.9	−12.0 ± 0.6	0.1	−1.6	1.5	−13.6	1.6
I50L	$(6.1 \pm 0.5) \times 10^{10}$	0.016 ± 0.001	4.0 ± 2.0	−13.4 ± 0.9	−1.3	−1.1	2.0	−14.5	0.7
ATV									
WT	$(4.5 \pm 2.4) \times 10^9$	0.23 ± 0.12	1.0	−1.1 ± 0.2	-	−11.8	-	−12.9	-
I50V	$(0.4 \pm 0.2) \times 10^9$	0.03 ± 0.01	0.13 ± 0.08	−7.3 ± 1.1	−6.2	−6.6	5.2	−13.9	−1.0
I50L	$(1.5 \pm 0.1) \times 10^9$	0.65 ± 0.05	2.8 ± 1.5	−3.6 ± 0.4	−2.5	−8.7	3.1	−12.3	0.6

RESULTS

Binding thermodynamics

To determine the effects of resistance mutations at position 50 in HIV-1 protease on inhibitor binding, the binding thermodynamic parameters of APV, DRV and ATV binding to WT, I50V and I50L variants were determined by ITC (Table IV-2). The I50V variant showed 2.2-fold and 16.7-fold weaker affinity for APV and DRV respectively when compared to the binding affinities of the WT protease. In contrast, the I50V variant had a 7.7-fold improved binding affinity for ATV (K_d = 0.03 nM) when compared with the WT protease (0.23 nM). Thus, the binding data indicated that while the I50V variant had reducing affinity for APV and DRV, the protease was more susceptible to ATV.

The I50L variant had a 2.8-fold and 2.0-fold weaker affinity for ATV and DRV respectively when compared with the WT protease. Despite the weaker affinity for DRV, the I50L variant nonetheless exhibits a K_d for DRV (K_d = 0.016 nM) that is 40.6-fold better than for ATV (K_d = 0.65 nM). In contrast, APV bound with higher affinity to the I50L variant when compared with WT protease and indicated that the I50L variant was more susceptible to APV.

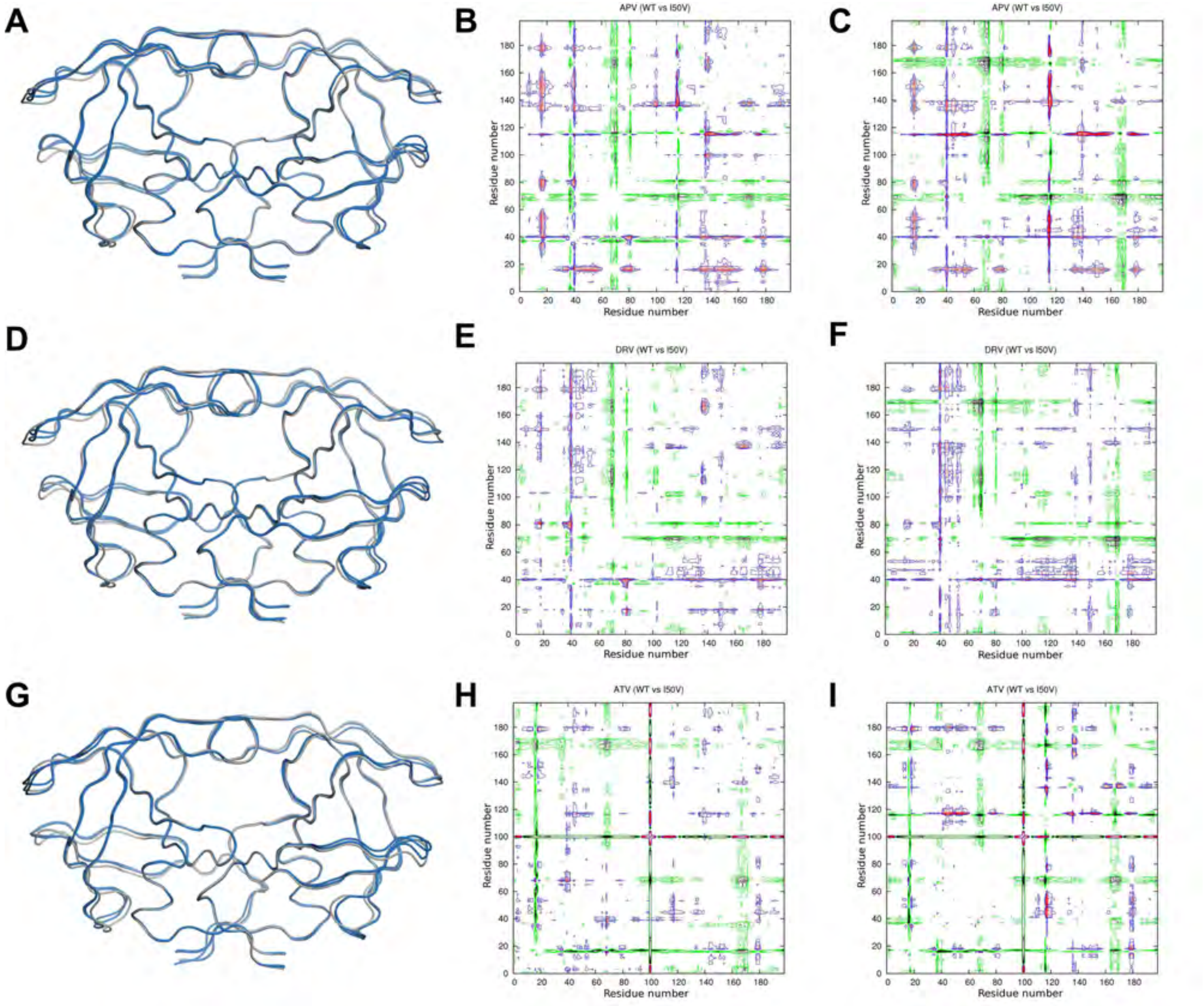
No significant changes were observed in the free energies of binding for APV, DRV or ATV with I50V or I50L when compared to the WT protease. The binding of DRV to all protease variants was enthalpically favorable and the drug resistance mutations had negligible effects on the binding thermodynamic parameters. The binding of APV and ATV was enthalpically favorable with all protease variants. However, with the drug resistant variants, the enthalpic contribution was reduced and was compensated by an increase of the entropic contribution when compared to the binding to WT protease.

Crystal Structures

To evaluate effect of substitutions at residue 50 in protease I50V and I50L protease variants were co-crystallized with APV, DRV and ATV and compared to the WT inhibitor of each complex. The APV_{WT}, DRV_{WT} and ATV_{WT} complexes were previously solved in the lab. The DRV_{I50V}, ATV_{I50V} and ATV_{I50L} protease complexes crystallized with two protease dimers in the asymmetric unit. The APV_{I50V} protease had a second inhibitor molecule bound outside the active site of one of the dimers and is likely a crystallographic artifact since ITC data for APV binding to the I50V variant does not support the presence of a second binding site. Structural comparisons were carried out for each inhibitor-protease resistant variant and corresponding WT protease complex by pairwise structural superpositions and double difference plots as described in the Materials and Methods section (Figure IV-2).

The superposed structures and double difference plots for APV_{I50V}, DRV_{I50V} and ATV_{I50V} revealed several regions of the protease backbone that differed between the mutant and WT structures (Figure IV-2 A-I). These differences predominantly were concentrated around Gly40-Arg57 (flap region), which included the I50V substitution, and Gly68-Gly71, which included the A71V substitution. The differences seen in the flap region were more prominent in one monomer of the DRV_{I50V} structure. Similar structural changes were observed in the APV_{I50L}, DRV_{I50L} and ATV_{I50L} structures (Figure IV-2 J-

FIGURE IV-2


A
B
APV (WT vs I50V)
C
APV (WT vs I50V)
D
E
DRV (WT vs I50V)
F
DRV (WT vs I50V)
G
H
ATV (WT vs I50V)
I
ATV (WT vs I50V)
Residue number
Residue number
0
20
40
60
80
100
120
140
160
180

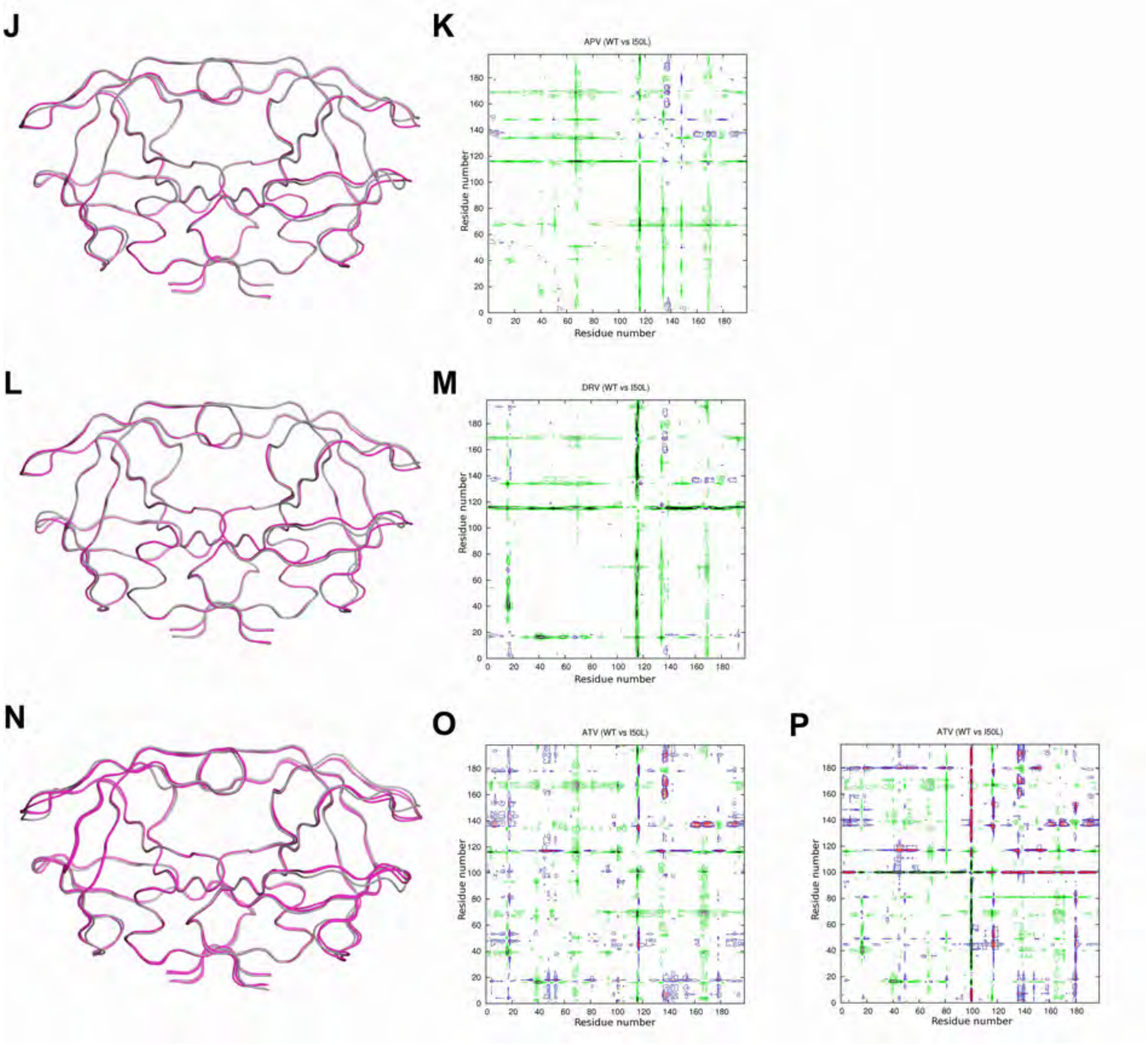

J
K
APV (WT vs I50L)
Residue number
Residue number
L
M
DRV (WT vs I50L)
Residue number
Residue number
N
O
ATV (WT vs I50L)
Residue number
Residue number
P
ATV (WT vs I50L)
Residue number
Residue number
0
20
40
60
80
100
120
140
160
180

FIGURE IV-2. Ribbon representations of structural superpositions and double difference plots. WT structures are colored in grey, I50V in blue and I50L in magenta in the ribbon diagrams. The color contours in the double difference plots indicate distance differences of < -1.0 Å (black), -1.0 to -0.5 Å (green), 0.5 to 1.0 Å (blue) and > 1.0 Å (red). **(A)** APV_{WT} and APV_{I50V}. **(B)** Double difference plot for APV_{WT} and APV_{I50V} dimer 1. **(C)** Double difference plot for APV_{WT} and APV_{I50V} dimer 2. **(D)** DRV_{WT} and DRV_{I50V}. **(E)** Double difference plot for DRV_{WT} and DRV_{I50V} dimer 1. **(F)** Double difference plot for DRV_{WT} and DRV_{I50V} dimer 2. **(G)** ATV_{WT} and ATB_{I50V} **(H)** Double difference plot for ATV_{WT} and ATB_{I50V} dimer 1. **(I)** Double difference plot for ATV_{WT} and ATV_{I50V} dimer 2. **(J)** APV_{WT} and APV_{I50L}. **(K)** APV_{WT} and APV_{I50L}. **(L)** Double difference plot for DRV_{WT} and DRV_{I50L}. **(M)** DRV_{WT} and DRV_{I50L}. **(N)** Double difference plot for ATV_{WT} and ATV_{I50L}. **(O)** Double difference plot for ATV_{WT} and ATV_{I50L} dimer 1. **(P)** Double difference plot for ATV_{WT} and ATV_{I50L} dimer 2.

P). However the, APV_{I50L} and DRV_{I50L} structures had contours in the double difference plots that map to backbone shifts greater than 1.0 Å when compared to the WT complexes (Figure IV-2 J and L). Furthermore, difference peaks observed for the APV_{I50L} and DRV_{I50L} complexes were significantly less than for that of ATV_{I50L} (Figure IV-2 K, M, O and P) and indicated that backbone shifts resulting from the I50L substitution were smaller than those observed for ATV_{I50L}.

To investigate if residues 50 and 71 play a role in modifying inhibitor binding, hydrogen bond interactions between the inhibitors and protease variants were analyzed. The APV_{I50V} structure showed no significant difference in the hydrogen bonding pattern when compared to the APV_{WT} structure. The APV_{I50L} structure exhibited a loss in one hydrogen bond between Asp29 N of the protease and O6 of APV while two new hydrogen bond interactions were observed between N3 of APV and Asp30 OD2, and with O1 of APV and a water molecule. Hydrogen bond interactions between DRV_{wt}, DRV_{I50V} and DRV_{I50L} were almost identical with the notable difference being that the N1 atom of DRV in the DRV_{I50V} and DRV_{I50L} complexes directly interacted with the Asp30' side chain whereas in the DRV_{WT} structure, the N1-Asp30' interaction was bridged via a water molecule. Analyses of the ATV complexes revealed no significant changes in the hydrogen bonding interactions between the inhibitor and protease. Thus, the varied inhibitor susceptibilities observed with the binding data are not a direct result of changes in protease-inhibitor hydrogen bonding.

Van der Waals interactions between the protease and inhibitors were analyzed for each complex (Figure IV-3A-I). Ile50 interacts with the sulfonyl moiety in the APV and

FIGURE IV-3

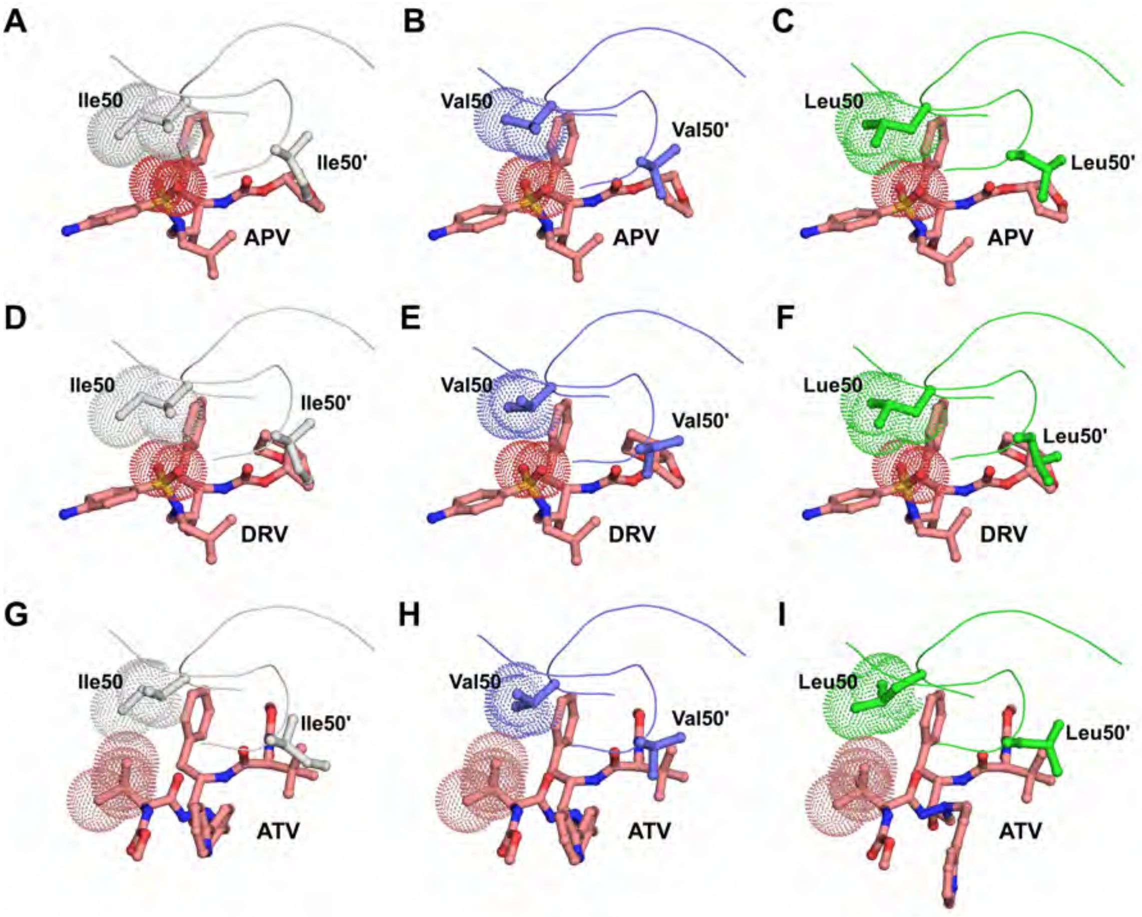

A
Ile50
Ile50'
APV
B
Val50
Val50'
APV
C
Leu50
Leu50'
APV
D
Ile50
Ile50'
DRV
E
Val50
Val50'
DRV
F
Lue50
Leu50'
DRV
G
Ile50
Ile50'
ATV
H
Val50
Val50'
ATV
I
Leu50
Leu50'
ATV

FIGURE IV-3. Van der Waals interactions of residue 50 with APV, DRV and ATV with the protease variants. Protease variants are colored: WT (gray), I50V (blue) and I50L (green). The inhibitors are colored in salmon. **(A)** $\mathrm{APV}_{\mathrm{WT}}$. **(B)** $\mathrm{APV}_{\mathrm{I50V}}$. **(C)** $\mathrm{APV}_{\mathrm{I50L}}$. **(D)** $\mathrm{DRV}_{\mathrm{WT}}$. **(E)** $\mathrm{DRV}_{\mathrm{I50V}}$. **(F)** $\mathrm{DRV}_{\mathrm{I50L}}$. **(G)** $\mathrm{ATV}_{\mathrm{WT}}$. **(H)** $\mathrm{ATV}_{\mathrm{I50V}}$. **(I)** $\mathrm{ATV}_{\mathrm{I50L}}$.

DRV complexes and packs against the *t*-butyl group at the P2' position in the ATV complex (Figure IV-3). A significant loss of van der Waals interactions was observed at Val50 in the APV_{I50V}, DRV_{I50V} and ATV_{I50V} complexes when compared to the WT complexes (Figure IV-3A, B, D, E, G and H). Similarly, Leu50 in the ATV_{I50L} variant had reduced van der Waals interactions with the inhibitor when compared to the WT complex (Figure IV-3I). However, in comparison, the change in van der Waals interactions between the I50L variant and APV and DRV was negligible (Figure IV-3A, C, D and F). Overall, amino acid changes at residue 50 resulted in changes in van der Waals contacts with the inhibitor predominantly around the P2 position of the inhibitors.

DISCUSSION

APV, DRV and ATV are potent inhibitors that are used as a part of HAART to slow down the progression of HIV-1 to AIDS (King et al., 2004d; Robinson et al., 2000; Sadler and Stein, 2002). However, amino acid substitutions accumulating at residue 50 in protease have been observed to affect susceptibility to these inhibitors. The I50V/A71V substitutions have been observed predominantly in patients failing APV and DRV therapy while the substitutions I50L/A71V occur in patients failing ATV therapy (Colonno et al., 2004a; Colonno et al., 2004b; Partaledis et al., 1995b; Tisdale et al., 1999; Van Marck et al., 2007; Vermeiren et al., 2007). To examine how substitutions at residue 50 affect inhibitor susceptibility we have performed detailed analyses on binding thermodynamics and structural data of APV, DRV and ATV interactions with I50V and I50L protease variants. The results presented in this study provide insights into the molecular basis for differential susceptibility profiles observed in HIV-1 protease resulting from the I50V and I50L resistance mutations.

The I50V/A71V substitutions reduced affinity of both APV and DRV, which have closely related chemical structures (Table IV-1 and Figure IV-1A). Thus, the weaker affinity observed for the I50V variant provides a likely explanation for therapy failure in patients receiving APV and DRV therapy. Conversely, the I50V variant was more susceptible to ATV. The idea of hypersusceptibility of the I50V variant to ATV has not been widely established. However, there have been clinical reports of increased or no change in susceptibility to ATV in patients who carry the I50V mutation in the protease gene (Colonno et al., 2004a; Elston et al., 2006; Wainberg, Martinez-Cajas, and Brenner, 2007). In one instance the hypersusceptibility was observed in the presence of Gag cleavage site mutations (Wainberg et al., 2007). Our binding thermodynamics data, in combination with the clinical observations, suggest that the I50V substitution has the potential to make the protease more susceptible to ATV. However, secondary mutations or cleavage site mutations might be required in order to enhance the binding of ATV to I50V protease *in-vivo*.

Binding thermodynamics data indicate the the I50L variant has reduced affinity for ATV (Table IV-2). This result is consistent with previously published reports and supports clincal observations where patients who accumulate the I50L mutation in the protease gene respond poorly to ATV therapy (Colonno et al., 2004b). The increased affinity of the I50L variant to APV is also consistent with previously published data and corroborates the idea that the I50L substitution makes the protease hypersusceptible to

APV (Colonno et al., 2004b; Yanchunas et al., 2005). While a similar result was expected for DRV given its high degree of similarity to APV, a two-fold loss of affinity was observed with DRV binding to the I50L variant. Despite the loss of affinity DRV maintained a high affinity (K_d = 16.0 pM) to the I50L variant which is 14-fold higher than ATV binding WT protease (K_d = 230.0 pM) . Thus, despite the loss of affinity for the I50L variant, DRV could still effectively inhibit the protease.

Based on our structural analysis the I50V subsitution loses a signficant amount of van der Waals interactions with the sulfonyl group of APV and DRV (Figure IV-3). With the I50V substitution, residue 50 loses the ability to form better interactions with thus sulfonyl moeity due to the loss of the branched methyl group in valine. This loss of interactions could contribute to the reduced binding affinities observed for APV and DRV with the I50V variant. However, reasons for the increased suceptibility of the I50V variant to ATV are not evident from the crystal structures. The shorter side chain of Val50 is unlikely to have steric clashes with the bulky *t*-butyl group of ATV at P2' position. Thus, Val50 can pack well against the *t*-butyl group of ATV and allow the inhibitor to bind with minimal structural perturbations which is likely reflected in the gain of free energy of binding ($\Delta\Delta G = -1.0$ kcal mol^{-1}).

The longer side chain of Leu50 likely results in steric clashes with the bulky *t*-butyl groups of ATV at the P2' positions in the ATV_{I50L} complex. Therefore, in order to accommodate the inhibitor residue 50 is slightly shifted outward relative to Ile50 in the WT complex. This shift in the flap backbone is relected in the structural superpositions and double difference plots (Figure IV-2 N-P). Furthermore, the backbone shift results in the loss of favorable van der Waals contacts with the inhibitor particularly at Ile50 and P2’ position. These observations are similar to results obtained by modeling studies on ATV with I50L protease (Yanchunas et al., 2005). Conversely, with the I50L substitution, the additional methyl group in Leu allows residue 50 to maintain its interactions with the sulfonyl group at P2 and likely contributes to the increased affinity observed for APV.

In conclusion, our findings show that amino acid substitutions at residue 50 in HIV-1 protease can alter APV, DRV and ATV susceptibility and are consistent with observations seen in clinical studies. Results presented in this study suggest that the varied susceptibility to APV, DRV and ATV are likely a consequence from structural perturbations resulting from substitutions at residue 50 that alter interactions between inhibitors and protease. Furthermore, the accumulation of I50V or I50L as inhibitor resistance substitutions is likely influenced by the chemical moeities that interact with residue 50. While the results presented in this study provide insights into the molecular basis for APV, DRV and ATV resistance through changes at residue 50, further studies are needed to evaluate secondary resistance mutations that occur together with I50V and I50L as well as cleavage site mutations order to better understand observations seen in clinical settings.

ACKNOWLEDGEMENTS

This work was supported by the National Institutes of Health grant (P01-GM66524) and Tibotec Inc. Use of the Advanced Photon Source for x-ray data collection was supported by the U.S. Department of Energy, Basic Energy Sciences, Office of Science, under Contract No. DE-AC02-06CH11357. Use of the BioCARS Sector 14 was supported by the National Institutes of Health, National Center for Research Resources, under grant number RR007707. The protease inhibitors used in this study were obtained through the NIH AIDS Research and Reference Reagent Program, Division of AIDS, NIAID, NIH. I would also like to thank Drs. William Royer and Madhavi Nalam for helpful discussion when solving the crystal structures.

CHAPTER V

Discussion

The HIV-1 and AIDS pandemic continues to be a serious global health issue despite efforts for almost thirty years to suppress the virus. With no cure available to treat HIV-1 infections and AIDS, inhibiting the spread of the virus has been a key focus in efforts to slow down the progression of the infection to AIDS. HIV-1 protease is an important component of the viral replication cycle and thus, has been a key anti-HIV drug target. Currently there are nine FDA approved protease inhibitors that are used successfully as a part of HAART. However, as is with all anti-HIV drugs, the emergence of drug resistance substitutions within the protease has been a major obstacle with the use of PIs. These substitutions can significantly reduce the efficacy of the PI being used and resistance substitutions emerge against all currently available PIs. The emergence of resistance substitutions gives rise to a constant need to design new inhibitors to help patients combat drug resistant viral strains. Thus, it is important to understand how resistance substitutions modulate inhibitor binding in order to design robust inhibitors that are not influenced by resistance substitutions.

The primary focus of my dissertation research was to try and understand the molecular basis for drug resistance due to amino acid substitutions in protease. A combination of biochemical and biophysical studies were performed to gain mechanistic insights into how amino acid substitutions within protease affect enzyme function as well as to understand how they contribute to PI resistance. Furthermore, as a part of these studies, I have analyzed the role of clade specific protease sequence variations in modulating enzyme activity and PI susceptibility.

Drug resistance in non-B clade HIV-1 proteases

Until recently, the genetic diversity of the HIV-1 genome was not considered to be important in formulating anti-HIV treatment strategies. However, as the incidence of HIV-1 continues, the genetic variability of HIV-1 has been identified to be an important factor that influences viral transmission and disease progression (Hu et al., 1999; Kaleebu et al., 2002a; Kanki et al., 1999a; Spira et al., 2003b). Clinical observations show that clade specific sequence variations within HIV-1 protease can affect inhibitor susceptibility as well as the development of inhibitor resistance (Kantor and Katzenstein, 2003; Kantor and Katzenstein, 2004). Furthermore, different HIV-1 clades have been observed to develop altered pathways to inhibitor resistance (Ariyoshi et al., 2003a; Grossman et al., 2004b). Clade B is the most widely studied HIV-1 clade and at present, a vast body of biochemical and structural knowledge exists on clade B protease. At the time this dissertation research was undertaken, limited biochemical data and no structural data were available on non-B clade proteases.

Work carried out by Valzaquez-Campoy and colleagues on a number of HIV-1 clades from the sub-Saharan Africa region initially showed that clade specific sequence variation had varied enzymatic activity and inhibitor susceptibility when compared to clade B protease (Velazquez-Campoy et al., 2003; Velazquez-Campoy, Vega, and Freire, 2002). In work presented in chapters II and III, I have focused on the HIV-1 CRF01_AE (AE) protease, which is one of the prototypic recombinant HIV-1 variants, and have examined how sequence variations specific to AE protease affected protease structure and function as well as their role in influencing the pathway to nelfinavir (NFV) resistance.

AE protease has seven amino acid positions that can differ from the clade B protease sequence. A number of these amino acid differences seen in the wild type AE protease sequence, are associated with inhibitor resistance in clade B protease. Therefore, one working hypothesis in this study was that AE protease has a baseline level of resistance when compared to clade B protease. Moreover, this baseline level of resistance should enable AE protease to confer resistance much faster than clade B protease. From results presented in Chapter III, I was able to show that AE protease had an inherent weaker affinity for the PIs NFV and darunavir (DRV) and thus confirmed the hypothesis. The weaker affinity observed for NFV is also consistent with previously published data for another AE protease variant as well as for clade A protease, which closely related to AE (Velazquez-Campoy et al., 2003). Furthermore, preliminary binding studies on wild type clade C protease (Appendix I) strengthen the idea that non-B clade proteases can have varied resistance profiles when compared to clade B protease.

Structural data presented in chapters II and III show that sequence polymorphisms in AE protease result in structural changes within the protease. These changes are primarily centered around the flap hinge region and core region of the protease. The crystal structures of AE protease show that the flap hinge is packed against the core region and this packing is stabilized by hydrogen bonds. The relatively short side chain of Ile36 likely facilitates the packing of the flap hinge against the core region of the protease. Conversely, Met36 in clade B protease has a longer side chain and prevents the flap hinge from collapsing against the core region. Thus, the AE protease flap hinge is more restrained when compared to the clade B protease flap hinge. Recently published crystal structures of clades A and F protease show flap hinge restrictions similar to AE protease (Robbins et al., 2010; Sanches et al., 2007a). The protease from clades A and F also have Ile36, which appears to facilitate the packing of the flap hinge against the core region of the protease.

The movement of the protease flaps is essential for allowing substrates to bind and for cleaved products to be released (Foulkes-Murzycki, Scott, and Schiffer, 2007). Thus, the flap hinge region of HIV-1 protease plays an important role in modulating flap dynamics (Rose, Craik, and Stroud, 1998b; Scott and Schiffer, 2000; Todd and Freire, 1999b). The structural data presented in chapters II and III suggest that the AE protease flap hinge is more restricted when compared to the clade B protease. With enzyme kinetics data described in Chapter III, I was able to show that the catalytic efficiency of AE protease was significantly less than clade B protease. The decreased catalytic efficiency could be a result of the restricted flap hinge of the AE protease.

The protease needs to undergo significant conformational changes in order to facilitate substrate binding. Molecular dynamics studies have suggested that hydrophobic sliding of residues within the core region of the protease may be important to facilitate the conformational changes the protease needs to undergo to allow substrate binding and product release (Foulkes-Murzycki, Scott, and Schiffer, 2007). The hydrogen bond interactions between the flap hinge and core region of the AE protease have the ability to disrupt the hydrophobic sliding or the core. Thus, the constrained movement of the core region may well be reflected in the decreased catalytic efficiency values observed for AE protease.

NFV resistance in clade B protease occurs through a combination of an active site mutation D30N and a non-active site mutation N88D. However, NFV resistance in AE occurs mainly through the non-active site N88S substitution. In data presented in Chapter III, I provide structural explanations for the molecular basis for NFV resistance resulting from the two different pathways. Clinical observations report that AE infected patients who accumulate the N88S substitution in response to NFV, quickly develop an additional L10F substitution in protease (Ariyoshi et al., 2003a). *In vitro* viral replication capacity studies have indicated that accumulation of the N88S substitution in protease significantly reduces the replicative capacity of the virus (Dr. Wataru Sugiura, personal communication). L10F is a non-active site substitution and therefore, it is not clear how this substitution enhances replicative capacity of the N88S variant. Kinetics studies have shown that the L10F substitution by itself did not have any significant effect on protease activity (Pazhanisamy et al., 1996). Thus, future enzyme kinetics and inhibitor binding studies need to evaluate the role of L10F in AE protease combination with the N88S substitution. Additional crystallographic studies on AE protease with N88S and L10F substitutions may provide structural insights into the role of L10F in enhancing the effects of the N88S substitution.

At present, very little is known about the variability of substrate sequences between HIV-1 clades. A study analyzing a set of 89 HIV-1 sequences from a number of clades revealed that while the majority of substrate cleavage sites were conserved between clades, the p2/NC, TFP/p6^{pol}, and p6^{pol}/PR sites were variable between clade B and C HIV-1 variants (de Oliveira et al., 2003). Recent years have seen an increase in the number of full-length non-B clade HIV-1 sequences available through sequence repositories. Therefore, as a future direction, I propose to examine how substrate cleavage sites differ between clades. Substrate cleavage sites often co-evolve in response to drug resistance substitutions in protease (Kolli, Lastere, and Schiffer, 2006; Kolli et al., 2009). Thus, analysis of non-B substrate sites might reveal substrates that have co-evolved in response to clade specific sequence variations in protease. The sequence in which the Gag and Gag-Pro-Pol polyproteins are processed is essential for the proper maturation of HIV virions. The decreased catalytic efficiency of AE protease may well alter how Gag and Gag-Pro-Pol polyproteins are processd. If substrates co-evolve in response to clade specific variations in protease, they likely restore proper Gag and Gag-Pro-Pol processing and this can be tested experimentally.

The role of residue 50 in altering inhibitor susceptibility

Residue 50 in protease is one of the few active site residues that change to confer differential PI resistance. Ile50 is often seen substituted to valine in patients failing amprenavir (APV) and DRV therapy (Partaledis et al., 1995a; Tisdale et al., 1999; Van Marck et al., 2007). Patients failing atazanavir (ATV) therapy predominantly accumulate the I50L substitution in protease (Colonno et al., 2004a; Colonno et al., 2004b). Using binding thermodynamics data described in Chapter IV, I have shown that changes at residue 50 change the susceptibility profiles of APV, DRV and ATV. By analyzing crystal structures I have shown that changes at residue 50 can cause subtle structural changes that alter protease-inhibitor interactions. These altered protease-inhibitor interactions are the likely cause for the varied susceptibility profiles observed with the binding data.

Clinical observations show that the I50L substitution makes the protease hypersusceptible to other inhibitors (Colonno et al., 2004b). Binding data obtained for the I50L variant appear to support this observation. Interestingly, the binding data show that the I50V substitution makes the protease more susceptible to ATV. Hypersusceptibility of the I50V variant to ATV has not been well described in the literature. However, a recent report describes a case where ATV hypersusceptibility was observed in combination with cleavage site mutations (Wainberg, Martinez-Cajas, and Brenner, 2007).

Since residue 50 is within the active site, substitutions at this position can significantly affect protease activity. A number of studies have indeed shown that enzymatic activity and replicative capacity are affected by substitutions to Ile50. The A71V substitution partially restores defects resulting from substitutions at residue 50 (Colonno et al., 2004b; Pazhanisamy et al., 1996). Furthermore, analysis of Gag cleavage sites has shown that the I50V substitution is often associated with co-evolution of the p1-p6 substrate (Kolli et al., 2009). Moreover, substitutions at residue 50 often occur in combination with a number of secondary mutations (Colonno et al., 2004b; Pazhanisamy et al., 1996). Therefore, future directions need to examine substitutions at residue 50 in combination with common secondary mutations as well as in combination with cleavage site substitutions.

Conclusions

Collectively, my dissertation research provides mechanistic insights into how amino acid substitutions within protease modulate protease activity and contribute to inhibitor resistance. Furthermore, these studies have examined how clade specific sequence variations contribute to inhibitor resistance. The genetic diversity of HIV-1 continues to grow as HIV-1 prevails across the world. There is evidence to suggest that genetic variation is a factor in viral transmission and disease progression as well as in the emergence of inhibitor resistance. Thus, it is essential that we understand how genetic variations modulate HIV drug targets and contributes to inhibitor resistance. Being able to understand how drug resistance occurs at a molecular level can provide valuable insights that can help in the design of robust PIs that are not susceptible to resistance substitutions. Furthermore, until a definitive cure is developed to treat HIV-1 infection and AIDS, mechanistic insights into drug resistance can aid in the formulation of better treatment strategies to slow down the progression of HIV infections to AIDS.

APPENDIX I

Structural and Biochemical Studies on HIV-1 Clade C Protease

INTRODUCTION

Based on genomic diversity the HIV-1 is classified into nine major clades (A, B, C, D, F, G, H, J and K) and 43 circulating recombinant forms (Los Alamos Laboratory, 2010; Robertson et al., 2000). While the majority of knowledge available on HIV-1 is based on studies carried out on clade B viral strains, non-B clade infections account for over 80% of HIV-1 infections in the world. In recent years, several studies have observed that disease progression and modes of transmission can differ between clades (Hu et al., 1999; Kanki et al., 1999a; Kiwanuka et al., 2008). Thus, given that currently available anti-HIV therapeutics were developed using studies on clade B strains, these observations raise legitimate questions regarding the efficacy of currently available anti-HIV therapies against different clades.

Clade C is the most predominant form and accounts for over 50% of the infections in the world and is primarily seen is sub-Saharan regions and in south and central Asia (Buonaguro, Tornesello, and Buonaguro, 2007; Wainberg, 2004). Several clinical studies have observed that clade C has differences in the mode of viral transmission as well as progression to AIDS when compared to other clades (Essex, 1996; Kanki et al., 1999a; Renjifo et al., 2004). Furthermore, the clade C protease has also been observed to have a different resistance pathway from that of clade B to confer resistance to the protease inhibitor nelfinavir (NFV). In patients infected with clade C, the protease predominantly accumulates the L90M substitution while D30N/ N88D is the most common NFV resistance profile in patients with clade B. Thus, additional studies on non-B clades are warranted in order to better understand the reasons for these observed differences so that better treatment strategies can be formulated.

The amino acid sequence of clade C protease differs by ~10% when compared to the clade B protease with all the amino acid differences mapping to positions outside the active site (Figure AI-1). Several of these amino acid differences are associated with protease inhibitor resistance in clade C. The biochemical and biophysical data presented in this appendix are an effort to better understand how sequence variations in clade C when compared to clade B affect enzyme activity and inhibitor binding as well as to determine how they might influence pathway to NFV resistance.

MATERIALS AND METHODS

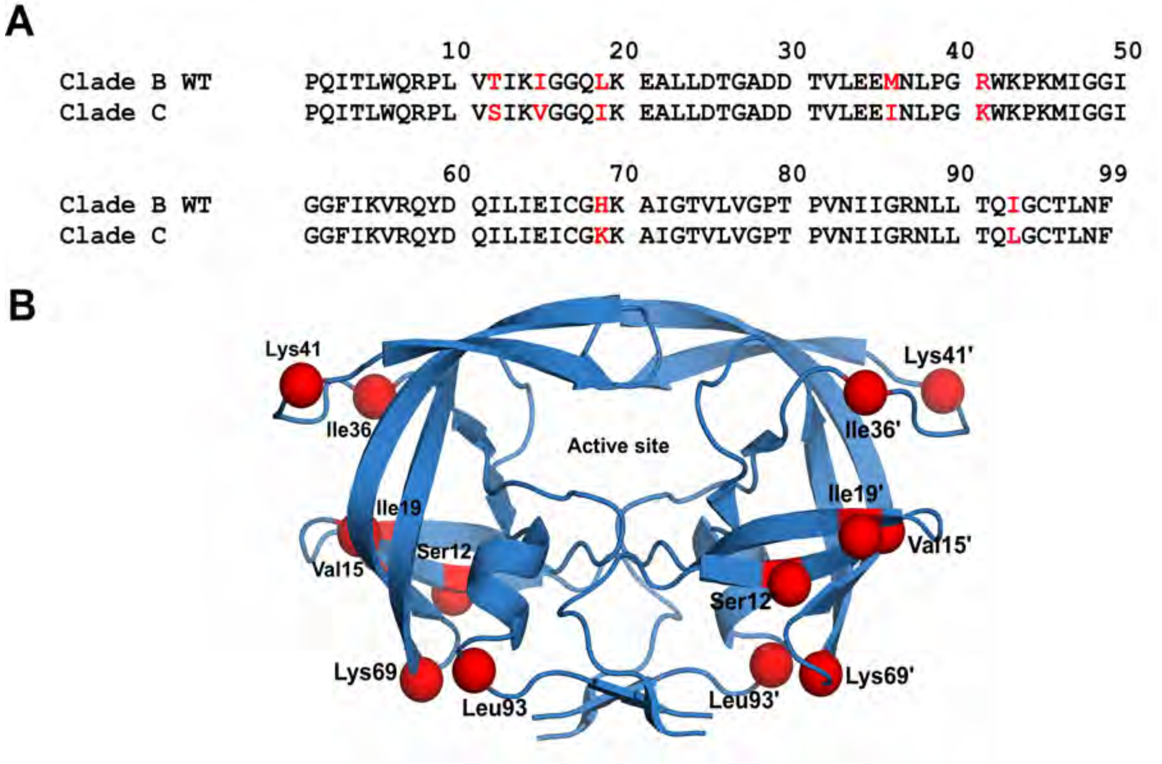
Protease-gene construction

The clade C wild type (AE-WT) protease gene was synthesized (Integrated DNA Technologies, Coralville, IA), with codons optimized for expression in *Escherichia coli*. Q7K, L33I and L63I mutations were introduced into the protease gene to prevent autoproteolysis. The gene was subcloned into the pET11a expression vector (Novagen/EMD Chemicals, Gibbstown, NJ). The protease sequence was confirmed by DNA sequencing.

Protein expression and purification

The clade C gene was subcloned into the heat-inducible pXC35 expression vector (American Type Culture Collection (ATCC), Manassas, VA) and transformed into *E. coli* TAP-106 cells. Protein overexpression, purification and refolding was carried out, as

FIGURE AI-1



```
                           10         20         30         40         50
Clade B WT       PQITLWQRPL VTIKIGGQLK EALLDTGADD TVLEEMNLPG RWKPKMIGGI
Clade C          PQITLWQRPL VSIKVGGQIK EALLDTGADD TVLEEINLPG KWKPKMIGGI

                           60         70         80         90         99
Clade B WT       GGFIKVRQYD QILIEICGHK AIGTVLVGPT PVNIIGRNLL TQIGCTLNF
Clade C          GGFIKVRQYD QILIEICGKK AIGTVLVGPT PVNIIGRNLL TQLGCTLNF
```

A
B
Lys41
Ile36
Lys41'
Ile36'
Active site
Ile19
Ser12
Val15
Ile19'
Val15'
Ser12'
Lys69
Leu93
Leu93'
Lys69'

FIGURE AI-1. A. Amino acid sequence alignment of clade B and C protease. Sequence differences between the two clades are indicated in red. **B.** Amino acid sequence differences in clade C map to positions outside the active site. Sequence differences are indicated in red spheres.

previously described (King et al., 2002). Protein used for crystallographic studies was further purified with a Pharmacia Superdex 75 fast-performance liquid-chromatography column (GE Healthcare, Chalfont St. Giles, U.K.) equilibrated with refolding buffer (50mM sodium acetate pH 5.5, 10% glycerol, 5% ethylene glycol, and 5mM dithiothreitol).

Isothermal titration calorimetry (ITC)

Binding affinities and thermodynamic parameters of inhibitor binding to clade C protease were determined by ITC on a VP isothermal titration calorimeter (MicroCal, LLC, Northampton, MA). The buffer used for all protease and inhibitor solutions consisted of 10 mM sodium acetate (pH 5.0), 2% dimethyl sulfoxide, and 2 mM *tris*[2-carboxyethyl]phosphine. Binding affinities for all protease variants were obtained by competitive-displacement titration using acetyl-pepstatin as the weaker binder. A solution of 30–45 μM protease was titrated with 10-μL injections of 200 μM acetyl-pepstatin to saturation. The pepstatin was then displaced by titrating 36 8-μL injections of 200 μM amprenavir (APV) or NFV or 41 7-μL injections of 40 μM DRV. Heats of dilution were subtracted from the corresponding heats of reaction to obtain the heat resulting solely from the binding of the ligand to the enzyme. Data were processed and analyzed with the ITC data analysis module (Microcal) for Origin 7 data analysis and graphing software (OriginLab, Northampton, MA). Final results represent the average of at least two measurements. ITC data were compared with previously described clade B data (Chapter 3).

Crystallization trials

Protease solutions between 1.0 and 2.0 mg mL^{-1} were equilibrated with a fivefold molar excess of NFV, darunavir (DRV) for 1 hour on ice. Crystal trials were set up with a reservoir solution consisting of 126-mM phosphate buffer (pH 6.2), 63-mM sodium citrate and 18% to 23% ammonium sulfate by the hanging-drop vapor-diffusion method. Additional crystal screens were set up with the Hampton Crystal Screen and Crystal Screen 2 (Hampton Research, Aliso Viejo, CA) by the sitting-drop vapor diffusion method using a PHENIX crystallization robot (Art Robbins Instruments, Sunnyvale, CA).

Measurement of protease activity

Clade B and C protease activity was assayed by following each variant's ability to hydrolyze the fluorogenic substrate HiLyte Fluor 488-Lys-Ala-Arg-Val-Leu-Ala-Glu-Ala-Met-Ser-Lys (QXL-520) (AnaSpec Inc., Fremont, CA) that corresponds to the HIV-1 Ca-p2 substrate. The assay was carried out in a 96-well plate, and the enzymatic reaction was initiated by adding 20 μL of a solution of 100–250 nM protease to 80 μL of substrate solution. The buffer used in all reactions consisted of 10 mM sodium acetate (pH 5.0), 2% dimethyl sulfoxide, and 2 mM *tris*[2-carboxyethyl]phosphine. Final concentrations in each experiment were 0–40 μM substrate and 20–50 nM protease. Accurate concentrations of properly folded active protease were determined by carrying out ITC experiments for each variant with acetyl-pepstatin as described in the previous section. Fluorogenic response to protease cleavage was monitored at 23°C using a Victor^3 microplate reader (PerkinElmer, Waltham, MA) by exciting the donor molecule at 485 nM and recording emitted light at 535 nM. Data points were acquired every 30 seconds.

The data points in relative fluorescence units (RFU) were converted into concentrations using standard calibration curves generated for HiLyte Fluor 488 at each substrate concentration. In addition to converting RFUs to concentrations, generating calibration curves at each substrate concentration allowed us to correct for the inner filter effect (Copeland, 1996). Rates of each enzymatic reaction were determined from the linear portion of the data and were fitted against substrate concentrations to determine K_M and k_{cat} (catalytic-turnover rate) values using VisualEnyzmics enzyme-kinetics software (SoftZymics, Princeton, NJ). Final results for each variant represent the average of at least three experiments.

RESULTS

Binding thermodynamics

Binding thermodynamics parameters of NFV, DRV and APV binding to clade C and clade B protease were determined in order to evaluate the effects of background sequence polymorphisms on inhibitor binding in clade C. The binding thermodymics parameters are compared to values from clade B in Table AI-1. Clade C protease had a 14.6-fold weaker affinity (K_d = 5.7 nM) for NFV when compared to clade B protease (K_d = 0.39 nM). Similarly, DRV had a 28.5-fold reduced affinity (K_d = 0.114 nM) compared to clade B (K_d = 0.004 nM). However, the affinity of APV binding to clade C protease was comparable to clade B. Overall, the binding of NFV, DRV and APV to clade C protease was enthalpically unfavorable (Table AI-1). However, the loss in binding

TABLE AI-1. Binding thermodynamic parameters for NFV, DRV and APV binding to Clade C and clade B protease variants.

Protease Variant	K_a (M$^{-1)}$	K_d (nM)	K_d ratio	ΔH (kcal mol^{-1})	ΔΔH	–TΔS (kcal mol^{-1})	Δ(–TΔS)	ΔG (kcal mol^{-1})	ΔΔG
NFV									
Clade B WT	$(2.6 \pm 0.5) \times 10^9$	0.39 ± 0.07	1.0	4.4 ± 0.1	-	–17.0	-	–12.6 ± 0.1	-
Clade C WT	$(1.7 \pm 1.8) \times 10^8$	5.7 ± 5.9	14.6	8.3 ± 0.3	3.9	–19.3	–2.3	–11.1 ± 0.05	1.5
DRV									
Clade B WT	$(2.2 \pm 1.1) \times 10^{11}$	0.004 ± 0.002	1.0	–12.1 ± 0.9	-	–3.1	-	–15.2 ± 0.3	-
Clade C WT	$(8.8 \pm 6.1) \times 10^9$	0.114 ± 0.080	28.5	–9.1 ± 1.2	3.0	–4.6	–1.5	–13.3 ± 0.04	1.9
APV									
Clade B WT	$(2.6 \pm 1.3) \times 10^9$	0.39 ± 0.20	1.0	–7.3 ± 0.9	-	–5.3	-	–12.6 ± 0.3	-
Clade C WT	$(1.8 \pm 0.1) \times 10^9$	0.56 ± 0.03	1.4	–4.73 ± 0.02	2.6	–7.7	–2.4	–12.4 ± 0.03	0.2

enthalpy was not compensated through a gain in binding entropy with NFV and DRV binding to clade C protease. This resulted in a net gain in the free energy of binding of NFV ($\Delta\Delta G$ = 1.5 kcal mol^{-1}) and DRV ($\Delta\Delta G$ = 1.9 kcal mol^{-1}) and thus, indicated that these binding reactions were unfavorable when compared to clade B.

Crystal trials

Clade B protease readily crystallizes in the phosphate buffer with ammonium sulfate as the precipitant. However, crystal trials with clade C protease with the same buffers yielded no crystal hits. Four crystal hits were obtained with the Hampton Crystal Screen and Crystal Screen 2. However, no x-ray diffraction was visible when the larger crystals were mounted on the in-house x-ray source.

Protease activity

The enzyme-kinetics experiments performed with the Ca-p2 fluorogenic substrate analog revealed that the clade B protease had a K_M of 16.7 ± 6.0 µM and a catalytic turnover rate (k_{cat}) of 1.79 ± 0.28 sec^{-1}. The catalytic efficiency was calculated to be 0.11 ± 0.04 sec^{-1} µM^{-1}. Using a substrate concentration range of 0 – 40 µM it was not possible to fit the initial velocities for clade C protease to the Michaelis–Menten equation.

DISCUSSION

HIV-1 clade C accounts for over 50% of HIV-1 infections in the world. Clade C protease carries a number of amino acid polymorphisms that are associated with inhibitor resistance in clade B. The primary pathway to NFV resistance in clade C protease is different from clade B. The results described in this section attempt to elucidate the effect of clade specific sequence polymorphisms in protease on enzyme activity and how they might influence inhibitor resistance pathways.

Based on the binding affinity data it is clear that clade C specific amino acid variations have a significant effect on NFV and DRV binding when compared to clade B protease. The weaker affinity for NFV likely permits the protease to confer resistance by accumulating the non-active site substitution, L90M. In comparison, the clade B protease that has a higher affinity for NFV requires a combination of an active site (D30N) and non-active site (N88S) substitutions in order to effectively gain NFV resistance. Therefore, by not altering the active site, the clade C protease maintains its ability to effectively recognize and cleave substrates while conferring inhibitor resistance.

Residue 90 is a non-active site residue and therefore, is not in a position to interact with ligands that bind within the protease active site (Figure AI-A). However, Leu90 is part of the termial helix and packs against the floor of the active site region that includes the catalytic aspartic acid residue (Figure AI-2B). Thus the L90M substitution most likely perturbs the packing interactions in this region. Mutating Leu90 in the wild type protease in complex with NFV (PDB code: 3EKX) to methionine *in silico* shows that Met90 side chain can pack against the floor region of the active site as well (Figure AI-C). However, based on this model the non-bonded electron pair in the sulfur atom in Met90 can serve as a hydrogen bond acceptor and interact the amide nitrogen of Asp25 (Gregoret et al., 1991; Zhou et al., 2009). Asp25 is important for inhibitor binding. Therefore, it is possible that Met90 can interact with the floor region of the active site and disrupt these interactions that are crucial for NFV binding.

FIGURE AI-2

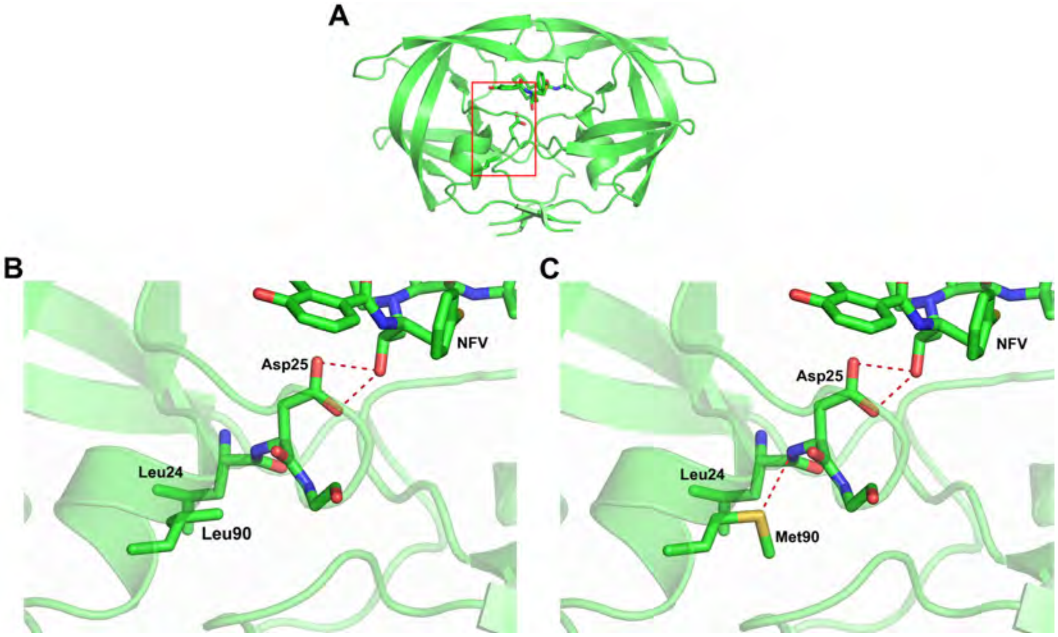

A
B
C
Asp25
NFV
Leu24
Leu90
Asp25
NFV
Leu24
Met90

FIGURE AI-2. A. L90M is a non-active site substitution. The red box indicates the region highlighted in panels B and C. **B.** The Leu90 side chain packs against the floor region of the active site in the clade B protease in complex NFV (PDB code: 3EKX). **C.** *In silico* modeling of methionine at residue 90 shows that the methionine side chain can interact with the active site backbone.

Based on the current data on clade C protease, background polymorphisms in clade C have an effect on inhibitor susceptibility. However, additional biochemical and structural studies that include the L90M protease variant are required to obtain a detailed understanding of the molecular basis for the altered resistance pathways in clade C protease.

APPENDIX II

Entropy-Enthalpy compensation in a Drug Resistant Variant of HIV-1 protease

ABSTRACT

The development of HIV-1 protease inhibitors has been the historic paradigm of rational structure based drug design, where structural and thermodynamic analyses have assisted in the development of novel inhibitors. Often the thermodynamics are analyzed in terms of particular properties of the inhibitors. In the current study profound changes in the thermodynamics of binding of a series of six FDA approved inhibitors are observed as a function of a drug resistant variant of HIV-1 protease. This protease variant has a combination of flap and active site mutations and exhibits extremely large entropy enthalpy compensation compared with wild-type, 5-15 kcal/mol, while only losing 1-3 kcal/mol in total binding energy for any given inhibitor. The co-crystal structures of this variant with four of the FDA approved inhibitors were determined and compared with both the wildtype complexes and complexes of another drug resistant variant that does not exhibit this energetic compensation. Conserved structural changes were observed, that likely contribute to the thermodynamic compensation. Although enthalpy-entropy compensation has been previously observed for a variety of systems, changes of this magnitude have not been reported. The fact that drug resistant mutations can profoundly modulate the relative thermodynamic properties of a therapeutic target independent of the inhibitor presents a new challenge for rational drug design.

INTRODUCTION

HIV-1 protease inhibitors (PI) were originally based on the substrate sequences as well as on the topology of the enzyme’s active site (Wlodawer and Erickson, 1993). The original structure-based drug design strategy was to optimize the entropy of binding by introducing conformational restraints into compounds so that they were pre-shaped to fit into the active site. In addition, these compounds are highly hydrophobic, resulting in an increase in solvation entropy upon binding. Thus, the first generation drugs bind with favorable entropy but with a corresponding loss in enthalpy (Todd et al., 2000). Some newer HIV-1 PIs (King et al., 2004c; Muzammil et al., 2007; Ohtaka et al., 2002; Surleraux et al., 2005a; Valzaquez-Campoy et al., 2000; Valzaquez-Campoy, Todd, and Freire, 2000)bound enthalpically favorably and often, as with Darunavir (DRV), with higher affinity, leading to the hypothesis that favorable enthalpy may aid in attaining better inhibitors that are less susceptible to drug resistance, however, Tipranavir (TPV) appears to be highly entropically driven (Muzammil et al., 2007).

The interplay between entropy and enthalpy in attaining high affinity is complex. Entropy-enthalpy compensation has been observed in many biological systems after relatively minor perturbations in the system including protein-metal interactions (Blasie and Berg, 2004; Kuroki, Nitta, and Yutani, 1992), cAMP receptor protein variants and RNA polymerase binding (Krueger et al., 2003), peptides binding to the Src Homology 2 domain of the Src kinase (Davidson et al., 2002), as well as ligands binding to cyclodextrin variants (Houk, 2003; Rekharsky and Inoue, 1998). This compensation, nearly equal and opposite, between the changes in $T\Delta S$ and ΔH usually of 1-2 kcal/mol often results in only minimal differences in the overall ΔG when comparing the binding of the different complexes (Sharp, 2001). The consequence of which makes it difficult to integrate the direct properties of enthalpy and entropy into rational drug design.

Drug resistant mutations in HIV protease throughout the enzyme can decrease the binding affinity with inhibitor molecules in a complex interdependent and cooperative manner (Ohtaka, Schon, and Freire, 2003; Olsen et al., 1999). Combinations of thermodynamic and structural studies to evaluate the consequences associated with drug-resistant mutations by many groups including our own (Bottcher et al., 2008; King et al., 2002; King et al., 2004b; King et al., 2004c; Lafont et al., 2007; Ohtaka et al., 2004; Steuber, Heine, and Klebe, 2007). Our earlier thermodynamic study on DRV and the chemically similar inhibitor amprenavir (APV), hypothesized a structural rationale for its unprecedented highly favorable enthalpy (King et al., 2004c). The single-ringed tetrahydrofuran (THF) group of APV was replaced with a double-ringed *bis*-THF in DRV, which formed additional protease-inhibitor interactions (King et al., 2004c) correlating with high affinity and highly favorable enthalpy. Although thermodynamics of binding is an equilibrium property between the liganded and unliganded forms of the enzymes, aspects of conformational changes in the bound structure may correlate with conserved thermodynamic changes.

In the present study, the crystal structures and thermodynamics are compared for the binding of the inhibitors APV, atazanavir (ATV), DRV, indinavir (IDV), nelfinavir (NFV) and saquinavir (SQV) of the wild-type (WT) HIV-1 protease with two multi-drug-resistant variants of the protease (Figure 1): (i) a variant (ACT) with the active site

FIGURE AII-1

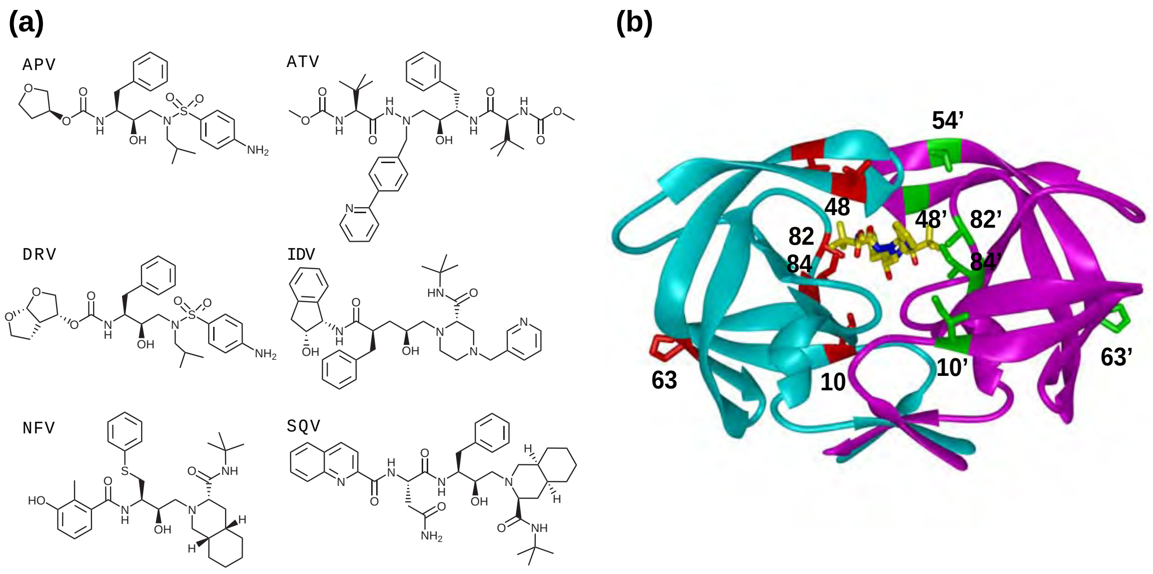

(a)
APV
ATV
DRV
IDV
NFV
SQV
(b)
54’
48
82
84
48’
82’
84’
63
10
10’
63’

FIGURE AI1-1 (a) Chemical structures of inhibitors. **(b)** Overview of the mutation sites of FLAP+ and ACT mutants mapped on an HIV-1 protease dimer. The monomers are distinguished in cyan and magenta, while the inhibitor is shown in yellow stick model. The mutation sites of FLAP+ (L10I/G48V/I54V/L63P/V82A) and ACT (L63P/V82T/I84V) are highlighted in red and green stick models.

mutations V82T and I84V; and (ii) a multi-drug-resistant variant (FLAP+) with L10I, G48V, I54V, and V82A derived as a combination of mutations that simultaneously occur in patients' sequences with a combination of flap mutations and active site mutations (Shafer, Stevenson, and Chan, 1999). Both drug resistant protease variants lose similar amounts of binding affinity relative to the WT protease. However the variant with flap mutations, FLAP+, exhibits extremely large and opposite changes in the entropy and enthalpy of interaction for all the inhibitors studied, indicating that the drug resistant mutations are directly modulating the relative thermodynamics of interactions.

METHODS

Protease gene construction

The WT protease gene was constructed using synthetic oligonucleotides optimized for *Escherichia coli* codon usage, and included a substitution of Q7K to prevent autoproteolysis (Rose, Salto, and Craik, 1993a). Mutations were introduced using the Quick Change site-directed mutagenesis kit (Stratagene, La Jolla, CA).

Protein expression and purification

HIV-1 protease was over-expressed in *E. coli* Tap106 cells using heat induction as previously described (King et al., 2002). The protease was extracted from inclusion bodies using 50% acetic acid (Hui et al., 1993). High molecular weight proteins were separated from the desired protease by size exclusion chromatography on a 2.1-L Sephadex G-75 superfine (Sigma Chemical) column equilibrated with 50% acetic acid. The protein was then refolded by rapid dilution in a 10-fold volume of 50 mM sodium acetate buffer at pH 5.5, also containing 10% glycerol, 5% ethylene glycol and 5 mM dithiothreitol (refolding buffer). The refolded protein was concentrated using an Amicon ultrafiltration cell, followed by dialysis to remove any residual acetic acid (King et al., 2002; King et al., 2004c; Prabu-Jeyabalan et al., 2004b). The protein was further concentrated to approximately 1 to 2 mg/ml and stored at -80℃. Protease used for crystallization was further purified with a Pharmacia Superdex 75 fast-performance liquid chromatography column equilibrated with refolding buffer.

Isothermal titration calorimetry

Isothermal titration calorimetry experiments were carried out at 20℃ using a VP-ITC microcalorimeter (MicroCal, Inc., Northampton, MA). The protease first underwent buffer exchange using PD10 gel filtration columns from Amersham Biosciences. All solutions were prepared in a buffer consisting of 10 mM sodium acetate (pH 5.0), 2% dimethyl sulfoxide, and 2 mM Tris (2-carboxyethyl) phosphine as final concentrations. Due to the high affinity of the WT protease for each of the inhibitors as well as the sharp transition to saturation, the thermodynamic parameters for these complexes were obtained by the displacement titration method (King et al., 2004c; Ohtaka et al., 2002; Sigurskjold, 2000; Valzaquez-Campoy et al., 2000). Likewise, affinities for the mutant enzymes with the tight binding inhibitors APV and DRV were performed using the competition experiments. The small magnitude of the signal for the ATV-ACT complex also necessitated the use of the displacement titration method in order to accurately determine its binding parameters. With one exception, acetyl-pepstatin (Bachem Bioscience, Inc., King of Prussia, PA) was used as the competing ligand at concentrations ranging from 250 to 400 μM. IDV was used as the weaker binder in displacement experiments with the ACT enzyme and DRV as described (King et al., 2004c) WT and mutant proteases at concentrations of approximately 6 to 20 μM pure enzyme were used in displacement experiments. Concentrations of reactive, folded protease reported here were determined during curve fitting using the MicroCal software Origin 7. This procedure involves adjusting the protease concentration to the value which results in a stoichiometric ratio of inhibitor to enzyme at half saturation. The protease in the calorimetric cell was initially titrated with the weaker binding inhibitor, followed by titration with the higher affinity ligand for displacement of acetyl-pepstatin (or IDV). APV, ATV, DRV, IDV, NFV and SQV were used at concentrations of 150 to 250 μM. The thermodynamic parameters for APV and DRV with the WT protease were obtained as previously described (King et al., 2004c).

Due to the decrease in the binding affinity of ATV, IDV, NFV and SQV with FLAP+, the thermodynamic parameters for these complexes were obtained by direct titrations. The same was true for experiments done on ACT with IDV, NFV, and SQV. In each case 200 μM inhibitor was titrated directly into the calorimetric cell containing protease at concentrations ranging from 16 to 26 μM. The affinities of the mutants for APV and DRV were of sufficient magnitude to require displacement titrations. (Experimental conditions used for the ACT enzyme with APV and DRV are described elsewhere (King et al., 2004c). For the FLAP+ displacement experiments, acetyl-pepstatin was used as the competing ligand as above at a concentration of 300 μM in experiments with 250 μM APV, while 900 μM acetyl-pepstatin was required for titrations with 66 μM DRV. (For experiments carried out using the displacement method, direct titrations were also performed in order to confirm the enthalpy changes acquired through competition experiments.) All experiments were carried out at least twice and their mean values are reported. Heats of dilution obtained after saturation were subtracted from the heats of reaction in order to obtain the heat due solely to the inhibitor binding to the enzyme. Data were analyzed using the Origin 7 software package provided by MicroCal.

In order to calculate the heat capacity change (ΔC_p) associated with binding of APV and DRV to WT and FLAP+ variants, the enthalpy change in binding was determined by titrating 200 μM APV or 84 μM DRV into 13 to 50 μM protein in the calorimetric cell between a temperature range of 10 to 42℃.

Crystal structures:

The following nomenclature will be followed to refer to each crystal structure: $\text{Inhibitor}_{\text{protease variant}}$. For example, APV_{WT}, APV_{ACT} and $\text{APV}_{\text{FLAP+}}$ represent the WT, ACT and FLAP+ complexes of APV, respectively. Inhibitors APV, ATV, DRV, IDV, NFV and SQV were cocrystallized with ACT and FLAP+. To preserve consistency with our analysis, WT protease complexes were also determined. However, crystals for the SQV_{WT} complex could not be grown and hence the structure available in PDB was used for the entire analysis (Krohn et al., 1991) (PDB code:1HXB). The $\text{NFV}_{\text{FLAP+}}$ complex exhibited a different flap conformation while crystals for $\text{IDV}_{\text{FLAP+}}$ complex could not be grown. Hence, structural analyses for NFV and IDV were omitted.

Crystallization and data collection

Crystal screens were set up with a three- to five-fold molar excess of inhibitor to protease to ensure ubiquitous binding. The final protein concentration ranged from 0.5-2.5 mg/ml in refolding buffer. The hanging drop vapor diffusion method was used for crystallization as previously described (King et al., 2004c). With two exceptions, crystal screens were set up at ambient temperature using reservoir solutions consisting of 126 mM phosphate buffer at pH 6.2, 63 mM sodium citrate and ammonium sulfate at a range of 25 to 33 %. Crystals for the $\mathrm{APV_{FLAP+}}$ and $\mathrm{DRV_{FLAP+}}$ complexes however, were grown at 4˚C using 10:1 molar ratios of inhibitor to protein. The buffer used in each case consisted of 50 mM citrate phosphate at pH 5.0 with 7% DMSO, and ammonium sulfate at concentrations of 38% and 28% for APV and DRV respectively.

Crystallographic data for all of the complexes discussed in this analysis was collected on an RAXIS IV. The raw frames were indexed and integrated using DENZO and subsequently scaled using ScalePack (Minor, 1993; Otwinowski, 1993). All of the complexes, except two, crystallized in the usual orthorhombic crystal form with isomorphous cell dimensions. $\mathrm{APV_{FLAP+}}$ and $\mathrm{DRV_{FLAP+}}$ crystallized in an unusual hexagonal space group with 12 HIV-1 protease dimers per unit cell. The data collection statistics are listed in Table AII-1.

TABLE AII-1. Crystallographic statistics.

Parameters	**FLAP+** (L10I/G48V/I54V/L63P/V82A)				**WT**				**ACT** (L63P/V82T/I84V)			
	APV	**ATV**	**DRV**	**SQV**	**APV**	**ATV**	**DRV**[a,b]	**SQV**[c]	**APV**[RT,a]	**ATV**	**DRV**[a,b]	**SQV**[RT]
Resolution (Å)	2.15	1.6	1.97	2.2	1.8	1.7	1.2	2.3	2.2	1.6	1.35	2.0
Space group	$P6_1$	$P2_12_12_1$	$P6_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P6_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Unit Cell a (Å)	92.0	51.2	92.0	51.2	50.7	51.0	54.9	63.3	51.1	50.7	50.9	51.3
b (Å)	-	58.5	-	59.3	57.4	58.8	57.8		59.3	58.1	58.0	59.3
c (Å)	106.2	61.3	106.2	61.4	61.7	61.3	62.0	83.5	61.8	61.4	61.7	62.1
Z[d]	6	4	6	4	4	4	4	6	4	4	4	4
Total reflections	161411	175086	202358	37340	49469	68018	302022		49950	87982	256671	81540
Unique reflections	27531	24185	34578	9789	14987	19646	55056		9705	23103	39998	12975
Completeness (%)	99.5	96.6	95.8	98.3	79.7	93.5	95.5		96.8	93.6	98.0	97.2
R_{merge} (%)	5.4	3.3	5.0	4.6	2.9	5.0	3.8		8.7	4.8	4.4	7.0
I/s_I	11.1	17.1	15.4	11.4	19.0	11.3	25.0		5.9	11.7	13.5	6.0
REFINEMENT	Refmac5	Refmac5	Refmac5	Refmac5	Refmac5	Refmac5	Refmac5	X-PLOR	CNS	Refmac5	Refmac5	CNS
R factor (%)	19.6	18.7	20.9	19.5	19.0	17.3	14.1	16.1	20.3	19.2	16.3	19.0
R_{free} (%)	25.5	22.2	25.6	24.6	22.6	20.5	17.9	NA	24.4	21.7	20.0	22.3
RMSD in												
bond length (Å)	0.009	0.006	0.007	0.008	0.007	0.007	0.004	0.013	0.007	0.006	0.005	0.007
bond angle (˚)	1.4	1.4	1.1	1.3	1.4	1.2	1.5	2.9	1.4	1.3	1.4	1.3

[a] PDB codes: DRV_{WT} – 1T3R; DRV_{ACT} – 1T7I; APV_{ACT} – 1T7J

[b] Data collection carried out at Advanced Light Source (ALS)

[c] PDB code 1HXB

[d] Number of dimeric molecules in the unit cell.

[RT] Room temperature

Structure solution and crystallographic refinement

The suite of programs from CCP4i (Collaborative-Computational-Project, 1994) was used for most of the crystallographic operations. Structure solution for all complexes was carried out using the molecular replacement package AMoRe (Navaza, 1994). The substrate complex of capsid-p2 bound with D25N inactive HIV-1 protease (Prabu-Jeyabalan, Nalivaika, and Schiffer, 2000a) (PDB.code: 1F7A) was used as the starting model. This substrate structure was used to solve inhibitor complexes to reduce model bias. For the orthorhombic crystals, the structure solution was straightforward. However, for the APV_{FLAP+} and DRV_{FLAP+} complexes, which crystallized in a hexagonal form, a combination of self-rotation maps were computed, revealing that the space group is $P6_1$ with two dimers per asymmetric unit. The subsequent refinement strategy involving ARP/wARP, TLS parameters and Refmac5 was similar to our earlier structural analyses (Prabu-Jeyabalan et al., 2006a). Interactive model building was conducted using the package O (Jones, Bergdoll, and Kjeldgaard, 1990) and the quality of the structures was assessed using PROCHECK (Laskowski et al., 1993). The dimers of the APV_{FLAP+} and DRV_{FLAP+} complexes within the asymmetric unit were restrained by non-crystallographic symmetry (NCS) restraints. However, no NCS restraints were imposed between the monomers of the dimer for any of the complexes. The refinement statistics are provided in Table AII-2.

Structural analysis

(i) Structural superimpositions: The twelve inhibitor complexes chosen for structural comparisons were superimposed on a complex containing the peptide fragment

representing the Gag substrate capsid-p2 (PDB code 1F7A (Prabu-Jeyabalan, Nalivaika, and Schiffer, 2000a)) using the protease terminal domain (Pro1-Pro9 & Arg87-Phe99). This capsid-p2 complex was chosen to preserve consistency with our previous analyses (King et al., 2004b; King et al., 2004c).

(ii) Double difference plots: Double-difference plots, computed using Ca-Ca distances, reveal structural differences between similar structures without the bias due to structural superimpositions or space group differences (Prabu-Jeyabalan et al., 2006a; Williams and Kelley, 1998a). Distances between all the Cα atoms within each dimer were computed (${}^{n}\boldsymbol{D}_{ij}$, where $\boldsymbol{D}$ is the double difference and i and j are residue numbers). This was repeated for each of the "*n*" structures. Double differences ($\boldsymbol{D}$) were generated as a (i x j) matrix by computing the difference of the differences between the two dimers ($\boldsymbol{D} = {}^{\boldsymbol{n}}\boldsymbol{D}_{ij} - {}^{n'}\boldsymbol{D}_{ij}$). The ($i$ x j) matrix was then displayed as a contour diagram using GnuPlot (Williams and Kelley, 1998a).

(iii) Estimation of van der Waals potential: Inhibitor-protease van der Waals contacts were estimated by a simplified Lennard-Jones potential $\boldsymbol{V}(r)$ using the relation $4\varepsilon[(\sigma/r)^{12} - (\sigma/r)^{6}]$; where r is the protease-inhibitor inter-atomic distances, and ε and σ are the well depth and hard sphere diameter, respectively, for each protease-inhibitor atom pair. $\boldsymbol{V}(r)$ is computed for all possible protease-inhibitor atom pairs within 5 Å and when potentials for non-bonded pairs separated less than a distance corresponding to ε, $\boldsymbol{V}(r)$ were equated to ε. Using this simplified potential for each non-bonded protease-inhibitor pair, $\sum\boldsymbol{V}(r)$ was then computed for each protease residue.

RESULTS

Thermodynamics of inhibitor binding

Isothermal titration calorimetry was used to determine the thermodynamics of inhibitor binding to the WT and multi-drug resistant (MDR) proteases. Table AII-2 lists the thermodynamic parameters for each protease-inhibitor pair along with their corresponding K_d ratios (K_{dMDR}/K_{dWT}). In each case, there is a decrease in binding affinity to the mutant proteases relative to that of the WT. APV binding to each of the MDR proteases is the least affected in terms of affinity, followed by DRV. In addition, the FLAP+ mutations appear to have less of an effect on the binding of these two inhibitors when compared to that of the ACT mutant. The K_d ratios for ACT-APV and ACT-DRV binding are 5.9 and 14.7 respectively, while the same ratios for FLAP+ binding are only 3.3 and 5.8. ATV binding of ACT and FLAP+ exhibit a decrease in affinity by 17.8 and 48.4 fold, respectively. The binding of IDV is further compromised with K_d ratios of 49.0 and 76.9 for the ACT and FLAP+ proteases, respectively. Also, NFV binding to ACT and FLAP+ shows similar reductions in affinity with values of 48.2 and 87.0, respectively. In contrast, SQV binding is the most compromised of all the inhibitors studied, with K_d ratios of 135 for ACT and 353 for FLAP+. Thus APV and DRV, which fit well within the substrate envelope (King et al., 2004b; King et al., 2004c), are the most robust against the drug-resistant variants.

A closer examination of the specific enthalpic and entropic contributions to the changes in free energy (Table AII-2) reveals that the binding of the first generation

TABLE AII-2 Thermodynamics of the binding of inhibitors to WT, ACT and FLAP+ variants of HIV-1 protease.

(Experiments were performed at 20˚C.)

Protease variant	ΔG (kcal mol^{-1})	ΔH (kcal mol^{-1})	-TΔS (kcal mol^{-1})	K_a (M^{-1})	K_d (nM)	K_d ratio	ΔC_p (cal K^{-1} mol^{-1})
			APV				
WT	-12.6	-7.3 ± 0.9	-5.3	(2.6 ± 1.3) x 10^9	0.39 ± 0.20	1	-429
ACT	-11.6	-6.0 ± 2.0	-5.6	(4.4 ± 1.5) x 10^8	2.3 ± 0.79	5.9	
FLAP+	-11.9	3.3 ± 0.5	-15.2	(7.6 ± 0.06) x 10^8	1.3 ± 0.01	3.3	-534
			ATV				
WT	-12.9	-1.1 ± 0.1	-11.8	(4.5 ± 2.4) x 10^9	0.23 ± 0.12	1	
ACT	-11.3	-0.33 ± 0.1	-10.9	(2.5 ± 0.13) x 10^8	4.0 ± 0.21	17.8	
FLAP+	-10.7	4.5 ± 0.1	-15.2	(9.1 ± 1.3) x 10^7	10.9 ± 1.6	48.4	
			DRV				
WT	-15.2	-12.1 ± 0.9	-3.1	(2.2 ± 1.1) x 10^{11}	0.0045 ± 0.0023	1	-373
ACT	-13.6	-10.0 ± 0.05	-3.7	(1.5 ± 0.42) x 10^{10}	0.066 ± 0.02	14.7	
FLAP+	-14.2	2.0 ± 0.6	-16.2	(3.9 ± 0.75) x 10^{10}	0.026 ± 0.005	5.8	-507
			IDV				
WT	-12.2	1.7 ± 0.2	-14.0	(1.4 ± 0.3) x 10^9	0.74 ± 0.14	1	
ACT	-10.0	7.4 ± 0.03	-17.4	(2.8 ± 1.1) x 10^7	36.0 ± 13.5	49.0	
FLAP+	-9.7	10.9 ± 0.08	-20.6	(1.8 ± 0.07) x 10^7	56.5 ± 22.9	76.9	
			NFV				
WT	-12.6	4.4 ± 0.1	-17.0	(2.6 ± 0.5) x 10^9	0.39 ± 0.07	1	
ACT	-10.4	7.1 ± 0.1	-17.4	(5.3 ± 0.07) x 10^7	18.9 ± 0.25	48.2	
FLAP+	-10.0	10.8 ± 0.03	-20.8	(2.9 ± 0.11) x 10^7	34.1 ± 1.3	87.0	
			SQV				
WT	-12.5	3.6 ± 0.1	-16.1	(2.0 ± 0.1) x 10^9	0.50 ± 0.03	1	
ACT	-9.6	9.5 ± 0.03	-19.1	(1.5 ± 0.006) x 10^7	67.4 ± 0.28	135.1	
FLAP+	-9.1	12.9 ± 0.09	-21.9	(5.7 ± 0.14) x 10^6	176 ± 4.2	352.7	

inhibitors IDV, NFV, and SQV to each of the protease variants is entropically driven (negative TΔS) at the experimental temperature of 20˚C. In contrast, APV, ATV, and DRV are both enthalpically (negative ΔH) and entropically driven when binding to the WT and ACT proteases (Table1). However, the binding of these same compounds (APV, ATV, and DRV) becomes endothermic (positive ΔH) with the FLAP+ enzyme. In fact, the magnitude of the differences in both the enthalpic and entropic contributions to APV and DRV in particular are much greater for the FLAP+ variant (Figure 2a) when compared to those of the other inhibitors. The corresponding differences for APV and DRV when bound to the ACT variant are however less dramatic (Figure 2b) . The ΔΔH values for FLAP+ are 10.7 and 14.1 with APV and DRV respectively. The highly unfavorable enthalpy change is compensated for by a relatively large favorable change in the entropic component as demonstrated in the Δ-TΔS values of -10.0 and -13.1 for APV and DRV respectively. This enthalpy-entropy compensation gives rise to the very minimal overall change in the free energy for the binding of APV and DRV to the FLAP+ enzyme when compared to the same compounds binding to the WT protease (Figure 2a). Although inhibitors binding to the ACT protease also result in minimal overall ΔΔG values, the magnitude of the entropy-enthalpy compensation is much less than that of binding to the FLAP+ enzyme (Figure 2), thus suggesting the FLAP+ protease variant is energetically unique.

Heat capacities were measured to further elucidate these dramatic thermodynamic changes by measuring the enthalpy of the interactions at a variety of temperatures with APV and DRV. The temperature dependence of the enthalpy changes gives rise to heat

FIGURE AII-2

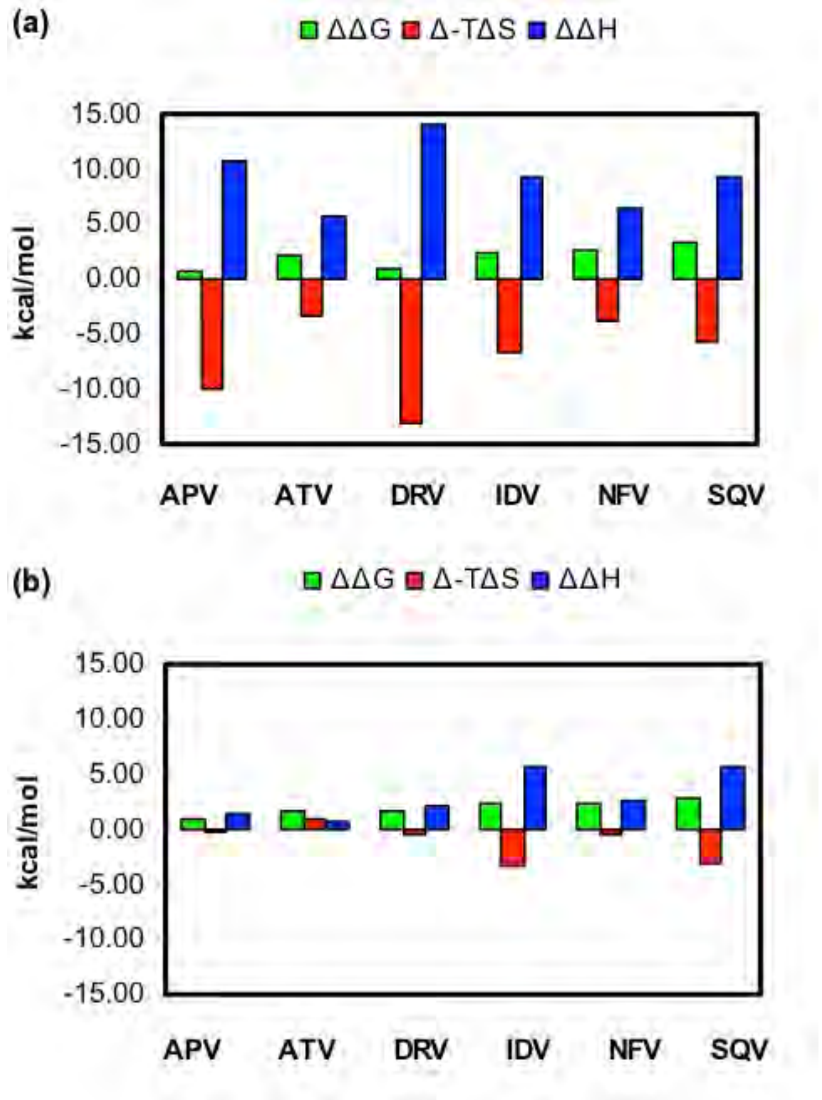

(a)
ΔΔG
Δ-TΔS
ΔΔH
15.00
10.00
5.00
0.00
-5.00
-10.00
-15.00
kcal/mol
APV
ATV
DRV
IDV
NFV
SQV
(b)
ΔΔG
Δ-TΔS
ΔΔH
15.00
10.00
5.00
0.00
-5.00
-10.00
-15.00
kcal/mol
APV
ATV
DRV
IDV
NFV
SQV

FIGURE AII-2 Differences in binding energetics between **(a)** WT and FLAP+ and **(b)** WT and ACT variants. The histograms representing the differences in ΔG, TΔS and ΔH are shown in green, red and blue, respectively.

capacity changes. The heat capacities obtained for the WT enzyme in association with APV and DRV were -429 and -373 cal.mol.K^{-1} respectively (Table AII-2). The substitutions in the FLAP+ protease gave rise to even larger (more negative) heat capacity changes for the corresponding inhibitor complexes with values of -534 and -507 cal.mol.K^{-1} for APVand DRV respectively. Thus, the five mutations in the FLAP+ variant have substantially changed the overall heat capacity for the binding of these two inhibitors by more than 100 cal.mol.K^{-1}. This change in overall heat capacity for binding further indicates that the mutations within FLAP+ fundamentally change the thermodynamic properties of this enzyme.

Overall features of the crystal structures:

The crystal structures of inhibitor-protease complexes involving WT, ACT and FLAP+ protease variants were determined. The IDV complex did not form diffraction quality crystals and the flaps in NFV_{FLAP+} exhibit an unusual conformation (Prabu-Jeyabalan et al., 2006a) and therefore, structural comparisons with those two inhibitor complexes are not included. Also, crystals for the SQV_{WT} complex could not be grown and therefore, the complex from the PDB was used (PDB code 1HXB (Krohn et al., 1991)). The three-dimensional structures for the twelve crystal complexes discussed here are consistent with other structures found in the database (Vondrasek and Wlodawer, 2002). Most of the structures have been determined at resolutions better than 2.0 Å The entire crystallographic statistics are listed in Table AII-2. The inhibitor in all of the complexes, except ATV_{WT} and SQV_{WT}, binds in a unique conformation. In the ATV_{WT} and SQV_{WT} complexes, the inhibitor binds in two orientations with nearly 50%

occupancy for each orientation. Also, in the ATV_{WT} structure, the second phenyl group at P1' does not have electron density, and therefore, it was not used during any of the subsequent structural analyses. Altogether, this analysis includes eight new HIV-1 protease-inhibitor complexes, three previously reported from our laboratory and one from the database, allowing for a detailed structural comparison of these drug resistant HIV-1 protease variants with the WT enzyme.

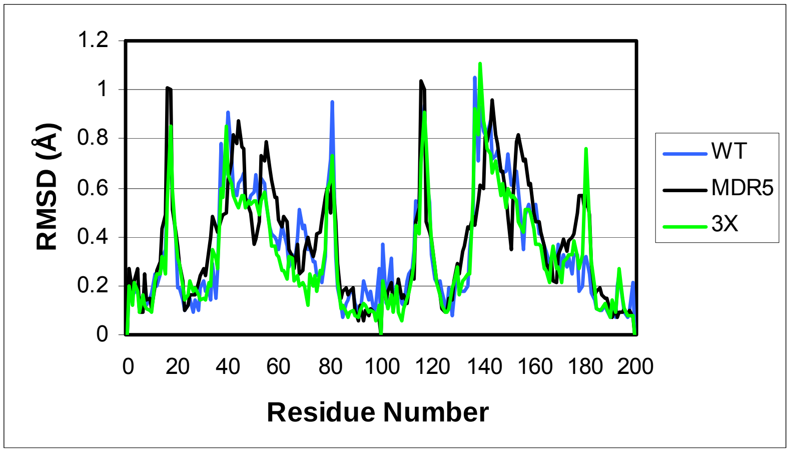
Structural comparisons between the protease complexes:

The twelve inhibitor complexes chosen for this structural analysis were superimposed as described in the Methods section. The root mean squared deviation (RMSD) in the coordinate positions for Ca atoms in each variant were calculated and plotted against residue numbers (Figure 3a). Although the distributions of RMSD values for the WT and ACT complexes follow the same pattern, the distribution observed in the FLAP+ complexes is significantly different. Not unsurprisingly, the FLAP+ structures exhibit the largest difference near the flap region. The RMSD in coordinates near Lys41 and Lys57 for both monomers of FLAP+ is the highest (> 0.8Å), while for the tip of the flap (Ile50) the RMSD value drops to <0.25Å. For the WT and ACT complexes, the RMSD values for the entire flap region remain uniform 0.5-0.65 Å. Thus the RMSD in coordinates suggest that the flap region of the FLAP+ complexes is invariant.

The RMSDs were mapped on an HIV-1 protease dimer model to illustrate which regions of the protease exhibit the most deviation (Figure 3b-d). This comparison further reveals that the FLAP+ (Figure 3b) complexes are different from the WT and ACT complexes (Figures c and d), especially at the flaps. Figure 3b demonstrates the high-

FIGURE AII-3

(a)



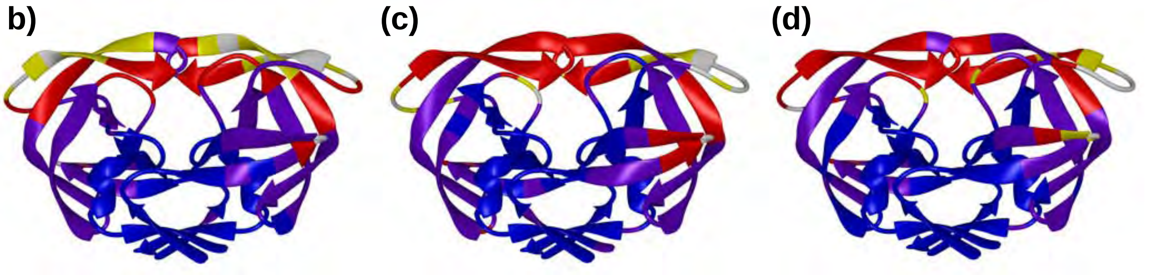


FIGURE AII-3 (a) Distribution of root mean squared deviations (RMSD) in Cα coordinates between the four inhibitor complexes of FLAP+ (black), WT (blue) and ACT (green) variants. The RMSD in Cα coordinates in **(b)** FLAP+, **(c)** WT and **(d)** ACT mapped on an HIV-1 protease dimer model. The color code for distinguishing the RMSD values are: blue – 0-0.25 Å; purple – 0.25-0.5 Å; red – 0.5-0.65 Å; yellow – 0.65-0.8 Å; white – 0.8 Å and above.

degree of structural conservation at the b-hairpin forming the tip of the flaps (Ile50-Gly51) in the FLAP+ complexes along with the significant variability in the b-strands flanking the tip. In contrast, the entire flap region in the WT and ACT complexes exhibits uniform structural variation (Figures 3c & d). Detailed comparison of the flaps in the FLAP+, WT and ACT complexes shows the conformational alterations in the FLAP+ structures (Figure 4a-c). While the structure of the flap tips for the WT and ACT complexes are conformationally variable, the FLAP+ complexes appear to converge to a single conformation regardless of which inhibitor is bound. Thus, the comparison of structures grouped by protease variant suggests that the flap tips of FLAP+ may have become rigid. The FLAP+ variant contains two flap mutations, G48V and I54V, flanking the flap tip, perhaps restricting the conformation of this region resulting in the apparent rigidity.

A pair-wise structural comparison was carried out for the three complexes of each inhibitor comparing WT vs FLAP+ and WT vs ACT using double-difference plots (Figure 5) (Prabu-Jeyabalan et al., 2006a; Williams and Kelley, 1998a). The double-difference plots displaying the deviations between the WT and FLAP+ complexes reveal several structural differences over 1 Å (Figures 5a-d). As illustrated in Figures 5a-d, the structural differences are chiefly concentrated around the flaps (Pro44-Lys57) and the P1-loop regions (Pro79-Ile84). In contrast with the changes observed between the WT-FLAP+ pairs, the plots computed for the WT-ACT pairs exhibit fewer peaks for all four inhibitors (Figures 5e-h), indicating that the structures of WT and ACT, when complexed to the same inhibitor, are similar to each other.

FIGURE AII-4

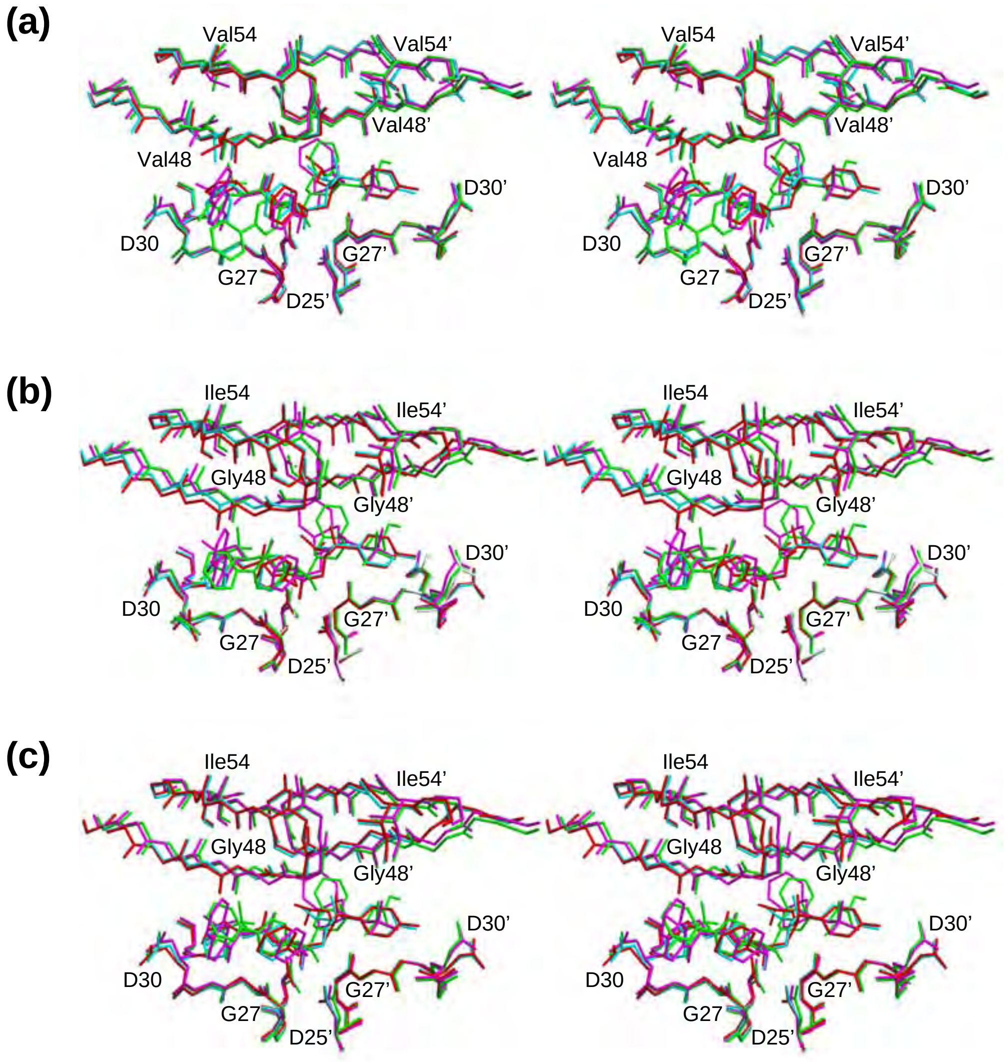

(a)
Val54
Val54’
Val48’
Val48
D30’
D30
G27’
G27
D25’
Val54
Val54’
Val48’
Val48
D30’
D30
G27’
G27
D25’
(b)
Ile54
Ile54’
Gly48
Gly48’
D30’
D30
G27’
G27
D25’
Ile54
Ile54’
Gly48
Gly48’
D30’
D30
G27’
G27
D25’
(c)
Ile54
Ile54’
Gly48
Gly48’
D30’
D30
G27’
G27
D25’
Ile54
Ile54’
Gly48
Gly48’
D30’
D30
G27’
G27
D25’

FIGURE AII-4 The active site region (Asp25-Asp30), flaps (Lys45-Lys55) and the inhibitors of **(a)** FLAP+, **(b)** WT and **(c)** ACT are superposed and illustrated as stereo pairs. The complexes involving APV, ATV, DRV and SQV are distinguished in red, green, cyan and magenta, respectively.

FIGURE AII-5

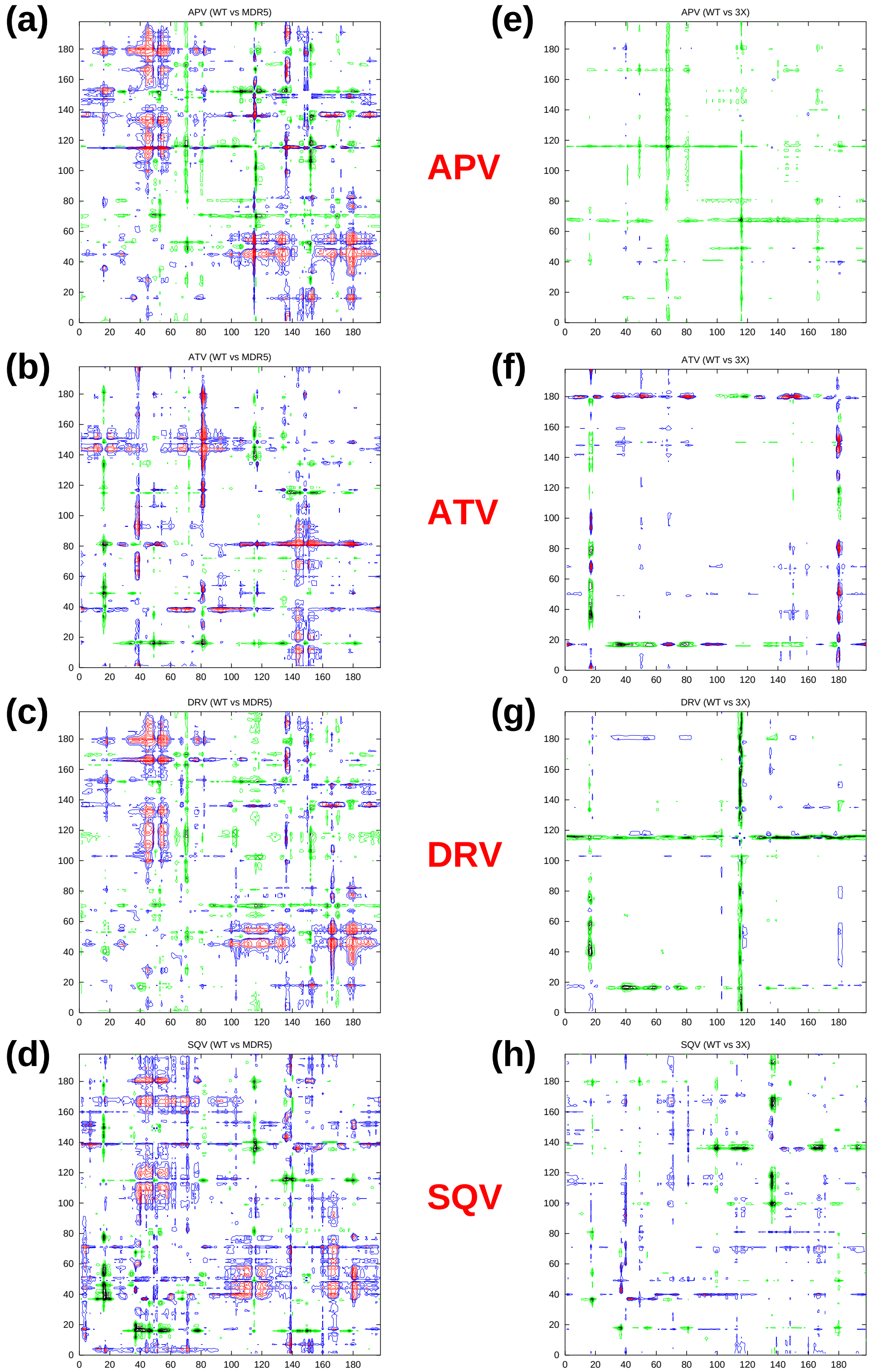

(a)
APV (WT vs MDR5)
(e)
APV (WT vs 3X)
APV
(b)
ATV (WT vs MDR5)
(f)
ATV (WT vs 3X)
ATV
(c)
DRV (WT vs MDR5)
(g)
DRV (WT vs 3X)
DRV
(d)
SQV (WT vs MDR5)
(h)
SQV (WT vs 3X)
SQV
0
20
40
60
80
100
120
140
160
180

FIGURE AII-5 Double difference plots illustrating the WT vs FLAP+ structural changes in **(a)** APV, **(b)** ATV, **(c)** DRV and **(d)** SQV, and WT vs ACT structural changes in **(e)** APV, **(f)** ATV, **(g)** DRV, **(h)** SQV. The key for contours: (i) black -5.0 to -1.0 Å and green -1.0 to -0.5Å (Corresponding residue distances in the mutant structures have increased); (ii) blue 0.5 to 1.0 Å and magenta 1.0 to 5.0 Å (Corresponding distances in the mutant structures have decreased).

The majority of the structural differences between the WT and FLAP+ structures occur primarily in one monomer. The flap of that monomer (Pro44-Lys57) of the FLAP+ variant, in relation to the WT complexes, has moved closer to the non-flap regions (1’-43’ and 58’-99’) of the opposite monomer by over 1 Å (Figure 5a-d). The flap-specific changes are observed between the FLAP+ and WT complexes and not between ACT and WT, in both the double-difference plots and RMSD plots (Figure 4). This conserved asymmetric structural change observed in the FLAP+ variant could be a factor leading to its unique thermodynamic characteristic.

Inhibitor-Protease Hydrogen bonds:

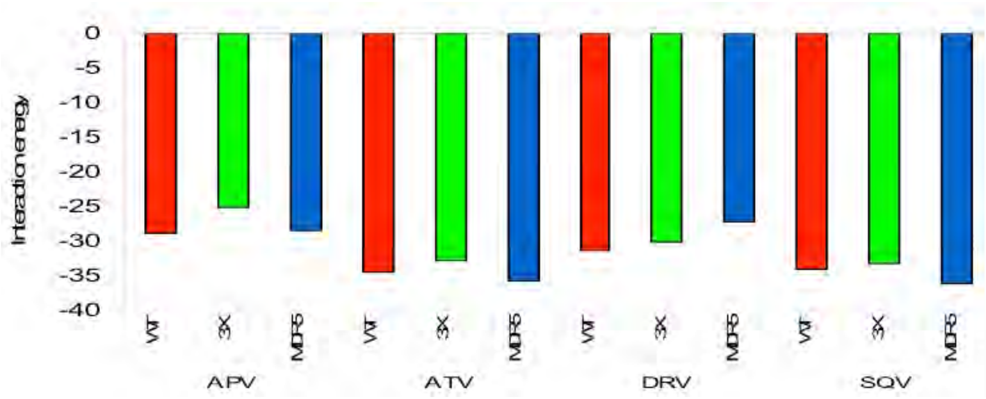
Hydrogen bonds between the inhibitor and the protease atoms for each variant were also computed. In each protease variant, the inhibitors APV, ATV, DRV and SQV form 9, 13, 11, and 12 hydrogen bonds, respectively. The APV_{WT} and APV_{ACT} complexes form two hydrogen bonds between Asp29 N and Asp30 N and the inhibitor atom O6 which are absent in the FLAP+ complex. However, these two interactions are maintained in all the complexes of DRV. One new hydrogen bond is observed in both the APV and DRV FLAP+ complexes between Asp30’ OD1 and APV_{FLAP+} (N3) or DRV_{FLAP+} (N1), respectively. Another feature in all of the APV and DRV complexes is the absence hydrogen bonds with the flaps, while ATV and SQV inhibitors form 4 and 2 hydrogen bonds, respectively, to the backbone of residue 48. Also, the N to Gly48 O hydrogen bond found in the SQV_{WT} and SQV_{ACT} does not exist in SQV_{FLAP+}. Thus specific changes in the hydrogen bond patterns for each of the inhibitors are observed in the FLAP+ complexes that are not observed in the corresponding inhibitor complexes of the other two protease variant complexes.

Inhibitor-Protease van der Waals contacts:

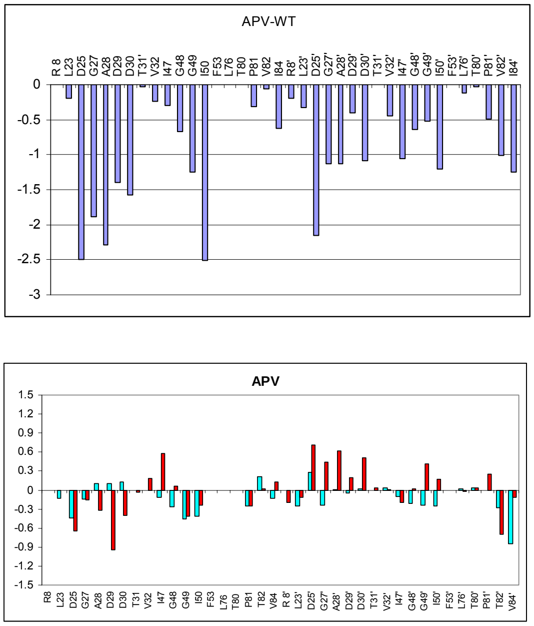
Estimates of total van der Waals contacts between the inhibitor and the protein were calculated with a simplified Lennard-Jones potential. Not unsurprisingly, the larger inhibitors ATV and SQV inhibitors had more favorable van der Waals interactions then the smaller ones like APV and DRV, especially in the FLAP+ complexes. Probably indicative of the locations of the mutations, the total van der Waals inhibitor-protease contacts for the ACT complexes of APV, ATV and SQV are slightly less favorable than in the corresponding WT and FLAP+ complexes (Figure 6a). The difference in van der Waals interaction for each protease residue ($\Delta \boldsymbol{V}(r)$) for the FLAP+ and ACT complexes with respect to the WT were also examined (Figure 6b-e). The figure panels also show the actual van der Waals for the WT ($\boldsymbol{V}(r)_{WT}$) for comparison. This analysis reveals that increase in van der Waals contacts by one inhibitor-protease atom pair is usually compensated by a decrease in contacts between another pair and therefore, the net change in van der Waals contacts ($\Sigma\Delta \boldsymbol{V}(r)$) is negligible. However, the absolute change in van der Waals contacts ($\Sigma|(\Delta \boldsymbol{V}(r)|$) will be significant. Thus, the absolute values of the difference in van der Waals contacts relative to the APV_{WT} and DRV_{WT} were analyzed. The APV_{FLAP+} and DRV_{FLAP+} complexes exhibit the largest differences (APV:9.1 and DRV:13.1), while smaller differences were observed in APV_{ACT} and DRV_{ACT} (APV:5.7 and DRV:4.4) (Figures 6b & d).Thus, the inhibitor-protease packing in the APV and

FIGURE AII-6

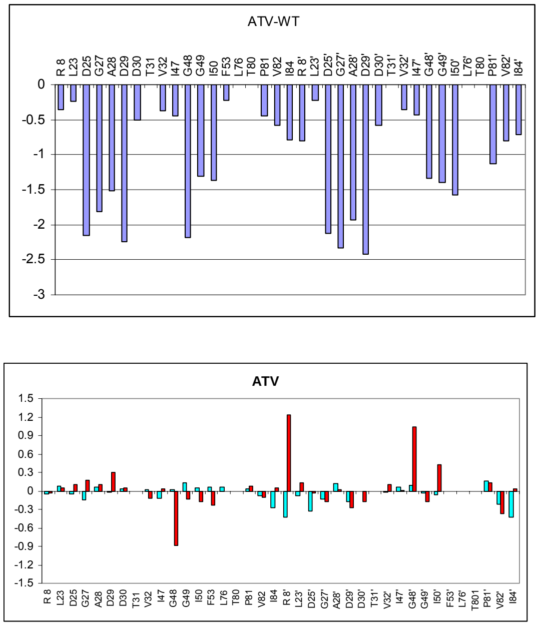
(a)



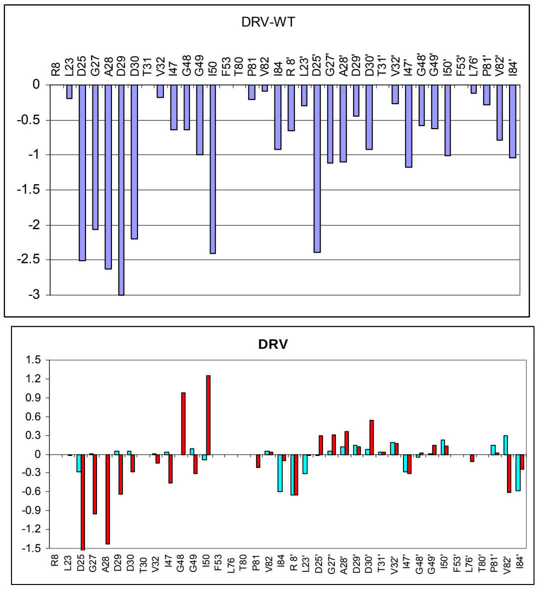
(b)


APV-WT
0
-0.5
-1
-1.5
-2
-2.5
-3
APV
1.5
1.2
0.9
0.6
0.3
0
-0.3
-0.6
-0.9
-1.2
-1.5

(c)


ATV-WT
0
-0.5
-1
-1.5
-2
-2.5
-3
ATV
1.5
1.2
0.9
0.6
0.3
0
-0.3
-0.6
-0.9
-1.2
-1.5

(d)


DRV-WT
0
-0.5
-1
-1.5
-2
-2.5
-3
DRV
1.5
1.2
0.9
0.6
0.3
0
-0.3
-0.6
-0.9
-1.2
-1.5

(e)

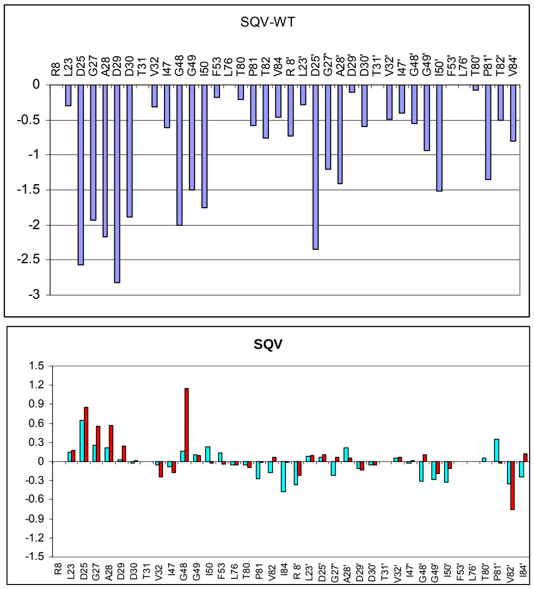

SQV-WT
0
-0.5
-1
-1.5
-2
-2.5
-3
SQV
1.5
1.2
0.9
0.6
0.3
0
-0.3
-0.6
-0.9
-1.2
-1.5

FIGURE 6a Histogram representation of the total inhibitor-protease van der Waals interaction energies for WT (gray), ACT (white) and FLAP+ (black). Residue-wise distribution of interaction energy is shown in **(b)** APV, **(c)** ATV, **(d)** DRV and **(e)** SQV. The distribution of energies in the WT complexes are shown in the upper panels while the lower panels illustrate the WT vs FLAP+ (black) and WT vs ACT (white) differences in energy distribution.

DRV complexes of the FLAP+ variant are significantly different. Similar comparisons for ATV and SQV reveal smaller changes in absolute differences in van der Waals contacts either between WT and FLAP+ (ATV:6.5 and SQV:6.9) or WT and ACT (ATV:3.6 and SQV:6.2) (Figures 6c and e). The van der Waals interactions for the active site protease residues were examined (Figures 6b-e) in detail. In the APV_{FLAP+} and DRV_{FLAP+} complexes, an increase in van der Waals interactions occurs in residues Asp25-Asp30, while a decrease in interactions is seen for the same residues in the second monomer (Asp25'-Asp30'). However, these changes in interactions are not observed in the ATV and SQV complexes. Interestingly, although the flap regions were observed to have large structural changes (Figure 3 and 4) these actually resulted in almost no changes in the extent of van der Waals changes. Overall, APV_{FLAP+} and DRV_{FLAP+} complexes undergo several structural changes to preserve their net van der Waals contacts.

To assess the direct affect of the resistance mutations, inhibitor atoms that lie within van der Waals distance of the sites of mutations, L10I, G48V, I54V, L63P, and V82A, were examined in detail. The sidechain of Val48 forms interactions with the inhibitor in SQV_{FLAP+}, while the inhibitor in all of the other FLAP+ complexes interact only with the backbone of Val48. The V82A mutation site is the only other substitution involving direct contact with inhibitors in FLAP+ while in ACT the substitution is V82T. In relation to Thr82 in the ACT complexes (0.3), the FLAP+ complexes exhibit a loss in contacts by Ala82 (0.7). Despite this loss in interaction, due to the valine to alanine substitution, the three-dimensional arrangement of this region, among the FLAP+

complexes, is highly conserved. The loss/gain due to the I84V substitution in the ACT complexes was also assessed. Like the V82A mutation in the FLAP+ complexes, the I84V mutation in the ACT complexes results in a reduction in van der Waals contact by ~0.5 although the absolute values vary. The V82T mutation in the ACT complexes results in minor changes in van der Waals contacts. Overall, the impact of the mutations appears not to be a direct change in van der Waals contact but rather an indirect change subtly rearranging the active site and therefore the energetics of the inhibitor binding.

Comparison of ΔC_p and change in Accessible Surface Area on binding inhibitor :

Heat capacity changes for the binding of APV and DRV were computed (Table AII-2). The ΔC_p for both inhibitors binding to FLAP+ was more negative relative to WT. FLAP+ exhibits a $\Delta\Delta C_{pAPV} = -105$ cal·K^{-1}mol^{-1} and $\Delta\Delta C_{pDRV} = -134$ cal·K^{-1}mol^{-1}. Negative ΔC_p values generally correlate with a loss in accessible surface area (Ladbury and Chowdhry, 1998), as well as a more favorable entropy change as was observed here (Table AII-2). Changes in accessible surface area for APV and DRV were computed using the crystal structures. In relation to the WT structures, the FLAP+ complexes of APV and DRV have a change in accessible surface area of -36 Å^2 and -27 Å^2, respectively. These values are not large enough to explain the considerable differences in ΔC_p values found experimentally. Therefore, the large compensatory entropic changes found for APV and DRV binding to the FLAP+ mutant are likely not limited to changes in accessible surface area but likely arise from other sources such as subtle structural rearrangement in the flaps or changes in the dynamics of the flaps in the free form of the enzyme.

Crystallographic waters and entropy-enthalpy compensation:

One possible explanation for entropy-enthalpy compensation involves desolvation. The loss of ordered water molecules upon binding an inhibitor gives rise to a favorable change in entropy, however this is compensated by a loss in enthalpy as specific hydrogen bonds with water molecules are broken. The magnitudes of the differences in the entropy and enthalpy changes associated with the binding of APV and DRV to FLAP+ when compared to WT are greater than those observed for the binding of the other inhibitors, and thus are of particular interest. The significant differences in the enthalpy changes observed for APV and DRV binding to FLAP+ relative to WT ($\Delta\Delta H_{APV}$ =10.7 $kcal \cdot mol^{-1}$and 14.1 $kcal \cdot mol^{-1}$) (Table AII-2 and Figure AII-2), may be associated with the loss of bound water molecules. The approximate energy of a hydrogen bond (5 $kcal \cdot mol^{-1}$) (Dunitz, 1995), could potentially account for the release of two to four water molecules for APV and DRV binding to FLAP+ respectively. In an attempt to search for a structural correlation for the observed changes in enthalpy, the crystallographic waters within a 4.2Å hydration shell around the protein were analyzed. Detailed comparisons were conducted to elucidate possible changes in water structure between the FLAP+ and the WT or ACT complexes, with particular focus on the APV and DRV complexes. However, no systematic change in the water structure of the FLAP+ complexes was observed. Nevertheless these large changes in entropy and enthalpy could still be due to differential solvation of the free state of the FLAP+ enzyme.

DISCUSSION

A major challenge during the treatment of HIV infections is the emergence of drug-resistant viruses. Substitutions in the protease that arise as a result of these mutations lead to a reduction in the affinity of the inhibitor, thereby maintaining the viral load. The majority of thermodynamic studies involving drug-resistant variants of HIV-1 protease have revealed a loss in the enthalpic contribution resulting in the loss of overall free energy of binding (King et al., 2004c; Lafont et al., 2007; Muzammil, Ross, and Freire, 2003; Ohtaka and Freire, 2005; Ohtaka, Schon, and Freire, 2003; Ohtaka et al., 2002; Stoica, Sadiq, and Coveney, 2008). In the current study, a variant of HIV-1 protease, FLAP+ (L10I, G48V, I54V, L63P, and V82A) is characterized that exhibits an extremely large entropy-enthalpy compensation of a magnitude not previously reported for any system. This compensation is observed to various extents with six different HIV-1 protease inhibitors, and most dramatic for APV and DRV. In these cases, the loss in the enthalpic interactions are 11 and 14 kcal·mol^{-1} for APV and DRV, respectively, and are compensated by large similar magnitude favorable entropy. These compensating effects resulted in low overall losses in binding affinity.

This extreme entropy-enthalpy compensation is a function of the physical characteristics of the FLAP+ variant of HIV-1 protease. This variant has a combination of mutations in the flaps at G48V, I54V, an active site mutation V82A and a surface mutation just outside the active site L10I. Some conserved changes in the overall structure are observed between the FLAP+ versus either WT or ACT variants. Changes in solvent accessibility (Ladbury and Chowdhry, 1998), hydrogen bonding, van der

Waals interactions and water structure (Dunitz, 1995), all which have been previously implicated as potential reasons for entropy-enthalpy compensation, were analyzed. However, particularly for the changes observed in the APV and DRV complexes, these properties of the inhibited forms of the enzyme do not fully explain the observed large changes in enthalpy, entropy and heat capacity.

The changes in thermodynamics are likely caused by highly interdependent alterations in the enzyme, involving both the inhibited enzyme and changes in the dynamics and hydrophilicity of the free state of the FLAP+ HIV-1 protease. This seems likely as two of the mutations are in the flexible flap regions, G48V and I54V. Molecular dynamics simulations comparing the WT protease with a G48V mutant indicated a marked difference in the flexibility of the flap tips (Hamelberg and McCammon, 2005), reducing the frequency of *trans-cis* isomerization of the ω-bond for Val48 relative to Gly48. To completely integrate thermodynamics in structure based drug design, a comprehensive approach is necessary involving structures and dynamic information of both the free and the bound states of the inhibitor and the therapeutic target. The binding thermodynamic properties can be profoundly modulated not only by the inhibitor, but also by alteration of the target, such as in the evolution of drug resistance. The challenge to rational drug design is to truly integrate interdependent sequence, structure, energetic and dynamic data in a productive manner.

ACKNOWLEDGEMENTS

The authors thank Deyna Cooper, Christina Ng, Vincent Chou, Colleen McRell, Sagar Kathuria, Madhavi Kolli for help with experiments, Yufeng Cai for the force field, and Balaji Bhyravbhatla and Luca Leone for technical assistance. The following reagents were obtained through the NIH AIDS Research and Reference Reagent Program, Division of AIDS, NIAID, NIH: amprenavir, atazanavir, indinavir, nelfinavir and saquinavir. Darunavir was kindly provided by Tibotec, Inc. This work has been funded by National Institutes of Health (P01-GM66524).

APPENDIX III

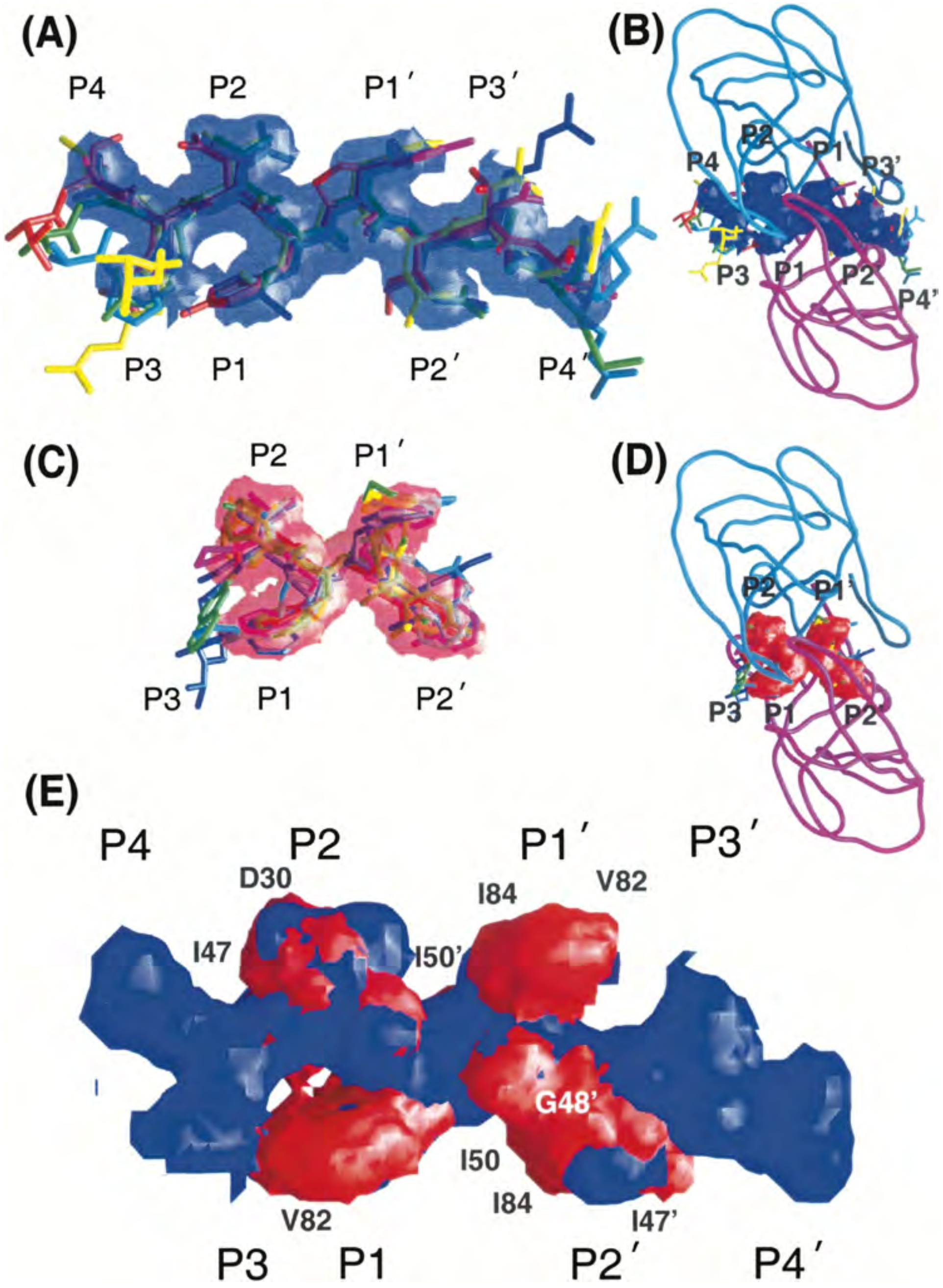
Binding Thermodynamics Profiles of Inhibitors that Bind Within the Substrate Envelope

INTRODUCTION

The homo-dimeric aspartyl protease of HIV-1 recognizes and cleaves at least ten non-homologous and asymmetric substrates in the Gag and Gag-Pro-Pol polyproteins and is essential for the formation of infectious viral particles (Chou, 1996a; Henderson et al., 1988; Pettit et al., 1998). Given the vital role played by the protease in the viral life cycle, HIV-1 protease has been a key target in the development of anti HIV-1 therapies in an effort to slow down the progression of the infection to AIDS. At present, there are nine FDA approved inhibitors (PIs) that are successfully being used as a part of highly active anti-retroviral therapy (HAART). However, the accumulation of inhibitor resistance mutations within the protease gene can significantly reduce the efficacy of the inhibitors.

Based on crystallographic data on HIV-1 protease in complex with substrate peptides, it has been established that HIV-1 protease recognizes a specific asymmetric shape the substrates conform to rather than a specific sequence (Prabu-Jeyabalan, Nalivaika, and Schiffer, 2002). The specific shape taken up by the substrates has been defined as the substrate envelope (Figure AIII-1A and B). Similarly, using crystal structures of HIV-1 protease in complex with PIs, an inhibitor envelope defining the shape taken up by the bound inhibitors has been generated (Figure AIII-1C and D). By superposing the inhibitor envelope on to the substrate envelope, it was observed that a significant number of inhibitor resistance substitutions within the protease active site occurred at positions where the inhibitor envelope protruded beyond the substrate

FIGURE AIII-1


(A)
P4
P2
P1′
P3′
P3
P1
P2′
P4′
(B)
P4
P2
P1
P3'
P3
P1
P2'
P4'
(C)
P2
P1′
P3
P1
P2′
(D)
P2
P1'
P3
P1
P2'
(E)
P4
P2
P1′
P3′
D30
I84
V82
I47
I50'
G48'
I50
I84
V82
I47'
P3
P1
P2′
P4′

Reprinted from Chemistry & Biology, Vol. 11, *"Combating Susceptibility to Drug Resistance: Lessons from HIV-1 Protease"*, Pages 1333-8, © 2004, with permission from Elsevier.

FIGURE AIII-1. Substrate and Inhibitor Envelopes of HIV-1 Protease. **A.** The substrate envelope calculated with GRASP (Nicholls, Sharp, and Honig, 1991b)from the overlapping van der Waals volume of four or more substrate peptides. The colors of the substrate peptides are red, matrix-capsid; green, capsid-p2; blue, p2-nucleocapsid; cyan, p1-p6; magenta, reverse-transcriptase-ribonucleaseH; and yellow, ribonucleaseH-integrase. **B.** The substrate envelope as it fits within the active site of HIV-1 protease. The α-carbon trace is of the CA-p2 substrate peptide complex (Prabu-Jeyabalan, Nalivaika, and Schiffer, 2000b). **C.** The inhibitor envelope calculated from overlapping van der Waals volume of five or more of eight inhibitor complexes. The colors of the inhibitors are yellow, Nelfinavir (NFV); gray, Saquinavir (SQV); cyan, Indinavir (IDV); light blue, Ritonavir (RTV); green, Amprenavir (APV); magenta, Lopinavir (LPV); blue, Atazanavir (ATV); and red, darunavir (DRV). **D.** The inhibitor envelope as it fits within the active site of HIV-1 protease. **E.** Superposition of the substrate envelope (blue) with the inhibitor envelope (red). Residues that contact the inhibitors where the inhibitors protrude beyond the substrate envelope and confer drug resistance when they mutate are labeled.

envelope to interact with the protein (Figure AIII-1E) (King et al., 2004a). Thus, if inhibitors can be designed to fit within the substrate, amino acid substitutions that accumulate within the active site will no longer be able to influence inhibitor binding.

To test the hypothesis that the protease will not be able to develop resistant mutations in response to inhibitors that fit within substrate envelope, large libraries of protease inhibitors were designed using computational approaches. A major focus in the design effort has been with using the *(R)-* (hydroxyethylamino) sulfonamide isostere as a scaffold which is used in two potent FDA approved protease inhibitors, amprenavir (APV) and darunavir (DRV) (Figure AIII-2A). This has resulted in several APV and DRV analogs that have now been synthesized and characterized using structural and biochemical methods. Crystal structures of protease variants in complex with many of these inhibitors indicate that they fit well within the substrate envelope (Nalam et. al., unpublished data). Data from enzyme and viral inhibition assays have revealed that several of these inhibitors were able to maintain exceptional potency against a number of inhibitor resistant enzyme variants and viral strains (unpublished data). Data presented in this section are a part of an ongoing effort to characterize the thermodynamic binding parameters of these newly synthesized APV and DRV analog inhibitors with a number of drug resistant protease variants.

FIGURE AIII-2

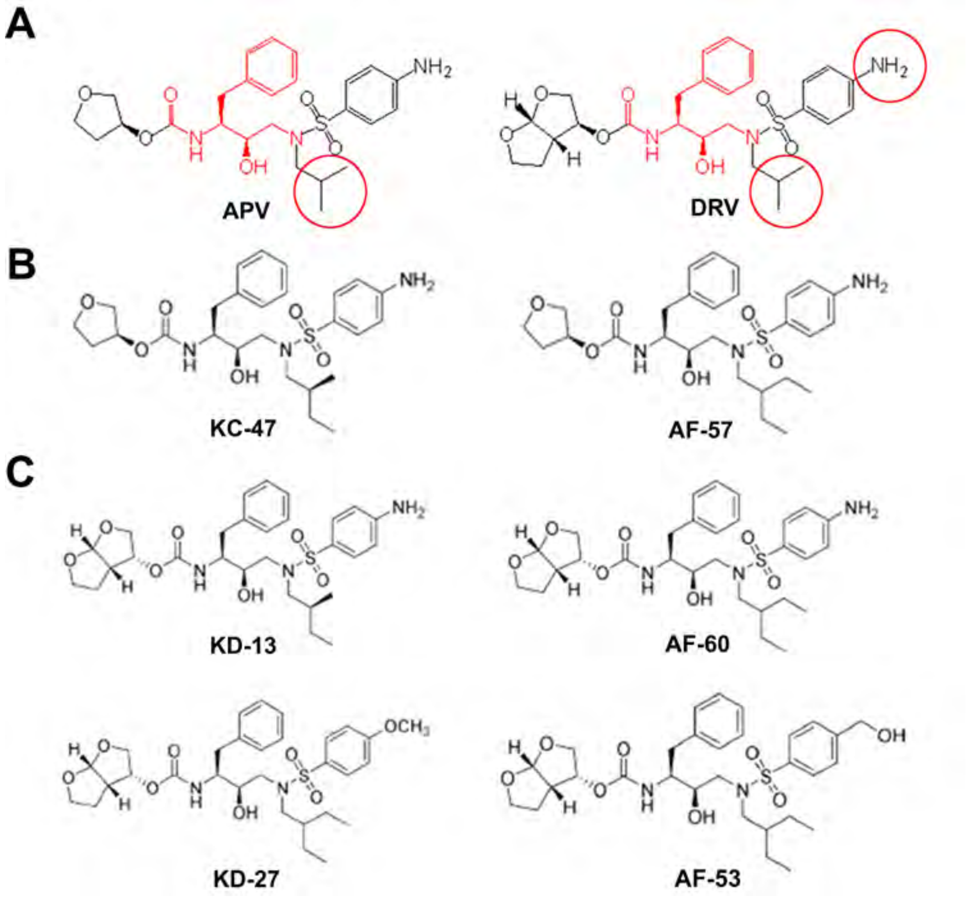

A
APV
DRV
B
KC-47
AF-57
C
KD-13
AF-60
KD-27
AF-53

FIGURE AIII-2. A. Chemical structures of APV and DRV with the scaffold colored in red. Red circles indicate chemical moieties that were changed to generate the APV and DRV analogs. **B.** Chemical structures of APV analogs. **C.** Chemical structures of DRV analogs.

MATERIALS AND METHODS

Protease gene construction, protein expression and purification and displacement isothermal titration calorimetry (ITC) experiments for wild type, FLAP+ (L10I, G48V, I54V, V82A), I50V, D30N/N88D and M3 (L10I, A71V, G73S, I84V, L90M) protease variants were performed as described in previous chapters.

RESULTS AND FUTURE DIRECTIONS

Thermodynamic binding parameters for APV and DRV analogs determined up to now with protease variants are summarized in Table AIII-1 and AIII-2. Both APV analogs had improved affinity for WT protease. The binding affinities of KC-47 to most variants were not significantly different from that of APV. Currently, ITC experiments are being performed to determine binding thermodynamic parameters for AF-57 with the inhibitor resistant protease variant panel.

The affinities of DRV analogs binding to WT protease appeared to be significantly lower than the DRV affinity for WT protease. However, K_i data determined for all the DRV analogs indicated that they bound to WT protease with a greater affinity (sub-pico molar or better) than DRV (Nalam *et al.*, unpublished data). Furthermore, IC_{50} values determined for the DRV analogs against a panel of viruses indicated that the inhibitory concentrations were much lower than that for DRV. Together, the K_i and IC_{50} data suggest that the DRV analogs are more effective that DRV.

TABLE AIII-1 Binding thermodynamic parameters for APV, KC47 and AF57.

Protease Variant	K_a (M^{-1})	K_d (nM)	K_d ratio	ΔH (kcal mol^{-1})	ΔΔH	–TΔS (kcal mol^{-1})	Δ (–TΔS)	ΔG (kcal mol^{-1})	ΔΔG
				APV					
WT	(2.6 ± 1.3) × 10^9	0.39 ± 0.20	1.0	–7.3 ± 0.9	-	–5.3	-	–12.6 ± 0.3	-
FLAP+	(7.62 ± 0.06) × 10^8	1.3 ± 0.01	3.3	3.3 ± 0.5	10.6	–15.2	–9.9	–11.9 ± 0.3	0.7
I50V, A71V	(1.14 ± 0.03) × 10^8	0.88 ± 0.02	2.2	–11.8 ± 0.3	–4.5	–0.3	5.0	–12.1 ± 0.01	0.5
D30N, N88D	(1.19 ± 0.18) × 10^{10}	0.08 ± 0.01	0.2	–10.2 ± 1.5	–2.9	–3.3	2.0	–13.5 ± 0.09	–0.9
M3	(4.99 ± 0.05) × 10^8	2.0 ± 0.02	5.1	–7.2 ± 0.04	–0.1	–11.7	–6.4	–11.7 ± 0.006	0.9
				KC-47					
WT	(5.4 ± 2.7) × 10^9	0.18 ± 0.09	1.0	–6.1 ± 0.5	-	–7.0	-	–13.1 ± 0.2	-
FLAP+	(1.2 ± 0.4) × 10^9	0.84 ± 0.26	4.6	2.7 ± 0.3	8.8	−14.8	–7.8	–12.2 ± 0.2	0.9
I50V, A71V	(9.6 ± 1.3) × 10^8	1.04 ± 0.15	5.7	–8.4 ± 0.2	–2.3	–3.6	3.4	–12.0 ± 0.2	1.1
D30N, N88D	(5.8 ± 0.3) × 10^8	0.173 ± 0.009	1.0	–8.2 ± 0.6	–2.1	–3.6	3.4	–13.1 ± 3.0	0.0
M3	(3.2 ± 0.5) × 10^8	3.1 ± 0.46	17.2	–9.3 ± 0.9	–3.2	–2.1	4.9	–11.4 ± 0.9	–1.7
				AF-57					
WT	(3.9 ± 1.2) × 10^9	0.26 ± 0.07	1.0	–5.9 ± 1.3	-	–7.02	-	–12.9 ± 0.2	-

TABLE AIII-2 Binding thermodynamic parameters DRV analogs binding to wild type protease.

Protease Variant	K_a (M^{-1})	K_d (pM)	ΔH (kcal mol^{-1})	–TΔS (kcal mol^{-1})	ΔG (kcal mol^{-1})
WT	$(2.6 \pm 1.3) \times 10^{9}$	4.0 ± 2.0	–12.1 ± 0.9	–3.1	–15.2 ± 0.3
WT	$(1.18 \pm 1.56) \times 10^{12}$	85.0 ± 113.0	–7.3 ± 0.5	–6.2	–13.5 ± 0.3
WT	$(9.6 \pm 0.96) \times 10^{9}$	104.0 ± 11.0	–8.5 ± 0.8	–4.9	–13.4 ± 5.9
WT	$(1.2 \pm 0.3) \times 10^{10}$	83.0 ± 20.0	–5.7 ± 0.5	–7.9	–13.5 ± 0.2
WT	$(1.8 \pm 0.2) \times 10^{10}$	53.2 ± 4.6	–11.4 ± 1.3	–2.4	–13.8 ± 5.3

One explanation the weaker affinities observed in the ITC experiments could be the poor solubility of these analogs at the concentrations needed to carry out ITC experiments. While problems related to solubility are not obvious when carrying out the ITC experiments, it is possible that the inhibitors precipitate out of the inhibitor solutions to a certain level during the course of the experiment and thereby change the effective inhibitor concentration of the experiment. Another possible explanation could be the choice of the weak inhibitor used for the displacement experiments. Acetyl-pepstatin, which was used to initially saturate the protein, has a binding affinity for protease in the micro molar range. The protease inhibitory assay data show that the affinities of the DRV analogs are in the sub-pico molar range or better. Thus, given that the DRV analogs bind to protease about six orders of magnitude or better than acetyl-pepstatin, it might not be possible to accurately determine the affinity constants by using acetyl-pepstatin as the weak binding inhibitor for the displacement experiments. At present, experiments are being carried out using nelfinavir (NFV) as the weak binding inhibitor for the displacement experiments. The affinity of NFV to protease, which in the sub nano molar range, will be about three orders of magnitude tighter than acetyp-pepstatin and should not allow the DRV analogs to displace it too fast. This slow displacement should allow for the accurate determination of affinity constants for the DRV analogs.

Overall, determining the binding thermodynamic parameters for the APV and DRV analogs will provide insights into the how thermodynamics of binding are affected by modifying chemical groups within PIs that fit well within the substrate envelope.

Thus, data gathered from these studies will aid in the design of inhibitors that have improved binding thermodynamics while fitting well within the substrate envelope.

REFERENCES

Adachi, M., Ohhara, T., Kurihara, K., Tamada, T., Honjo, E., Okazaki, N., Arai, S., Shoyama, Y., Kimura, K., Matsumura, H., Sugiyama, S., Adachi, H., Takano, K., Mori, Y., Hidaka, K., Kimura, T., Hayashi, Y., Kiso, Y., and Kuroki, R. (2009). Structure of HIV-1 protease in complex with potent inhibitor KNI-272 determined by high-resolution X-ray and neutron crystallography. *Proc Natl Acad Sci U S A* **106**(12)**,** 4641-6.

Alimonti, J. B., Ball, T. B., and Fowke, K. R. (2003). Mechanisms of CD4+ T lymphocyte cell death in human immunodeficiency virus infection and AIDS. *Journal of General Virology* **84**(Pt 7)**,** 1649-61.

Ariyoshi, K., Matsuda, M., Miura, H., Tateishi, S., Yamada, K., and Sugiura, W. (2003a). Patterns of point mutations associated with antiretroviral drug treatment failure in CRF01_AE (subtype E) infection differ from subtype B infection. *J Acquir Immune Defic Syndr* **33**(3)**,** 336-42.

Ariyoshi, K., Matsuda, M., Miura, H., Tateishi, S., Yamada, K., and Sugiura, W. (2003b). Patterns of point mutations associated with antiretroviral drug treatment failure in CRF01_AE (subtype E) infection differ from subtype B infection. *Journal of Acquired Immune Deficiency Syndromes* **33**(3)**,** 336-42.

Asamoah-Odei, E., Garcia Calleja, J. M., and Boerma, J. T. (2004). HIV prevalence and trends in sub-Saharan Africa: no decline and large subregional differences. *Lancet* **364**(9428)**,** 35-40.

Barre-Sinoussi, F., Chermann, J. C., Rey, F., Nugeyre, M. T., Chamaret, S., Gruest, J., Dauguet, C., Axler-Blin, C., Vezinet-Brun, F., Rouzioux, C., Rozenbaum, W., and Montagnier, L. (1983). Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science* **220**(4599)**,** 868-71.

Bebenek, K., Abbotts, J., Wilson, S. H., and Kunkel, T. A. (1993). Error-prone polymerization by HIV-1 reverse transcriptase. Contribution of template-primer misalignment, miscoding, and termination probability to mutational hot spots. *Journal of Biological Chemistry* **268**(14)**,** 10324-34.

Becker-Pergola, G., Kataaha, P., Johnston-Dow, L., Fung, S., Jackson, J. B., and Eshleman, S. H. (2000). Analysis of HIV type 1 protease and reverse transcriptase in antiretroviral drug-naive Ugandan adults. *AIDS Res Hum Retroviruses* **16**(8)**,** 807-13.

Berg, J. M., Tymoczko, J. L., and Stryer, L. (2007). Biochemistry. 6 ed. *In* "Biochemistry", pp. 252. Tenney, S.

Blasie, C. A., and Berg, J. M. (2004). Entropy-enthalpy compensation in ionic interactions probed in a zinc finger peptide. *Biochemistry* **43**(32)**,** 10600-10604.

Bonfanti, P., Capetti, A., and Rizzardini, G. (1999). HIV disease treatment in the era of HAART. *Biomedicine and Pharmacotherapy* **53**(2)**,** 93-105.

Bottcher, J., Blum, A., Heine, A., Diederich, W. E., and Klebe, G. (2008). Structural and Kinetic Analysis of Pyrrolidine-Based Inhibitors of the Drug-Resistant Ile84Val Mutant of HIV-1 Protease. *J Mol Biol***,** (in press) doi:10.1016/j.jmb.2008.07.062

Brik, A., and Wong, C. H. (2003). HIV-1 protease: mechanism and drug discovery. *Org Biomol Chem* **1**(1)**,** 5-14.

Buonaguro, L., Tornesello, M. L., and Buonaguro, F. M. (2007). Human immunodeficiency virus type 1 subtype distribution in the worldwide epidemic: pathogenetic and therapeutic implications. *Journal of Virology* **81**(19)**,** 10209-19.

Carr, J. K., Salminen, M. O., Koch, C., Gotte, D., Artenstein, A. W., Hegerich, P. A., St Louis, D., Burke, D. S., and McCutchan, F. E. (1996). Full-length sequence and mosaic structure of a human immunodeficiency virus type 1 isolate from Thailand. *J Virol* **70**(9)**,** 5935-43.

Chan, D. C., and Kim, P. S. (1998). HIV entry and its inhibition. *Cell* **93**(5)**,** 681-4.

Chou, K. (1996a). Prediction of human immunodeficiency virus protease cleavage sites in proteins. *Anal Biochem* **233,** 1-14.

Chou, K. (1996b). Prediction of human immunodeficiency virus protease cleavage sites in proteins. *Analytical Biochemistry* **233,** 1-14.

Clemente, J. C., Coman, R. M., Thiaville, M. M., Janka, L. K., Jeung, J. A., Nukoolkarn, S., Govindasamy, L., Agbandje-McKenna, M., McKenna, R., Leelamanit, W., Goodenow, M. M., and Dunn, B. M. (2006). Analysis of HIV-1 CRF_01 A/E protease inhibitor resistance: structural determinants for maintaining sensitivity and developing resistance to atazanavir. *Biochemistry* **45**(17)**,** 5468-77.

Collaborative Computational Project Number 4 (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* **50,** 760-763.

Collaborative-Computational-Project, N. (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* **50,** 760-763.

Colonno, R., Parkin, N., McLaren, C., Seekins, D., Hodder, S., Schnittman, S., and Kelleher, T. (2004a). Pathways to Atazanavir Resistance in Treatment-experienced Patients and Impact of Residue 50 Substitutions. *In* "Conference on Retroviruses and Opportunistic Infections", San Francisco, California.

Colonno, R., Rose, R., McLaren, C., Thiry, A., Parkin, N., and Friborg, J. (2004b). Identification of I50L as the signature atazanavir (ATV)-resistance mutation in treatment-naive HIV-1-infected patients receiving ATV-containing regimens. *Journal of Infectious Diseases* **189**(10)**,** 1802-10.

Coman, R. M., Robbins, A., Goodenow, M. M., McKenna, R., and Dunn, B. M. (2007). Expression, purification and preliminary X-ray crystallographic studies of the human immunodeficiency virus 1 subtype C protease. *Acta Crystallograph Sect F Struct Biol Cryst Commun* **63**(Pt 4)**,** 320-3.

Condra, J. H., Schleif, W. A., Blahy, O. M., Gabryelski, L. J., Graham, D. J., Quintero, J. C., Rhodes, A., Robbins, H. L., Roth, E., Shivaprakash, M., Titus, D., Yang, T., Teppler, H., Squires, K. E., Deutsch, P. J., and Emini, E. (1995). *In vivo* emergence of HIV-1 variants resistant to multiple protease inhibitors. *Nature* **374,** 569-571.

Copeland, R. A. (1996). "Enzymes: A practical introduction to structure, mechanism, and data analysis." Wiley-VCH.

Davidson, J. P., Lubman, O., Rose, T., Waksman, G., and Martin, S. F. (2002). Calorimetric and Structural Studies of 1,2,3-Trisubstituted Cyclopropanes as

Conformationally Constrained Peptide Inhibitors of Src SH2 Domain Binding. *J. Am. Chem. Soc.* **124**(2)**,** 205-215.

De Clercq, E. (2009). The history of antiretrovirals: key discoveries over the past 25 years. *Rev Med Virol* **19**(5)**,** 287-99.

De Meyer, S., Azijn, H., Surleraux, D., Jochmans, D., Tahri, A., Pauwels, R., Wigerinck, P., and de Bethune, M. P. (2005). TMC114, a novel human immunodeficiency virus type 1 protease inhibitor active against protease inhibitor-resistant viruses, including a broad range of clinical isolates. *Antimicrobial Agents and Chemotherapy* **49**(6)**,** 2314-21.

de Oliveira, T., Engelbrecht, S., Janse van Rensburg, E., Gordon, M., Bishop, K., zur Megede, J., Barnett, S. W., and Cassol, S. (2003). Variability at human immunodeficiency virus type 1 subtype C protease cleavage sites: an indication of viral fitness? *Journal of Virology* **77**(17)**,** 9422-30.

Debouck, C. (1992). The HIV-1 protease as a therapeutic target for AIDS. *AIDS Res Hum Retroviruses* **8**(2)**,** 153-164.

Debouck, C., and Metcalf, B. W. (1990). Human Immunodeficiency Virus Protease: A target for AIDS therapy. *Drug Develop Res* **21,** 1-17.

The PyMol Molecular Graphics System. San Carlos, CA, USA.: DeLano Scientific.

Dimitrov, D. S. (1997). How do viruses enter cells? The HIV coreceptors teach us a lesson of complexity. *Cell* **91**(6)**,** 721-30.

Dorsey, B. D., Levin, R. B., McDaniel, S. L., Vacca, J. P., Guare, J. P., Darke, P. L., Zugay, J. A., Emini, E. A., Schleif, W. A., Quintero, J. C., and et al. (1994). L-735,524: the design of a potent and orally bioavailable HIV protease inhibitor. *Journal of Medicinal Chemistry* **37**(21)**,** 3443-51.

Dunitz, J. D. (1995). Win some, lose some: enthalpy-entropy compensation in weak intermolecular interactions. *Chem Biol* **2**(11)**,** 709-712.

Elston, R., Scherer, J., Schapiro, J., Bethell, R., Kohlbrenner, V., and Mayers, D. (2006). De-Selection fo the I50V Mutation Occurs in Clinical Isolates During Aptivus/r (Tipranavir/Ritonavir) Based Therapy. *In* "15th International HIV Drug Resitance Workshop", Sitges, Spain.

Emini, E. A., Schleif, W. A., Deutsch, P., and Condra, J. H. (1996). In vivo selection of HIV-1 variants with reduced susceptibility to the protease inhibitor L-735,524 and related compounds. *Adv Exp Med Biol* **394,** 327-331.

Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**(Pt 12 Pt 1)**,** 2126-32.

Essex, M. (1996). Retroviral vaccines: challenges for the developing world. *AIDS Research and Human Retroviruses* **12**(5)**,** 361-3.

Fackler, O. T., and Peterlin, B. M. (2000). Endocytic entry of HIV-1. *Current Biology* **10**(16)**,** 1005-8.

Flexner, C. (2007). HIV drug development: the next 25 years. *Nat Rev Drug Discov* **6**(12)**,** 959-66.

Foulkes-Murzycki, J. E., Scott, W. R., and Schiffer, C. A. (2007). Hydrophobic sliding: a possible mechanism for drug resistance in human immunodeficiency virus type 1 protease. *Structure* **15**(2)**,** 225-33.

Gao, F., Bailes, E., Robertson, D. L., Chen, Y., Rodenburg, C. M., Michael, S. F., Cummins, L. B., Arthur, L. O., Peeters, M., Shaw, G. M., Sharp, P. M., and Hahn, B. H. (1999). Origin of HIV-1 in the chimpanzee Pan troglodytes troglodytes. *Nature* **397**(6718)**,** 436-41.

Gao, F., Robertson, D. L., Morrison, S. G., Hui, H., Craig, S., Decker, J., Fultz, P. N., Girard, M., Shaw, G. M., Hahn, B. H., and Sharp, P. M. (1996). The heterosexual human immunodeficiency virus type 1 epidemic in Thailand is caused by an intersubtype (A/E) recombinant of African origin. *J Virol* **70**(10)**,** 7013-29.

Gomes, P., Diogo, I., Goncalves, M. F., Carvalho, P., Cabanas, J., Lobo, M. C., and Camacho, R. (2002). *Conference on Retroviruses and Opportunistic Infections (CROI), Seatle, WA.*

Gonzales, M. J., Machekano, R. N., and Shafer, R. W. (2001). Human immunodeficiency virus type 1 reverse-transcriptase and protease subtypes: classification, amino acid mutation patterns, and prevalence in a northern California clinic-based population. *J Infect Dis* **184**(8)**,** 998-1006.

Gregoret, L. M., Rader, S. D., Fletterick, R. J., and Cohen, F. E. (1991). Hydrogen bonds involving sulfur atoms in proteins. *Proteins* **9**(2)**,** 99-107.

Grossman, Z., Paxinos, E. E., Averbuch, D., Maayan, S., Parkin, N. T., Engelhard, D., Lorber, M., Istomin, V., Shaked, Y., Mendelson, E., Ram, D., Petropoulos, C. J., and Schapiro, J. M. (2004a). Mutation D30N is not preferentially selected by human immunodeficiency virus type 1 subtype C in the development of resistance to nelfinavir. *Antimicrob Agents Chemother* **48**(6)**,** 2159-65.

Grossman, Z., Paxinos, E. E., Averbuch, D., Maayan, S., Parkin, N. T., Engelhard, D., Lorber, M., Istomin, V., Shaked, Y., Mendelson, E., Ram, D., Petropoulos, C. J., and Schapiro, J. M. (2004b). Mutation D30N is not preferentially selected by human immunodeficiency virus type 1 subtype C in the development of resistance to nelfinavir. *Antimicrobial Agents and Chemotherapy* **48**(6)**,** 2159-65.

Gulnik, S. V., Suvorov, L. I., Liu, B., Yu, B., Anderson, B., Mitsuya, H., and Erickson, J. W. (1995). Kinetic characterization and cross-resistance patterns of HIV-1 protease mutants selected under drug pressure. *Biochemistry* **34,** 9282-9287.

Hamelberg, D., and McCammon, J. A. (2005). Fast Peptidyl cis-trans Isomerization within the Flexible Gly-Rich Flaps of HIV-1 Protease. *J. Am. Chem. Soc.* **127**(40)**,** 13778-13779.

Hemelaar, J., Gouws, E., Ghys, P. D., and Osmanov, S. (2006). Global and regional distribution of HIV-1 genetic subtypes and recombinants in 2004. *AIDS* **20**(16)**,** W13-23.

Henderson, L. E., Copeland, T. D., Sowder, R. C., Schultz, A. M., and Oraszlan, S. (1988). "Human retroviruses, cancer and AIDS: Approaches to prevention and therapy." Liss, New York.

Hirsch, V. M., Olmsted, R. A., Murphey-Corb, M., Purcell, R. H., and Johnson, P. R. (1989). An African primate lentivirus (SIVsm) closely related to HIV-2. *Nature* **339**(6223)**,** 389-92.

Holguin, A., Ramirez de Arellano, E., Rivas, P., and Soriano, V. (2006). Efficacy of antiretroviral therapy in individuals infected with HIV-1 non-B subtypes. *AIDS Rev* **8**(2)**,** 98-107.

Houk, K. N., Leach, A.G, Kim, S. P, Zhang, X (2003). Binding Affinities of Host-Guest, Protein-Ligand, and Protein-Transition-State Complexes. *Angewandte Chemie International Edition* **42**(40)**,** 4872-4897.

Hu, D. J., Buve, A., Baggs, J., van der Groen, G., and Dondero, T. J. (1999). What role does HIV-1 subtype play in transmission and pathogenesis? An epidemiological perspective. *AIDS* **13**(8)**,** 873-81.

Huet, T., Cheynier, R., Meyerhans, A., Roelants, G., and Wain-Hobson, S. (1990). Genetic organization of a chimpanzee lentivirus related to HIV-1. *Nature* **345**(6273)**,** 356-9.

Hui, J. O., A.G., T., Reardon, I. M., Lull, J. M., Brunner, D. P., Tomich, C. C., and Heinrikson, R. L. (1993). Large scale purification and refolding of HIV-1 protease from *Escherichia coli* inclusion bodies. *J. Prot. Chem.* **12,** 323-327.

Hyland, L. J., Tomaszek, T. A., Jr., and Meek, T. D. (1991). Human immunodeficiency virus-1 protease. 2. Use of pH rate studies and solvent kinetic isotope effects to elucidate details of chemical mechanism. *Biochemistry* **30**(34)**,** 8454-63.

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., and et al. (1991). Human immunodeficiency virus-1 protease. 1. Initial velocity studies and kinetic characterization of reaction intermediates by 18O isotope exchange. *Biochemistry* **30**(34)**,** 8441-53.

James, M. N., Sielecki, A. R., Hayakawa, K., and Gelb, M. H. (1992). Crystallographic analysis of transition state mimics bound to penicillopepsin: difluorostatine- and difluorostatone-containing peptides. *Biochemistry* **31**(15)**,** 3872-86.

Jones, T. A., Bergdoll, M., and Kjeldgaard, M. (1990). O: A macromolecular modeling environment. *In* "Crystallographic and Modeling Methods in Molecular Design" (C. Bugg, and S. Ealick, Eds.), pp. 189-195. Springer-Verlag Press, Berlin.

Julg, B., and Goebel, F. D. (2005). HIV genetic diversity: any implications for drug resistance? *Infection* **33**(4)**,** 299-301.

Kaldor, S. W., Kalish, V. J., Davies, J. F., 2nd, Shetty, B. V., Fritz, J. E., Appelt, K., Burgess, J. A., Campanale, K. M., Chirgadze, N. Y., Clawson, D. K., Dressman, B. A., Hatch, S. D., Khalil, D. A., Kosa, M. B., Lubbehusen, P. P., Muesing, M. A., Patick, A. K., Reich, S. H., Su, K. S., and Tatlock, J. H. (1997). Viracept (nelfinavir mesylate, AG1343): a potent, orally bioavailable inhibitor of HIV-1 protease. *Journal of Medicinal Chemistry* **40**(24)**,** 3979-85.

Kaleebu, P., French, N., Mahe, C., Yirrell, D., Watera, C., Lyagoba, F., Nakiyingi, J., Rutebemberwa, A., Morgan, D., Weber, J., Gilks, C., and Whitworth, J. (2002a). Effect of human immunodeficiency virus (HIV) type 1 envelope subtypes A and D on disease progression in a large cohort of HIV-1-positive persons in Uganda. *Journal of Infectious Diseases* **185**(9)**,** 1244-50.

Kaleebu, P., French, N., Mahe, C., Yirrell, D., Watera, C., Lyagoba, F., Nakiyingi, J., Rutebemberwa, A., Morgan, D., Weber, J., Gilks, C., and Whitworth, J. (2002b).

Effect of human immunodeficiency virus (HIV) type 1 envelope subtypes A and D on disease progression in a large cohort of HIV-1-positive persons in Uganda. *J Infect Dis* **185**(9)**,** 1244-50.

Kanki, P. J., Hamel, D. J., Sankale, J. L., Hsieh, C., Thior, I., Barin, F., Woodcock, S. A., Gueye-Ndiaye, A., Zhang, E., Montano, M., Siby, T., Marlink, R., I, N. D., Essex, M. E., and S, M. B. (1999a). Human immunodeficiency virus type 1 subtypes differ in disease progression. *Journal of Infectious Diseases* **179**(1)**,** 68-73.

Kanki, P. J., Hamel, D. J., Sankale, J. L., Hsieh, C., Thior, I., Barin, F., Woodcock, S. A., Gueye-Ndiaye, A., Zhang, E., Montano, M., Siby, T., Marlink, R., I, N. D., Essex, M. E., and S, M. B. (1999b). Human immunodeficiency virus type 1 subtypes differ in disease progression. *J Infect Dis* **179**(1)**,** 68-73.

Kantor, R., and Katzenstein, D. (2003). Polymorphism in HIV-1 non-subtype B protease and reverse transcriptase and its potential impact on drug susceptibility and drug resistance evolution. *AIDS Rev* **5**(1)**,** 25-35.

Kantor, R., and Katzenstein, D. (2004). Drug resistance in non-subtype B HIV-1. *Journal of Clinical Virology* **29**(3)**,** 152-9.

Kantor, R., Katzenstein, D. A., Efron, B., Carvalho, A. P., Wynhoven, B., Cane, P., Clarke, J., Sirivichayakul, S., Soares, M. A., Snoeck, J., Pillay, C., Rudich, H., Rodrigues, R., Holguin, A., Ariyoshi, K., Bouzas, M. B., Cahn, P., Sugiura, W., Soriano, V., Brigido, L. F., Grossman, Z., Morris, L., Vandamme, A. M., Tanuri, A., Phanuphak, P., Weber, J. N., Pillay, D., Harrigan, P. R., Camacho, R., Schapiro, J. M., and Shafer, R. W. (2005). Impact of HIV-1 subtype and antiretroviral therapy on protease and reverse transcriptase genotype: results of a global collaboration. *PLoS Med* **2**(4)**,** e112.

Kempf, D. J., Marsh, K. C., Denissen, J. F., McDonald, E., Vasavanonda, S., Flentge, C. A., Green, B. E., Fino, L., Park, C. H., Kong, X. P., and et al. (1995). ABT-538 is a potent inhibitor of human immunodeficiency virus protease and has high oral bioavailability in humans. *Proceedings of the National Academy of Sciences of the United States of America* **92**(7)**,** 2484-8.

Kim, E. E., Baker, C. T., Dwyer, M. D., Murcko, M. A., Rao, B. G., Tung, R. D., and Navia, M. A. (1995). Crystal structure of HIV-1 protease in complex with VX-478, A potent and orally bioavailable inhibitor of the enzyme. *J Am Chem Soc* **117,** 1181.

King, N. M., Melnick, L., Prabu-Jeyabalan, M., Nalivaika, E. A., Yang, S.-S., Gao, Y., Nie, X., Zepp, C., Heefner, D. L., and Schiffer, C. A. (2002). Lack of synergy for inhibitors targeting a multi-drug-resistant HIV-1 protease. *Protein Sci* **11**(2)**,** 418-429.

King, N. M., Prabu-Jeyabalan, M., Nalivaika, E., and Schiffer, C. A. (2004a). Combating susceptibility to drug resistance: lessons from HIV protease. *Chemistry and Biology* **11,** 1333-1338.

King, N. M., Prabu-Jeyabalan, M., Nalivaika, E., and Schiffer, C. A. (2004b). Combating Susceptibility to Drug Resistance: Lessons from HIV-1 Protease. *Chemistry and Biology* **11,** 1333-1338.

King, N. M., Prabu-Jeyabalan, M., Nalivaika, E. A., Wigerinck, P., de Bethune, M.-P., and Schiffer, C. A. (2004c). Structural and thermodynamic basis for the binding of TMC114, a next-generation human immunodeficiency virus type-1 protease inhibitor. *J. Virol.* **78**(21)**,** 12012-12021.

King, N. M., Prabu-Jeyabalan, M., Wigerinck, P., De Bethune, M.-P., and Schiffer, C. A. (2004d). Structural and thermodynamic basis for the binding of TMC114, a next-generation human immunodeficiency virus type 1 protease inhibitor. *J Virol* **78**(21)**,** 12012-12021.

Kiwanuka, N., Laeyendecker, O., Robb, M., Kigozi, G., Arroyo, M., McCutchan, F., Eller, L. A., Eller, M., Makumbi, F., Birx, D., Wabwire-Mangen, F., Serwadda, D., Sewankambo, N. K., Quinn, T. C., Wawer, M., and Gray, R. (2008). Effect of human immunodeficiency virus Type 1 (HIV-1) subtype on disease progression in persons from Rakai, Uganda, with incident HIV-1 infection. *Journal of Infectious Diseases* **197**(5)**,** 707-13.

Knight, S. C., Macatonia, S. E., and Patterson, S. (1990). HIV I infection of dendritic cells. *International Reviews of Immunology* **6**(2-3)**,** 163-75.

Koenig, S., Gendelman, H. E., Orenstein, J. M., Dal Canto, M. C., Pezeshkpour, G. H., Yungbluth, M., Janotta, F., Aksamit, A., Martin, M. A., and Fauci, A. S. (1986). Detection of AIDS virus in macrophages in brain tissue from AIDS patients with encephalopathy. *Science* **233**(4768)**,** 1089-93.

Kohl, N. E., Emini, E. A., Schleif, W. A., Davis, L. J., Heimbach, J. C., Dixon, R. A., Scolnick, E. M., and Sigal, I. S. (1988). Active human immunodeficiency virus protease is required for viral infectivity. *Proc Natl Acad Sci U S A* **85,** 4686-4690.

Kolli, M., Lastere, S., and Schiffer, C. A. (2006). Co-evolution of nelfinavir-resistant HIV-1 protease and the p1-p6 substrate. *Virology* **347**(2)**,** 405-9.

Kolli, M., Stawiski, E., Chappey, C., and Schiffer, C. A. (2009). Human immunodeficiency virus type 1 protease-correlated cleavage site mutations enhance inhibitor resistance. *Journal of Virology* **83**(21)**,** 11027-42.

Krohn, A., Redshaw, S., Ritchie, J. C., Graves, B. J., and Hatada, M. H. (1991). Novel binding mode of highly potent HIV-proteinase inhibitors incorporating the (R)-hydroxyethylamine isostere. *J Med Chem* **34,** 3340-3342.

Krueger, S., Gregurick, S., Shi, Y., Wang, S., Wladkowski, B. D., and Schwarz, F. P. (2003). Entropic Nature of the Interaction between Promoter Bound CRP Mutants and RNA Polymerase. *Biochemistry* **42**(7)**,** 1958-1968.

Los Alamos National Laboratory (2008). HIV Sequence Compendium 2008. Kuiken, C., Foley, B., Preston, M., Wolinksy, S., Leitner, T., Hahn, B. H., McCutchan, F., and Korber, B.

Kuiken, C., Thakallapalli, R., Esklid, A., and de Ronde, A. (2000). Genetic analysis reveals epidemiologic patterns in the spread of human immunodeficiency virus. *American Journal of Epidemiology* **152**(9)**,** 814-22.

Kuroki, R., Nitta, K., and Yutani, K. (1992). Thermodynamic changes in the binding of Ca2+ to a mutant human lysozyme (D86/92). Enthalpy-entropy compensation observed upon Ca2+ binding to proteins. *J Biol Chem* **267**(34)**,** 24297-301.

Ladbury, J. E., and Chowdhry, B. Z., Eds. (1998). Biocalorimetry. Applications of Calorimetry in the Biological Sciences. . Chichester: John Wiley & Sons Ltd.

Lafont, V., Armstrong, A. A., Ohtaka, H., Kiso, Y., Mario Amzel L, and E., F. (2007). Compensating enthalpic and entropic changes hinder binding affinity optimization. *Chem. Biol. Drug Des.* **69,** 413-422.

Lama, J., and Planelles, V. (2007). Host factors influencing susceptibility to HIV infection and AIDS progression. *Retrovirology* **4,** 52.

Laskowski, R. A., Mac Arthur, M. W., Moss, D. S., and Thornton, J. M. (1993). PROCHECK. A program to check the stereochemical quality of protein structures. *J Appl Cryst* **26,** 283-291.

Liu, H., Muller-Plathe, F., and van Gunsteren, W. F. (1996). A combined quantum/classical molecular dynamics study of the catalytic mechanism of HIV protease. *Journal of Molecular Biology* **261**(3)**,** 454-69.

Los Alamos Laboratory (2010). HIV Sequence Database. http://www.hiv.lanl.gov/. Los Alamos National Laboratory.

Masquelier, B., Assoumou, K. L., Descamps, D., Bocket, L., Cottalorda, J., Ruffault, A., Marcelin, A. G., Morand-Joubert, L., Tamalet, C., Charpentier, C., Peytavin, G., Antoun, Z., Brun-Vezinet, F., and Costagliola, D. (2008). Clinically validated mutation scores for HIV-1 resistance to fosamprenavir/ritonavir. *J Antimicrob Chemother* **61**(6)**,** 1362-8.

McCutchan, F. E. (2006). Global epidemiology of HIV. *J Med Virol* **78 Suppl 1,** S7-S12. Purdue University. (1993).

XDISPLAYF program. Minor, W.

Miyauchi, K., Kim, Y., Latinovic, O., Morozov, V., and Melikyan, G. B. (2009). HIV enters cells via endocytosis and dynamin-dependent fusion with endosomes. *Cell* **137**(3)**,** 433-44.

Murphy, E., Korber, B., Georges-Courbot, M. C., You, B., Pinter, A., Cook, D., Kieny, M. P., Georges, A., Mathiot, C., Barre-Sinoussi, F., and et al. (1993). Diversity of V3 region sequences of human immunodeficiency viruses type 1 from the central African Republic. *AIDS Res Hum Retroviruses* **9**(10)**,** 997-1006.

Muzammil, S., Armstrong, A. A., Kang, L. W., Jakalian, A., Bonneau, P. R., Schmelmer, V., Amzel, L. M., and Freire, E. (2007). Unique thermodynamic response of tipranavir to human immunodeficiency virus type 1 protease drug resistance mutations. *J Virol* **81**(10)**,** 5144-54.

Muzammil, S., Ross, P., and Freire, E. (2003). A Major Role for a Set of Non-Active Site Mutations in the Development of HIV-1 Protease Drug Resistance. *Biochemistry* **42**(3)**,** 631-638.

Navaza, J. (1994). AMoRe: an automated package for molecular replacement. *Acta Crystallogr D Biol Crystallogr* **A50,** 157-163.

Navia, M. A., and McKeever, B. M. (1990). A role for the aspartyl protease from the human immunodeficiency virus type 1 (HIV-1) in the orchestration of virus assembly. *Annals of the New York Academy of Sciences* **616,** 73-85.

Nicholls, A., Sharp, K., and Honig, B. (1991a). Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* **11,** 281-296.

Nicholls, A., Sharp, K. A., and Honig, B. (1991b). Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* **11**(4)**,** 281-96.

Nijhuis, M., Schuurman, R., de Jong, D., Erickson, J., Gustchina, E., Albert, J., Schipper, P., Gulnik, S., and Boucher, C. A. (1999). Increased fitness of drug resistant HIV-1 protease as a result of acquisition of compensatory mutations during suboptimal therapy. *AIDS* **13**(17)**,** 2349-59.

Nunez, M., de Mendoza, C., Valer, L., Casas, E., Lopez-Calvo, S., Castro, A., Roson, B., Podzamczer, D., Rubio, A., Berenguer, J., and Soriano, V. (2002). Resistance mutations in HIV-infected patients experiencing early failure with nelfinavir-containing triple combinations. *Med Sci Monit* **8**(9)**,** CR620-3.

Ode, H., Matsuyama, S., Hata, M., Hoshino, T., Kakizawa, J., and Sugiura, W. (2007). Mechanism of drug resistance due to N88S in CRF01_AE HIV-1 protease, analyzed by molecular dynamics simulations. *J Med Chem* **50**(8)**,** 1768-77.

Ode, H., Ota, M., Neya, S., Hata, M., Sugiura, W., and Hoshino, T. (2005). Resistant mechanism against nelfinavir of human immunodeficiency virus type 1 proteases. *J Phys Chem B* **109**(1)**,** 565-74.

Ohtaka, H., and Freire, E. (2005). Adaptive inhibitors of the HIV-1 protease. *Prog Biophys Mol Biol.* **88**(2)**,** 193-208.

Ohtaka, H., Muzammil, S., Schön, A., Velazquez-Campoy, A., Vega, S., and Freire, E. (2004). Thermodynamic rules for the design of high affinity HIV-1 protease inhibitors with adaptability to mutations and high selectivity towards unwanted targets. *Int. J Biochem. Cell Biol.* **36,** 1787-1799.

Ohtaka, H., Schon, A., and Freire, E. (2003). Multidrug Resistance to HIV-1 Protease Inhibition Requires Cooperative Coupling between Distal Mutations. *Biochemistry* **42**(46)**,** 13659-13666.

Ohtaka, H., Velazquez-Campoy, A., Xie, D., and Freire, E. (2002). Overcoming drug resistance in HIV-1 chemotherapy: the binding thermodynamics of Amprenavir and TMC-126 to wild-type and drug-resistant mutants of the HIV-1 protease. *Protein Sci* **11**(8)**,** 1908-1916.

Olsen, D. B., Stahlhut, M. W., Rutkowski, C. A., Schock, H. B., vanOlden, A. L., and Kuo, L. C. (1999). Non-active site changes elicit broad-based cross-resistance of the HIV-1 protease to inhibitors. *J Biol Chem* **274,** 23699-23701.

Orenstein, J. M. (2001). The macrophage in HIV infection. *Immunobiology* **204**(5)**,** 598-602.

Osmanov, S., Pattou, C., Walker, N., Schwardlander, B., and Esparza, J. (2002). Estimated global distribution and regional spread of HIV-1 genetic subtypes in the year 2000. *Journal of Acquired Immune Deficiency Syndromes* **29**(2)**,** 184-90.

Otwinowski, Z. (1993). *CCP4 Study weekend: Data collection and processing, 29-30 Jan 1993, SERC Daresbury Laboratory, England.*

Otwinowski, Z., and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol* **276,** 307-326.

Painter, J., and Merritt, E. A. (2006). Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallogr D Biol Crystallogr* **62**(Pt 4)**,** 439-50.

Parkin, N., Stawiski, E., Chappey, C., and Coakley, E. (2007). Darunavir/amprenavir cross-resistance in clinical samples submitted for phenotype/genotype combination resistance testing. *In* "Conference on Retroviruses and Opportunistic Infections (CROI)", Vol. Abstract #607, San Francisco, CA.

Parris, K. D., Hoover, D. J., Damon, D. B., and Davies, D. R. (1992). Synthesis and crystallographic analysis of two rhizopuspepsin inhibitor complexes. *Biochemistry* **31**(35)**,** 8125-41.

Partaledis, J. A., Yamaguchi, K., Tisdale, M., Blair, E. E., Falcione, C., Maschera, B., Myers, R. E., Pazhanisamy, S., Futer, O., Cullinan, A. B., and et al. (1995a). In vitro selection and characterization of human immunodeficiency virus type 1 (HIV-1) isolates with reduced sensitivity to hydroxyethylamino sulfonamide inhibitors of HIV-1 aspartyl protease. *J Virol* **69**(9)**,** 5228-5235.

Partaledis, J. A., Yamaguchi, K., Tisdale, M., Blair, E. E., Falcione, C., Maschera, B., Myers, R. E., Pazhanisamy, S., Futer, O., Cullinan, A. B., and et al. (1995b). In vitro selection and characterization of human immunodeficiency virus type 1 (HIV-1) isolates with reduced sensitivity to hydroxyethylamino sulfonamide inhibitors of HIV-1 aspartyl protease. *Journal of Virology* **69**(9)**,** 5228-35.

Pauza, C. D. (1991). The endocytic pathway for human immunodeficiency virus infection. *Advances in Experimental Medicine and Biology* **300,** 111-38; discussion 139-44.

Pazhanisamy, S., Stuver, C. M., Cullinan, A. B., Margolin, N., Rao, B. G., and Livingston, D. J. (1996). Kinetic characterization of human immunodeficiency virus type-1 protease-resistant variants. *Journal of Biological Chemistry* **271**(30)**,** 17979-85.

Pearl, L. (1985). "Aspartic Proteinases and Their Inhibitors." (V. Kostka, Ed.) de Gruyter, Berlin.

Pettit, S. C., Sheng, N., Tritch, R., Erickson-Vitanen, S., and Swanstrom, R. (1998). The regulation of sequential processing of HIV-1 Gag by the viral protease. *Adv Exp Med Biol* **436,** 15-25.

Plantier, J. C., Leoz, M., Dickerson, J. E., De Oliveira, F., Cordonnier, F., Lemee, V., Damond, F., Robertson, D. L., and Simon, F. (2009). A new human immunodeficiency virus derived from gorillas. *Nature Medicine* **15**(8)**,** 871-2.

Popovic, M., Sarngadharan, M. G., Read, E., and Gallo, R. C. (1984). Detection, isolation, and continuous production of cytopathic retroviruses (HTLV-III) from patients with AIDS and pre-AIDS. *Science* **224**(4648)**,** 497-500.

Poveda, E., de Mendoza, C., Martin-Carbonero, L., Corral, A., Briz, V., Gonzalez-Lahoz, J., and Soriano, V. (2007). Prevalence of darunavir resistance mutations in HIV-1-infected patients failing other protease inhibitors. *J Antimicrob Chemother* **60**(4)**,** 885-8.

Prabu-Jeyabalan, M., King, N. M., Nalivaika, E., Heilek-Snyder, G., Cammack, N., and Schiffer, C. A. (2006a). Substrate Envelope and Drug Resistance: Crystal Structure of RO1 in Complex with Wild-Type Human Immunodeficiency Virus Type 1 Protease. *Antimicrob Agents Chemother* **50**(4)**,** 1518-1521.

Prabu-Jeyabalan, M., Nalivaika, E., King, N. M., and Schiffer, C. A. (2004a). Structual basis for coevolution of the human immunodeficiency virus type 1 nucleocapsid-p1 cleavage site with a V82A drug-resistant mutation in viral protease. *J Virol* **78**(22)**,** 12446-12454.

Prabu-Jeyabalan, M., Nalivaika, E., and Schiffer, C. A. (2000a). How does a symmetric dimer recognize an asymmetric substrate? A substrate complex of HIV-1 protease. *J Mol Biol* **301**(5)**,** 1207-20.

Prabu-Jeyabalan, M., Nalivaika, E., and Schiffer, C. A. (2000b). How does a symmetric dimer recognize an asymmetric substrate? A substrate complex of HIV-1 protease. *J Mol Biol* **301**(5)**,** 1207-20.

Prabu-Jeyabalan, M., Nalivaika, E. A., King, N. M., and Schiffer, C. A. (2004b). Structural basis for coevolution of a human immunodeficiency virus type 1 nucleocapsid-p1 cleavage site with a V82A drug-resistant mutation in viral protease. *J Virol* **78**(22)**,** 12446-54.

Prabu-Jeyabalan, M., Nalivaika, E. A., Romano, K., and Schiffer, C. A. (2006b). Mechanism of substrate recognition by drug-resistant human immunodeficiency virus type 1 protease variants revealed by a novel structural intermediate. *J Virol* **80**(7)**,** 3607-16.

Prabu-Jeyabalan, M., Nalivaika, E. A., and Schiffer, C. A. (2002). Substrate shape determines specificity of recognition for HIV-1 protease: Analysis of crystal structures of six substrate complexes. *Structure* **10**(3)**,** 369-381.

Preston, B. D., Poiesz, B. J., and Loeb, L. A. (1988). Fidelity of HIV-1 reverse transcriptase. *Science* **242**(4882)**,** 1168-71.

Rekharsky, M., and Inoue, Y. (1998). Complexation Thermodynamics of Cyclodextrins. . *Chem Rev.* **98,** 1875-1918.

Renjifo, B., Gilbert, P., Chaplin, B., Msamanga, G., Mwakagile, D., Fawzi, W., and Essex, M. (2004). Preferential in-utero transmission of HIV-1 subtype C as compared to HIV-1 subtype A or D. *AIDS* **18**(12)**,** 1629-36.

Robbins, A. H., Coman, R. M., Bracho-Sanchez, E., Fernandez, M. A., Gilliland, C. T., Li, M., Agbandje-McKenna, M., Wlodawer, A., Dunn, B. M., and McKenna, R. (2010). Structure of the unbound form of HIV-1 subtype A protease: comparison with unbound forms of proteases from other HIV subtypes. *Acta Crystallographica. Section D: Biological Crystallography* **66**(Pt 3)**,** 233-42.

Roberts, J. D., Bebenek, K., and Kunkel, T. A. (1988). The accuracy of reverse transcriptase from HIV-1. *Science* **242**(4882)**,** 1171-3.

Roberts, N. A. (1995). Drug-resistance patterns of saquinavir and other HIV proteinase inhibitors. *AIDS* **9 (supplement 2),** S27-S32.

Roberts, N. A., Martin , J. A., Kinchington, D., Broadhurst, A. V., Craig, J. C., Duncan, I. B., Galpin, S. A., Handa, B. K., Kay, J., Krohn, A., and et al. (1990). Rational design of peptide-based HIV proteinase inhibitors. *Science* **248**(4953)**,** 358-61.

Robertson, D. L., Anderson, J. P., Bradac, J. A., Carr, J. K., Foley, B., Funkhouser, R. K., Gao, F., Hahn, B. H., Kalish, M. L., Kuiken, C., Learn, G. H., Leitner, T., McCutchan, F., Osmanov, S., Peeters, M., Pieniazek, D., Salminen, M., Sharp, P. M., Wolinsky, S., and Korber, B. (2000). HIV-1 nomenclature proposal. *Science* **288**(5463)**,** 55-6.

Robinson, B. S., Riccardi, K. A., Gong, Y. F., Guo, Q., Stock, D. A., Blair, W. S., Terry, B. J., Deminie, C. A., Djang, F., Colonno, R. J., and Lin, P. F. (2000). BMS-232632, a highly potent human immunodeficiency virus protease inhibitor that can be used in combination with other available antiretroviral agents. *Antimicrobial Agents and Chemotherapy* **44**(8)**,** 2093-9.

Rose, J. R., Salto, R., and Craik, C. S. (1993a). Regulation of autoproteolysis of the HIV-1 and HIV-2 proteases with engineered amino acid substitutions. *J Biol Chem* **268,** 11939-11945.

Rose, J. R., Salto, R., and Craik, C. S. (1993b). Regulation of HIV-1 and HIV-2 proteases with engineered amino acid substitutions. *J Biol Chem* **268**(16)**,** 11939-11945.

Rose, R. B., Craik, C. S., and Stroud, R. M. (1998a). Domain flexibility in retroviral proteases: structural implications for drug resistant mutations. *Biochemistry* **37,** 2607-2621.

Rose, R. B., Craik, C. S., and Stroud, R. M. (1998b). Domain flexibility in retroviral proteases: structural implications for drug resistant mutations. *Biochemistry* **37**(8)**,** 2607-21.

Sadler, B. M., and Stein, D. S. (2002). Clinical pharmacology and pharmacokinetics of amprenavir. *Annals of Pharmacotherapy* **36**(1)**,** 102-18.

Sanches, M., Krauchenco, S., Martins, N. H., Gustchina, A., Wlodawer, A., and Polikarpov, I. (2007a). Structural characterization of B and non-B subtypes of HIV-protease: insights into the natural susceptibility to drug resistance development. *Journal of Molecular Biology* **369**(4)**,** 1029-40.

Sanches, M., Krauchenco, S., Martins, N. H., Gustchina, A., Wlodawer, A., and Polikarpov, I. (2007b). Structural characterization of B and non-B subtypes of HIV-protease: insights into the natural susceptibility to drug resistance development. *J Mol Biol* **369**(4)**,** 1029-40.

Scott, W. R., and Schiffer, C. A. (2000). Curling of flap tips in HIV-1 protease as a mechanism for substrate entry and tolerance of drug resistance. *Structure* **8**(12)**,** 1259-65.

Shafer, R. W., Stevenson, D., and Chan, B. (1999). Human immunodeficiency virus reverse transcriptase and protease sequence database. *Nucleic Acids Res* **27**(1)**,** 348-352.

Shafer, R. W., and Vuitton, D. A. (1999). Highly active antiretroviral therapy (HAART) for the treatment of infection with human immunodeficiency virus type 1. *Biomedicine and Pharmacotherapy* **53**(2)**,** 73-86.

Sham, H. L., Kempf, D. J., Molla, A., Marsh, K. C., Kumar, G. N., Chen, C. M., Kati, W., Stewart, K., Lal, R., Hsu, A., Betebenner, D., Korneyeva, M., Vasavanonda, S., McDonald, E., Saldivar, A., Wideburg, N., Chen, X., Niu, P., Park, C., Jayanti, V., Grabowski, B., Granneman, G. R., Sun, E., Japour, A. J., Leonard, J. M., Plattner, J. J., and Norbeck, D. W. (1998). ABT-378, a highly potent inhibitor of the human immunodeficiency virus protease. *Antimicrobial Agents and Chemotherapy* **42**(12)**,** 3218-24.

Sharp, K. (2001). Entropy—enthalpy compensation: Fact or artifact? *Prot. Sci.* **10,** 661-667.

Sharp, P. M., Robertson, D. L., and Hahn, B. H. (1995). Cross-species transmission and recombination of 'AIDS' viruses. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **349**(1327)**,** 41-7.

Sheehy, A. M., Gaddis, N. C., Choi, J. D., and Malim, M. H. (2002). Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. *Nature* **418**(6898)**,** 646-50.

Sigurskjold, B. (2000). Exact analysis of competition ligand binding by displacement isothermal titration calorimetry. *Anal Biochem* **277**(2)**,** 260-266.

Silva, A. M., Cachau, R. E., Sham, H. L., and Erickson, J. W. (1996). Inhibition and catalytic mechanism of HIV-1 aspartic protease. *J Mol Biol* **255**(2)**,** 321-46.

Spira, S., Wainberg, M. A., Loemba, H., Turner, D., and Brenner, B. G. (2003a). Impact of clade diversity on HIV-1 virulence, antiretroviral drug sensitivity and drug resistance. *J Antimicrob Chemother* **51**(2)**,** 229-40.

Spira, S., Wainberg, M. A., Loemba, H., Turner, D., and Brenner, B. G. (2003b). Impact of clade diversity on HIV-1 virulence, antiretroviral drug sensitivity and drug resistance. *Journal of Antimicrobial Chemotherapy* **51**(2)**,** 229-40.

Steuber, H., Heine, A., and Klebe, G. (2007). Structural and thermodynamic study on aldose reductase: nitro-substituted inhibitors with strong enthalpic binding contribution. *J Mol Biol* **368**(3).

Stoica, I., Sadiq, S. K., and Coveney, P. V. (2008). Rapid and accurate prediction of binding free energies for saquinavir-bound HIV-1 proteases. *J Am Chem Soc* **130,** 2639-2648.

Subbramanian, R. A., and Cohen, E. A. (1994). Molecular biology of the human immunodeficiency virus accessory proteins. *Journal of Virology* **68**(11)**,** 6831-5.

Suguna, K., Padlan, E. A., Smith, C. W., Carlson, W. D., and Davies, D. R. (1987). Binding of a reduced peptide inhibitor to the aspartic proteinase from Rhizopus chinensis: implications for a mechanism of action. *Proc Natl Acad Sci U S A* **84**(20)**,** 7009-13.

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., and Wigerinck, P. B. (2005a). Design of HIV-1 protease inhibitors active on multidrug-resistant virus. *J Med Chem* **48**(6)**,** 1965-73.

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., and Wigerinck, P. B. (2005b). Discovery and selection of TMC114, a next generation HIV-1 protease inhibitor. *J Med Chem* **48**(6)**,** 1813-22.

Takeuchi, Y., Nagumo, T., and Hoshino, H. (1988). Low fidelity of cell-free DNA synthesis by reverse transcriptase of human immunodeficiency virus. *Journal of Virology* **62**(10)**,** 3900-2.

Tisdale, M., Myers, R. E., Ait-Khaled, M., and Snowden, W. A. (1999). HIV drug resistance analysis during clinical studies with the protease inhibitor amprenavir. *In* "Sixth Conference on Retroviruses and Opportunistic Infections", Chicago, IL.

Todd, M. J., and Freire, E. (1999a). The effect of inhibitor binding on the structural stability and cooperativity of the HIV-1 protease. *Proteins* **36**(2)**,** 147-56.

Todd, M. J., and Freire, E. (1999b). The effect of inhibitor binding on the structural stability and cooperativity of the HIV-1 protease. *Proteins* **36**(2)**,** 147-56.

Todd, M. J., Luque, I., Velazquez-Campoy, A., and Freire, E. (2000). Thermodynamic basis of resistance to HIV-1 protease inhibition: calorimetric analysis of the V82F/I84V active site resistant mutant. *Biochemistry* **39,** 11876-11883.

Turner, S. R., Strohbach, J. W., Tommasi, R. A., Aristoff, P. A., Johnson, P. D., Skulnick, H. I., Dolak, L. A., Seest, E. P., Tomich, P. K., Bohanon, M. J., Horng, M. M., Lynn, J. C., Chong, K. T., Hinshaw, R. R., Watenpaugh, K. D., Janakiraman, M. N., and Thaisrivongs, S. (1998). Tipranavir (PNU-140690): a potent, orally bioavailable nonpeptidic HIV protease inhibitor of the 5,6-dihydro-4-hydroxy-2-pyrone sulfonamide class. *Journal of Medicinal Chemistry* **41**(18)**,** 3467-76.

UNAIDS (2009). AIDS Epidemic Update 2009. UNAIDS.

Valzaquez-Campoy, A., Luque, I., Todd, M. J., Milutinovich, M., Kiso, Y., and Freire, E. (2000). Thermodynamic dissection of the binding energetics of KNI-272, a potent HIV-1 protease inhibitor. *Protein Sci* **9**(9)**,** 1801-1809.

Valzaquez-Campoy, A., Todd, M. J., and Freire, E. (2000). HIV-1 protease inhibitors: enthalpic versus entropic optimization of the binding affinity. *Biochemistry* **39,** 2201-7.

Valzaquez-Campoy, A., Todd, M. J., Vega, S., and Freire, E. (2001). Catalytic efficiency and vitality of HIV-1 proteases from African viral subtypes. *Proc Natl Acad Sci U S A* **98**(11)**,** 6062-7.

Van Marck, H., Dierynck, I., Kraus, G., Hallenberger, S., Pattery, T., Muyldermans, G., Van Vijmen, H., Hertogs, K., and Bethune, M. (2007). Unraveling the complex resistance pathways of darunavir using the bioinformatics resistance determination (BIRD). *In* "XVI International HIV Drug Resistance Workshop", Barbados.

Veerapandian, B., Cooper, J. B., Sali, A., Blundell, T. L., Rosati, R. L., Dominy, B. W., Damon, D. B., and Hoover, D. J. (1992). Direct observation by X-ray analysis of the tetrahedral "intermediate" of aspartic proteinases. *Protein Sci* **1**(3)**,** 322-8.

Velazquez-Campoy, A., Todd, M. J., Vega, S., and Freire, E. (2001). Catalytic efficiency and vitality of HIV-1 proteases from African viral subtypes. *Proc Natl Acad Sci U S A* **98**(11)**,** 6062-7.

Velazquez-Campoy, A., Vega, S., Fleming, E., Bacha, U., Sayed, Y., Dirr, H. W., and Freire, E. (2003). Protease inhibition in African subtypes of HIV-1. *AIDS Rev* **5**(3)**,** 165-71.

Velazquez-Campoy, A., Vega, S., and Freire, E. (2002). Amplification of the effects of drug resistance mutations by background polymorphisms in HIV-1 protease from African subtypes. *Biochemistry* **41**(27)**,** 8613-9.

Vermeiren, H., Van Craenenbroeck, E., Alen, P., Bacheler, L., Picchio, G., and Lecocq, P. (2007). Prediction of HIV-1 drug susceptibility phenotype from the viral genotype using linear regression modeling. *Journal of Virological Methods* **145**(1)**,** 47-55.

Vondrasek, J., and Wlodawer, A. (2002). HIVdb: a database of the structures of human immunodeficiency virus protease. *Proteins* **47,** 429-431.

Wainberg, M. A. (2004). HIV-1 subtype distribution and the problem of drug resistance. *AIDS* **18 Suppl 3,** S63-8.

Wainberg, M. A., Martinez-Cajas, J. L., and Brenner, B. G. (2007). Strategies for the optimal sequencing of antiretroviral drugs toward overcoming and preventing drug resistance. *Future HIV Therapy* **1**(3)**,** 291-313.

Wang, W. K., Chen, M. Y., Chuang, C. Y., Jeang, K. T., and Huang, L. M. (2000). Molecular biology of human immunodeficiency virus type 1. *J Microbiol Immunol Infect* **33**(3)**,** 131-40.

Williams, T., and Kelley, C. (1998a). GNUPLOT. gnuplot 1986-1993.

Williams, T., and Kelly, C. (1998). gnuplot 1986-1993, 1998, 2004. http://www.gnuplot.info.

Wlodawer, A., and Erickson, J. W. (1993). Structure-based inhibitors of HIV-1 protease. *Annu Rev Biochem* **62,** 543-585.

Wyatt, R., and Sodroski, J. (1998). The HIV-1 envelope glycoproteins: fusogens, antigens, and immunogens. *Science* **280**(5371)**,** 1884-8.

Yanchunas, J., Jr., Langley, D. R., Tao, L., Rose, R. E., Friborg, J., Colonno, R. J., and Doyle, M. L. (2005). Molecular basis for increased susceptibility of isolates with atazanavir resistance-conferring substitution I50L to other protease inhibitors. *Antimicrobial Agents and Chemotherapy* **49**(9)**,** 3825-32.

Zheng, Y. H., Lovsin, N., and Peterlin, B. M. (2005). Newly identified host factors modulate HIV replication. *Immunology Letters* **97**(2)**,** 225-34.

Zhou, P., Tian, F., Lv, F., and Shang, Z. (2009). Geometric characteristics of hydrogen bonds involving sulfur atoms in proteins. *Proteins* **76**(1)**,** 151-63.

Ziermann, R., Limoli, K., Das, K., Arnold, E., Petropoulos, C. J., and Parkin, N. T. (2000). A mutation in human immunodeficiency virus type 1 protease, N88S, that causes in vitro hypersensitivity to amprenavir. *J Virol* **74**(9)**,** 4414-9.