12-16-2019

X-Ray Crystal Structures And Enzyme Kinetic Investigations Of Drug-Resistant Mutants Of Hiv-1 Protease

Daniel Walter Kneller

X-RAY CRYSTAL STRUCTURES AND ENZYME KINETIC INVESTIGATIONS OF DRUG-RESISTANT MUTANTS OF HIV-1 PROTEASE

by

DANIEL W. KNELLER

Under the Direction of Irene T. Weber PhD

ABSTRACT

Globally, 62% of 38 million HIV-infected people are receiving antiretroviral therapy. Inhibitors targeting the viral protease have been clinically successful as 9 protease inhibitors (PIs) have been approved by the FDA since 1995. However, drug resistance arising by mutations in the protease undermines effective treatment. Analysis of protease and its mutants by structural biology methods paired with enzymology has given insight into the molecular mechanisms for drug resistance and guided new strategies for inhibitor design. Recently, highly resistant variants of HIV-1 protease from clinical isolates have been identified with ~20 mutations and several orders of magnitude worse binding affinity for clinical PIs such as darunavir. Three such mutants, PR20, PRS17, and PRS5B, are the focus of this body of work and show 800-10,000-fold less susceptibility to darunavir than wild-type protease.

Understanding the molecular mechanisms driving the extreme drug-resistance of these three protease mutants aid rational drug design efforts to fight the HIV/AIDS pandemic.

Structure-guided strategies for drug design have resulted in an investigational inhibitor, GRL-142, with modified ligands derived from the darunavir scaffold that shows 16-fold better inhibition than darunavir of resistant mutant PR20. The crystal structure of PR20 in complex with GRL-142 reveals how the expanded binding pocket, dynamic flaps, and faster dimer dissociation of PR20 are counteracted by the larger moieties of GRL-142.

Resistant variant PRS17, which was rationally selected from the HIVdb genotype-phenotype database by machine learning, shows ~3-fold better inhibition by peptide substrate analogs compared to wild-type protease. Crystal structures of PRS17 with substrate analogs show a major effect of drug-resistance mutations V82S and G48V improving interactions with substrates consistent with better inhibition, suggesting a novel mechanism for resistance.

Finally, structural studies of another mutant selected by machine learning, PRS5B, reveal coordinated structural changes leading to decreased intra-subunit interactions and intermediate levels of resistance to PIs. The sum of knowledge on resistant variants PR20, PRS17, and PRS5B illuminates the evolution of HIV-1 protease in the era of accessible PI treatments. These results illustrate the power of combining structural analysis of proteins with enzyme kinetics for combatting drug resistant HIV.

INDEX WORDS: HIV-1 Protease, Structure guided-drug design, Drug resistance, X-ray crystallography, Protease inhibitor, Enzyme Kinetics

X-RAY CRYSTAL STRUCTURES AND ENZYME KINETIC INVESTIGATIONS OF DRUG-RESISTANT MUTANTS OF HIV-1 PROTEASE

by

DANIEL W. KNELLER

A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

in the College of Arts and Sciences

Georgia State University

2019

X-RAY CRYSTAL STRUCTURES AND ENZYME KINETIC INVESTIGATIONS OF DRUG-RESISTANT MUTANTS OF HIV-1 PROTEASE

by

DANIEL W. KNELLER

Committee Chair: Irene Weber

Committee: Robert Harrison

Giovanni Gadda

Electronic Version Approved:

Office of Graduate Studies

College of Arts and Sciences

Georgia State University

December 2019

DEDICATION

To the 38 million individuals living with HIV and all those who dream of a world free of the disease.

ACKNOWLEDGEMENTS

The gratitude I hold for my wife, Bria, is impossible to overstate. Her unwavering support through a decade of my higher education defines the word love. These accomplishments would be still be dreams if not for her.

My most sincere appreciation goes to Dr. Irene Weber, who taught me how to ask scientific questions and navigate the answers. I am forever grateful for the opportunities to do science in such a genuine environment. Thank you for always treating me with respect when operating as both a mentor and a peer.

I would like to thank my committee members, Drs. Robert Harrison and Giovanni Gadda who shared their time and expertise in helping me complete projects. These works are better from their guidance. A special thanks to Dr. Johnson Agniswamy and Yuan-Fang Wang, who never once let me down during my research training. I have learned so much from their wisdom and experience.

To my labmates, Dr. Andres Wong, Shelley Burnaman, Rowan Brothers, Dr. Shrikant Pawar, and Dr. Brian Shen, thank you for your friendship in the trenches of the day-to-day research. I am appreciative of the technical assistance provided by GSU Department of Biology, SER-CAT staff, and scientists at Oak Ridge and Argonne National Laboratories. My research efforts were supported by the Molecular Basis of Disease program at GSU and grants from the National Institutes of Health.

Finally, I'd like to thank my parents, Bill and Doreen. My success is a reflection of their enduring dedication to selfless parenting. Thank you for always ensuring I could give my all.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

HIV	Human Immunodeficiency Virus
AIDS	Acquired Immunodeficiency Syndrome
WHO	World Health Organization
UNAIDS	Joint United Nations Programme on HIV/AIDS
ART	Antiretroviral therapy
HAART	Highly active antiretroviral treatment
PR	HIV-1 Protease
PI	Protease Inhibitor
PRS5B	HIV-1 protease with 22 mutations
PR20	HIV-1 protease with 20 mutations
PR^{S17}; PRS17	HIV-1 protease with 17 mutations
THF	Tetrahydrofuran
K_i	Kinetic inhibition value
DRV; **1**	darunavir
APV	amprenavir
SQV	saquinavir
ATV	atazanavir
TPV	tipranvir
LPV	lopinavir
IDV	indinavir
NFV	nelfinavir
GRL-142; **2**	GRL-142-13A
U = U	undetectable = untransmittable
NIH	National Institutes of Health
ENV	HIV envelope glycoprotein
RT	HIV reverse transcriptase
DNA	Deoxyribonucleic acid
IN	HIV integrase
MA	HIV matrix protein
TFP	Transframe protein
RAM	Resistance-associated mutation
ITC	Isothermal Titration Calorimetry
SAXS	Small angle X-ray scattering
SANS	Small angle neutron scattering
Cryo-EM	Cryogenic electron microscopy
AMMP	Another Molecular Mechanics Program
MD	Molecular Dynamics

1 INTRODUCTION

1.1 Human Immunodeficiency Virus and AIDS

Human Immunodeficiency Virus (HIV) is the etiological agent responsible for Acquired Immunodeficiency Syndrome (AIDS). No prophylactic vaccine for HIV is currently available and vaccine development remains a challenge[1]. Only two instances of long-term HIV remission, the famous London and Berlin Patients[2–4], have been described, which occurred through bone-marrow transplantation to treat cancer. Other than these exceptions, HIV infection is incurable and requires lifelong treatment. The World Health Organization (WHO) lists HIV/AIDS as one of the world's most serious public health challenges and consequently recommends a 26.2 billion dollar global investment to reach 2020 targets[5].

1.1.1 From mysterious beginnings to pandemic

Zoonotic transmission of the retrovirus from chimpanzees and gorillas to humans is speculated to originate from bushmeat practice[6]. Genetic studies place the most recent common ancestor of the most prevalent HIV subgroup (HIV-1 M) to ~1910 in what is now Kinshasa, Democratic Republic of the Congo[7]. From there, the first known cases of HIV disseminated through African colonialism-related urbanization, sex-trade, railways, and unsterilized medical activities in the 1920s[8].

Knowledge of the current outbreak of HIV began in 1981 in the United States as separate reports of rare *Pneumocystis carinii pneumonia* (PCP) lung infections and Kaposi's Sarcoma in young men who have sex with men (MSM) and IV drug users[9–11]. Mysterious cases of immune deficiency were reported in Europe as early as March 1982[12–14]. In September 1982, the United States Centers for Disease Control and Prevention (CDC) introduced the acronym AIDS to describe a defect in immunity associated with opportunistic infection primarily affecting MSM,

IV drug users, hemophiliacs, and Haitians with no known cause[15]. HIV was isolated from AIDS patients establishing the virus as the etiologic agent of AIDS in 1983[16–18]. Not long after, HIV antibody detection was developed allowing HIV diagnosis prior to presentation of immunodeficiency symptoms as well as blood screening[19,20]. From 1983 to 1984, the number of deaths from AIDS cases in the US had nearly tripled from 1,292 to 3,665[21] and by the end of 1985, the WHO reported at least one AIDS case in every major region on Earth in their total count of over 20,300 infections[22].

The first HIV antiretroviral drug, a reverse transcriptase inhibitor named azidothymidine (AZT), was approved by the FDA to treat adults in 1987[23]. However, HIV infection spread as researchers scrambled for new treatment options. By 1993, HIV infection was the leading cause of death among persons aged 25-44 in the US[24]. In June 1995, the first HIV protease (PR) inhibitor (PI), saquinavir (SQV), was approved by the FDA, ushering in the era of highly active antiretroviral treatment (HAART)[25]. The introduction of PIs in combination with other antiretrovirals brought about a drastic decrease in AIDS-related mortality[26].

When utilized properly, current pre-exposure prophylaxis (PrEP) and HAART treatment options have been successful at reducing sexual transmission of HIV in high risk groups[27–29]. However, sociopolitical and economic obstacles limit quality of life and treatment access for many infected individuals[30]. The United Nations Program on HIV and AIDS (UNAIDS) launched its ambitious 90-90-90 program in 2014 which aims for 90% of known infected individuals who are receiving antiretroviral therapy to have viral suppression[31]. In 2018,UNAIDS reported more than 62% of infected individuals (~22.3 million) were receiving antiretroviral treatment globally[32].

Widespread ART has fostered accumulation of drug resistance mutations for first line drugs leading to treatment failure[33,34]. The development of potent treatment options to combat drug-resistant virus is critical for continued success in suppressing the AIDS pandemic.

1.1.2 HIV transmission and replication cycle

HIV transmission often occurs via subclinically infected individuals because HIV is a lentivirus, a genus of retrovirus characterized by a long latency period. Infected individuals can rapidly accumulate large viral load for months or even years while remaining asymptomatic. This latency period is an important contributing factor in the spread of the virus across the globe.

The virus can be transmitted by sexual and non-sexual means. HIV is sexually transmitted from infected individuals to uninfected individuals through transfer of blood, semen, and vaginal fluids. HIV can be transmitted non-sexually from an infected mother during pregnancy, birth, and breast milk. Parenteral exposure can occur through needle-sharing during injection drug use and in rare cases of blood transfusions and percutaneous needle stick. HIV is sexually transmitted from infected individuals to uninfected individuals through transfer of blood, semen, and vaginal fluids.

Risk of sexual transmission is increased by individuals at a later stage of infection, sex act, higher viral load, and presence of other sexually transmitted diseases. Risk is decreased by condom use, male circumcision, PReP, and ART. A 2014 review by CDC researchers Patel *et al.* estimated per-act transmission probabilities for major exposure paths[35]. Per-act sexual intercourse transmission risk of HIV ranged from 0.04 to 1.38% with 99.2% attenuation with use of condoms and ART for the infected partner[36].

The current Prevention Access undetectable = untransmittable (U=U) campaign captures the current zeitgeist surrounding HIV treatment and transmission[37]. U=U describes HIV

treatment as a prevention strategy supported by 4 landmark studies showing that maintaining an undetectable viral load carries zero risk of HIV transmission in serodifferent couples during unprotected sex[38–41]. The treatment as prevention movement reinforces the need for effective ART for all at risk individuals in order to eradicate HIV.

HIV transmission is dependent on the viral load in an individual[42,43]. High viral titers are a result of successful integration and replication of the virus in $CD4^+$ T cells. Plasma HIV RNA levels are highest during the acute phase of infection (3-9 weeks after infection) then drop to the infection set point (4-6 months) for the chronic phase of infection (reviewed in [44]). The progressive depletion of $CD4^+$ cells leads to AIDS and chronic inflammation, risk of opportunistic infection, and oncological complication[45].

The basic replication cycle of HIV with major drug target steps[46] is shown in figure 1.1 and reviewed in detail by Freed in [47,48]. An infectious HIV particle enters the cell through attachment between the envelope glycoprotein (ENV) with the cell CD4 receptor and either CC-chemokine receptor 5 (CCR5) or the CXC-chemokine receptor 4 (CXCR4). The cell membranes of the incoming virion and target cell are fused together, and the viral core enters the host cytoplasm. The viral core is uncoated allowing the viral reverse transcriptase (RT) to convert the single-stranded, positive sense RNA genome into deoxyribonucleic acid (DNA). The viral DNA is then transported to the nucleus and the viral enzyme integrase (IN) inserted into the host DNA. The integrated provirus is able to serve as a template for human transcription machinery to generate viral RNAs. RNAs encoding the viral structure and enzymes (ENV, Gag, and Pol) as well as replication elements (Rev and Tat), are exported to the cytoplasm to be translated. The viral RNA genome is exported to the cytoplasm by Rev. New virus particles are produced when the Gag and Gag-Pol precursor polyproteins encapsulate the RNA genome, associate with ENV,

and assemble at the membrane using matrix protein (MA). A non-infectious immature particle is released in a process called budding. The final maturation step occurs when the viral HIV protease (PR) cleaves the Gag and Gag-Pol polyproteins into their individual viral components. Capsid (CA) and nucleocapsid (NC) encapsulate the RNA genome, RT, and IN forming a dense icosahedral core.

The now infectious mature virion is free to infect a new cell. An infected cell and a mature virion in plasma have a half-life of 2.2 and 0.24 days respectively[49]. The HIV-1 particle generation time, from budding of infected parent cell to budding from a newly infected daughter cell, is 2.6 days[50]. In an infected individual, the replication process will produce an average 10.3 billion HIV viral particles per day[50]. While ART is effective at inhibiting many steps in the HIV replication cycle, drug resistance[51] and residual viral replication in sanctuary sites[52] remain important issues for HIV research.

Four classes of ART drugs target four stages of the HIV replication cycle (figure 1.1) – entry/fusion, reverse transcription, integration, and maturation. Most modern ART regimens use combinations of inhibitors for two targets. Recommended first line therapy for most people with HIV is an integration inhibitor plus two nucleoside-RT inhibitors[53]. Second line ART options are patient specific but often are protease inhibitor (PI) based[53]. Demand for second-line treatments is increasing. In sub-Saharan Africa, an estimated 4.6 million people, or 20% of those receiving ART, are expected to be receiving second-line ART in 2030 – a 5% increase from 2020[54].

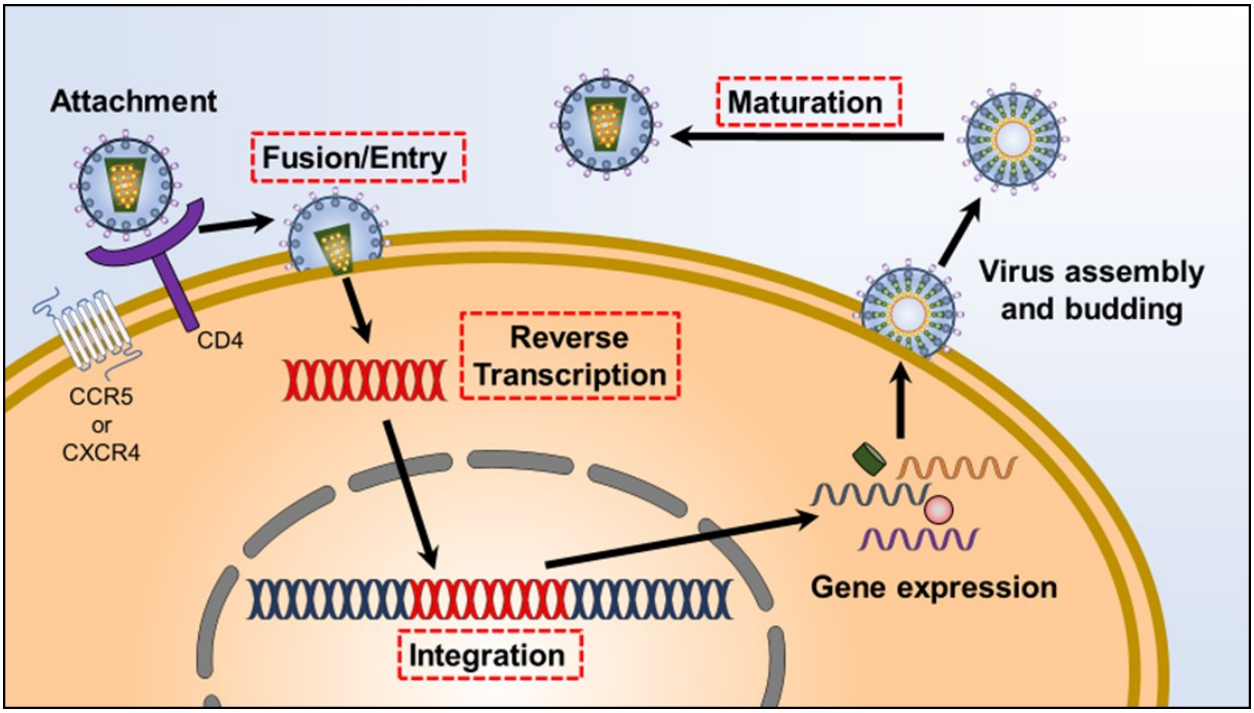


Figure 1.1: Major events in the HIV replication cycle.
After virion attachment and fusion, the viral core is released into the cytoplasm. The RNA genome is reverse transcribed into DNA which is then integrated into the host cell genome. Structural proteins and polyproteins are expressed and assemble at the cell membrane before an immature viral particle. Maturation via HIV PR processing of the Gag and Gag-Pol polyproteins produces a fully infectious HIV viral particle. Events outlined in red boxes are used as targets for inhibition by ART.

1.1.3 Current HIV prevalence and treatment accessibility

ART is recommended for all patients with HIV as soon as possible in an effort to reduce risk of morbidity and transmission[53]. ART can be effective in suppressing HIV replication but is not a cure. Stopping ART will allow the infection to rebound within weeks[55]. Current guidelines emphasize the importance for educating patients on proper ART adherence to maintain viral suppression and minimize the emergence of drug resistance[53].

Figure 1.2 shows the global statistics for HIV prevalence, ART accessibility, and deaths related to HIV from 2000 to 2018[56]. Right now, an estimated 37.9 million people globally are living with HIV and 23.3 million (62%) are accessing ART - up from 7.7 million in 2010[56]. The increase in ART access has improved clinical outcomes. In 2018, 770,000 people died from AIDS-related illnesses, down 55% compared to peak in 2004 (1.7 million)[56].

Global ART accessibility has improved over time but all populations of HIV infected and at risk people must be treated in order to eradicate the disease[57,58]. Figure 1.3 shows current statistics for ART accessibility between major populations in the world[56]. These data are also tabulated for new infections and AIDS-related deaths in Table 1.1[56]. The highest percentage of HIV-infected people accessing ART is in Europe and North America (79%). However, less than 40% of infected individuals in Eastern Europe, central Asia, the Middle East, and North Africa are accessing ART. In the last two decades, prolonged ART has become an effective means to control the disease and early diagnosis and treatment allows people with HIV to enjoy a life expectancy close to the general population[59–61].

Key populations and their sexual partners accounted for the majority of new infections in 2018[56,62,63]. Key populations are defined by UNAIDS as MSM, sex workers, transgender people, injection drug users, and prisoners[64]. The risk of acquiring HIV is over 20 times greater than the

general population for MSM, injection drug users, and sex workers[56]. All populations across all regions must have HIV diagnostic tests and life-long accessible ART in order for eradication strategies to succeed[65–68].

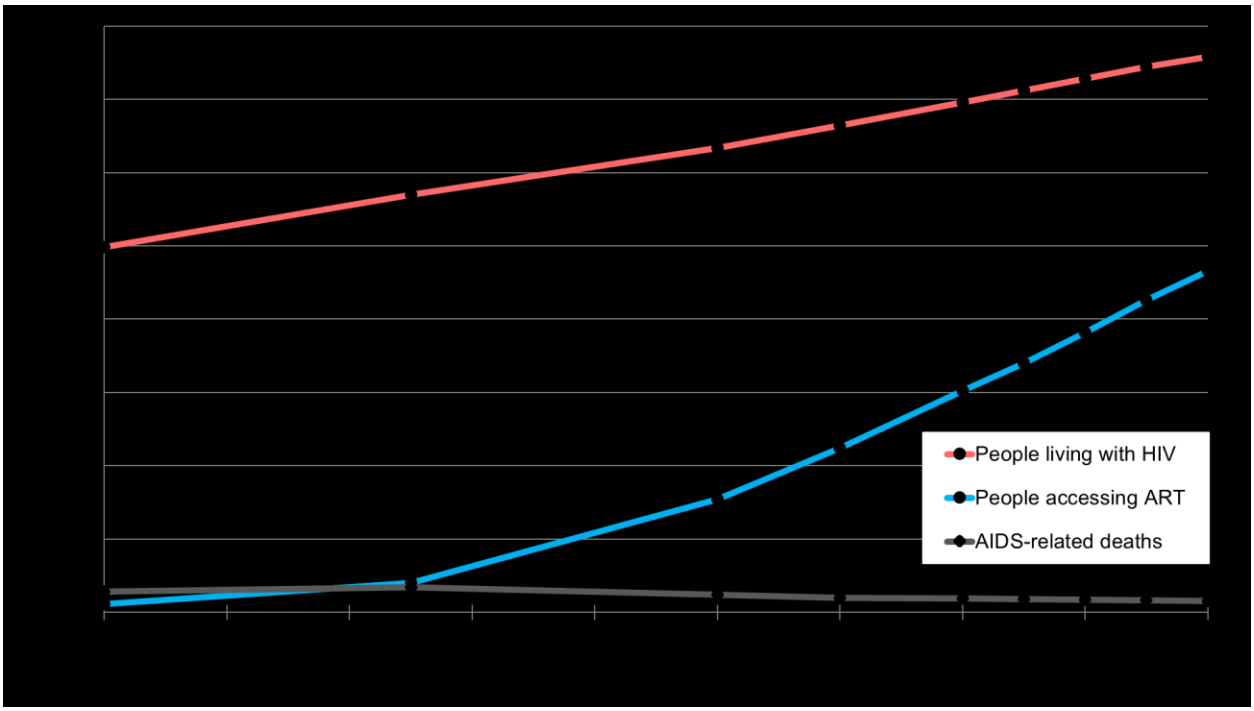


Figure 1.2: Worldwide number of people living with HIV, accessing ART, and AIDS-related deaths from 2000-2018.
Data adapted from UNAIDS at unaids.org, accessed 7/31/2019.

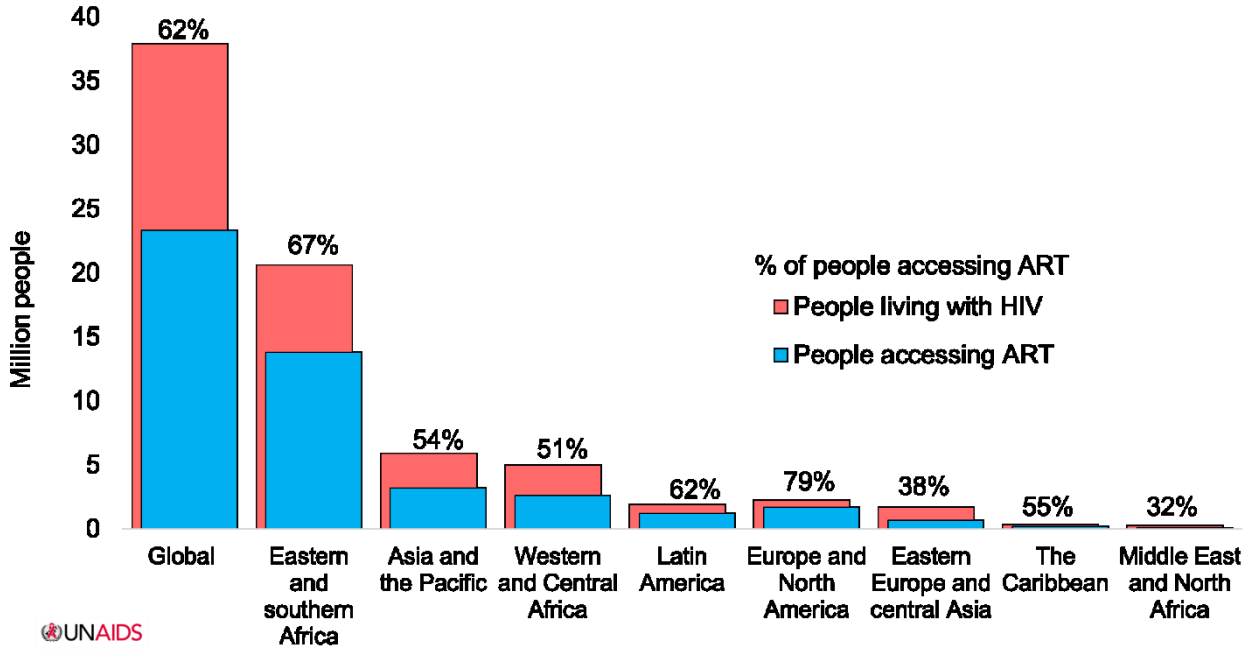


Figure 1.3: Number of people living with HIV and accessing ART for major geographical populations.
Percentage indicates number of people living with HIV that are accessing ART. Data is tabulated in table 1.1. Data adapted from UNAIDS at unaids.org accessed on 7/31/2019.

Table 1.1: Regional census of people living with HIV, accessing ART, new infections, and AIDS-related death in 2018.
Data adapted from UNAIDS at unaids.org accessed on 7/31/2019

	People living with HIV (million)	People accessing ART (million)	People living with HIV accessing ART (%)	New infections (million)	AIDS-related deaths (million)
Global	37.9	23.3	62	1.7	0.77
Eastern and southern Africa	20.6	13.8	67	0.8	0.31
Asia and the Pacific	5.9	3.2	54	0.31	0.2
Western and Central Africa	5.0	2.6	51	0.28	0.16
Latin America	1.9	1.2	62	0.1	0.035
Europe and North America	2.2	1.7	79	0.068	0.013
Eastern Europe and central Asia	1.7	0.648	38	0.15	0.038
The Caribbean	0.34	0.187	55	0.016	0.0067
Middle East and North Africa	0.24	0.0788	32	0.02	0.084

1.2 HIV Protease is an important drug target

HIV-1 protease (PR), also referred to as retropepsin or HIV proteinase, is one of the most widely studied enzymes in history. PR was first identified in 1985 when the HIV genome was sequenced[69] and quickly became an important target for inhibitors to prevent viral replication[70]. To date, the FDA has approved 9 protease inhibitors (PIs) for therapeutic use. The clinical success of PIs in combination therapy is considered a textbook example of the success of structure-based drug design[71–74]. Development of a new generation of PIs continues to innovate upon the success of current clinical drugs by improving drug potency and tolerability whilst combating drug resistant strains of PR[75].

1.2.1 HIV Protease function is essential for infectious virus

As shown in Figure 1.4, PR induces viral maturation by cleaving the Gag and Gag-Pol polyproteins into individual structural proteins and enzymes required for infectious virus. Chemical inhibition and inactivating mutations of HIV PR were first shown to prevent production of mature infectious viral particles in 1988[76–78]. Immature particles contain a spherical shell of Gag polyproteins with matrix protein (MA) bound to the lipid envelope. Capsid (CA) bridges MA and nucleocapsid protein (NC), which is bound to the viral RNA genome. Proteolytic maturation describes a process of the immature virus undergoing a morphological change into an infectious form. Mature viral particles have a condensed nucleoprotein core encapsulated by an icosahedral shaped fullerene-like shell. The shell is comprised of CA proteins hexamers with pentamers inducing curvature[79,80].

The 55 kDa Gag polyprotein consists of MA, CA, and NC structural proteins and the p6 membrane-interaction protein[81]. During translation of Gag, a frameshift occurs 5% of the time which reads through a stop codon producing a larger 160 kDa polyprotein. The Gag-Pol

polyprotein contains a MA, CA, and NC with an attached PR, RT, and IN proteins. PR embedded in Gag-Pol has intrinsic activity and is released by an autoproteolytic process to produce mature PR[82]. During maturation, the junctions of each component in the Gag and Gag-Pol polyproteins must be cleaved by the fully active PR dimer in a specific temporal and spatial processing cascade[83]. The timing of Gag and Gag-Pol processing must be coordinated with assembly and budding[84]. Furthermore, individual components must be processed and released in a specific order to allow for the structural rearrangement to occur within the small volume of the viral particle[85]. The cleavage order for the twelve different cleavage sites is regulated by the catalytic processing efficiency of PR for each site relative to one another[48]. In this way, HIV PR has been described as both a specific and non-specific enzyme. The direct involvement in HIV propagation makes PR an attractive target for antiviral inhibitors[86].

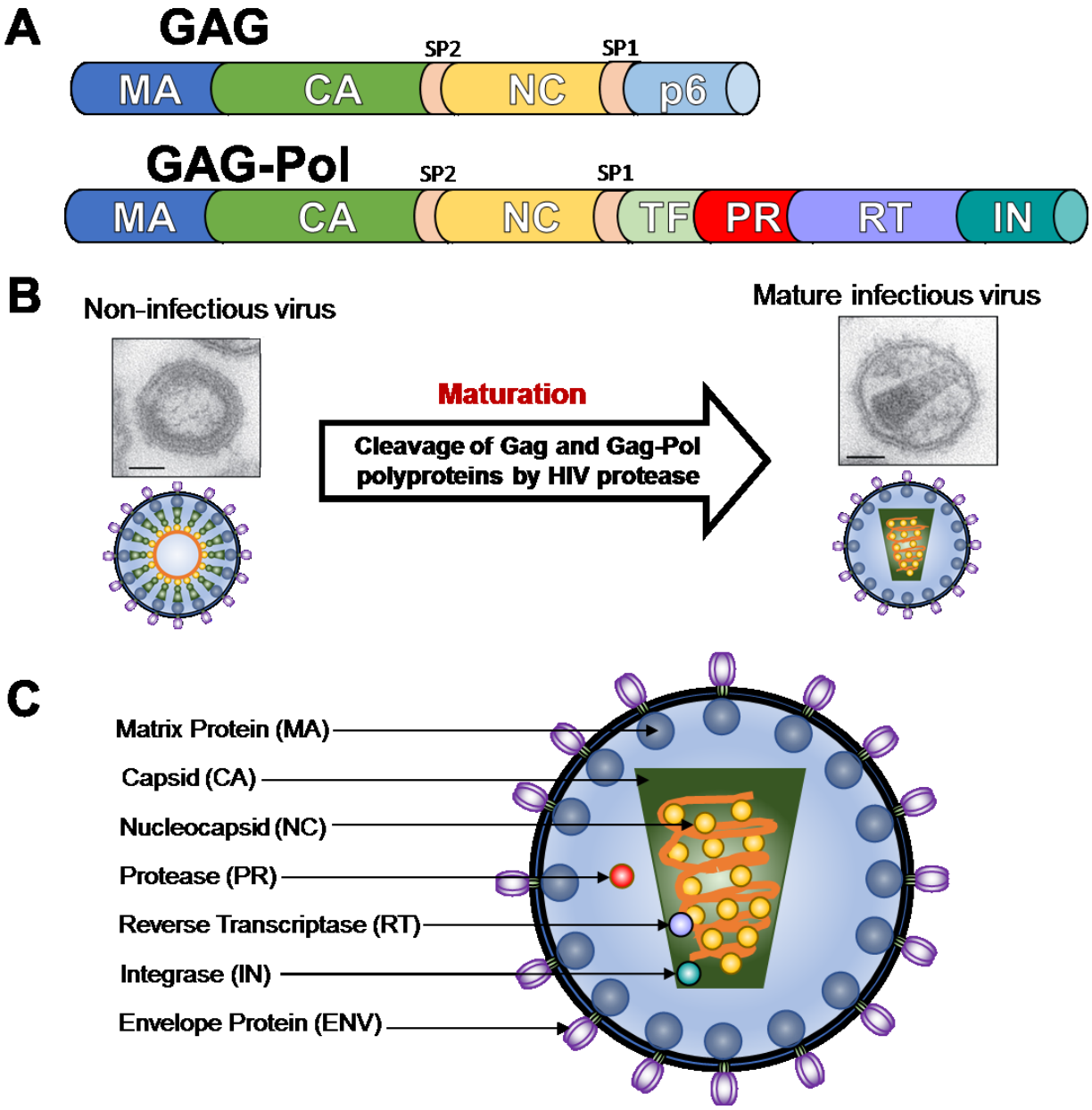


Figure 1.4: HIV maturation occurs from cleavage of Gag and Gag-Pol polyproteins by protease
A. Cartoon representation of the Gag and Gag-Pol polyproteins. PR cleaves the junctions between the individual structural and enzymatic components required for mature viral particles. sp=spacer peptide, TF=transframe protein.
B. Electron micrographs and cartoon representations of HIV particles before and after maturation. The Gag-RNA polyprotein border of in immature virus converts to an icosahedral core after maturation to infectious virus. Micrographs adapted from Konvalinka *et al*. 2015. https://doi.org/10.1016/j.virol.2015.03.021.
C. Cartoon of mature viral particle with labeled individual components. RNA genome colored orange.

1.2.2 HIV PR structure and mechanism

The tertiary structure of PR is fundamental for both its proteolytic function, its affinity for inhibitors and its propensity to evolve resistance to PIs. The structure of PR is well characterized. Over 400 X-ray crystal structures of PR and numerous mutant variants are available on the Protein Data Bank (www.rcsb.org) with others withheld for proprietary reasons. Early efforts to biochemically characterize PR were hampered by the inability to generate enough protein sample for crystallization experiments[87]. On February 9th, 1989, Rous Sarcoma Virus (RSV) protease became the first retroviral protease reported[88]. Just 7 days later, a crystal structure of HIV PR was published by a group from Merck[89]. However, much of this structure was later retracted. The first fully correct structure, consistent with earlier molecular models[90,91], was reported by Wlodawer *et al.* in August 1989[92]. Later that year, a 2.7 Å PR structure was released by Lapatto *et al.* (1989)[93]. These structures not only opened the door to structure-guided drug design of PIs but also showed that chemically synthesized protein can fold correctly in the absence of ribosomes[94]. The first structure was monomeric and ligand-free and thus limited insights into the exact mode of substrate binding. The first dimer structure of PR with a ligand, a hexapeptide inhibitor (MVT-101), was reported later in 1989[95]. Future structures with other ligands[96,97] combined with analysis of the mechanism of ligand binding[98,99] provided the foundation for rational drug-design of PIs[100,101]. Indeed, the PI Saquinavir (Ro 31-8959, Invirase) was studied by crystallography[102] and FDA approved in 1995[103].

HIV PR is an aspartic protease and its mature form has full catalytic activity as a homodimer. An illustration of PR dimer with PI is shown in figure 1.5. Two 99 amino acid subunits each contribute a catalytic triad (Asp25-Thr26-Gly27) to the active site for proteolytic function. The catalytic triad is highly conserved among aspartic proteases. The active site cavity

recognizes Gag, Gag-Pol, and Nef cleavage sites as decapeptides in subsites (S5-S5′). These sites are polymorphic[104], however the consensus sequence for each cleavage site for the reference HIV-1 genome (HXB2)[105] is shown in table 1.2. Subsites S2-S2′ are the target of competitive tight binding inhibitors currently used in the clinic.

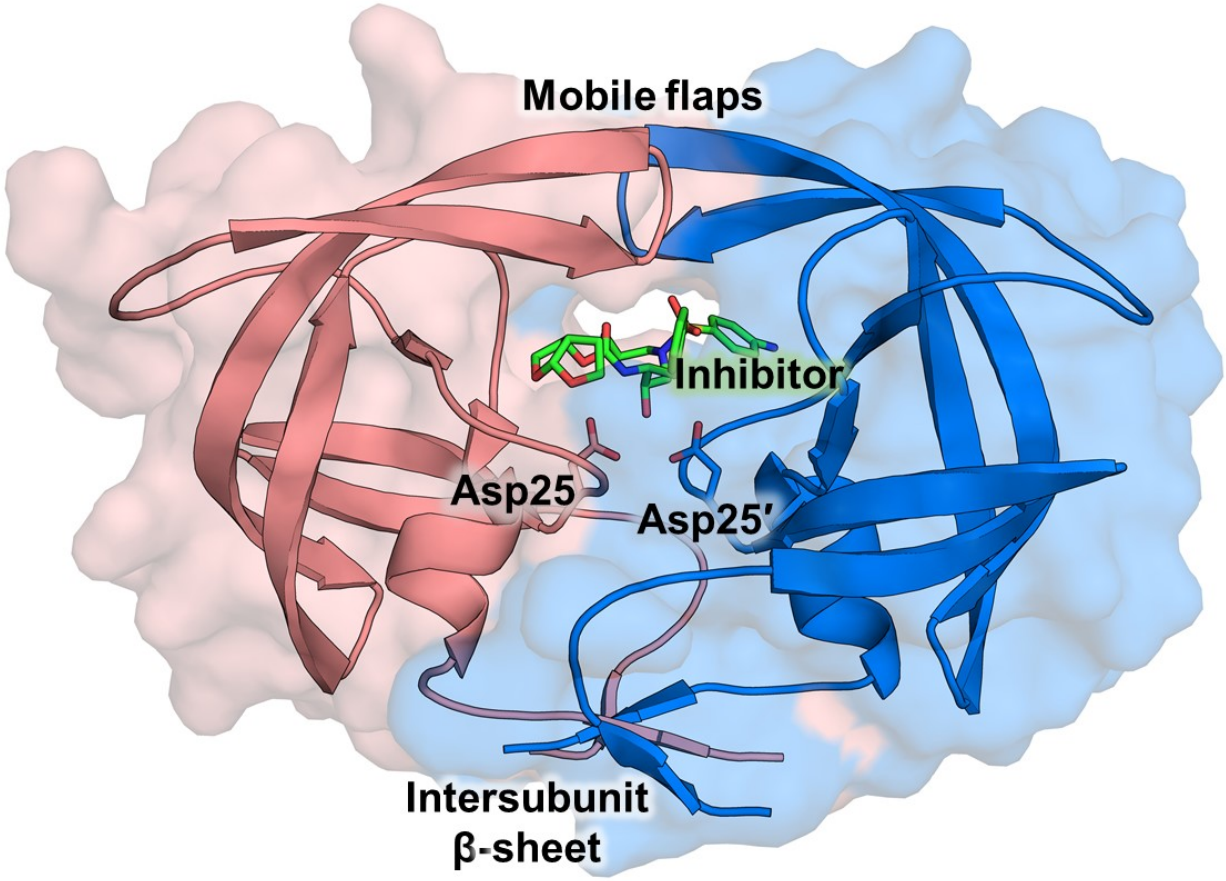
In addition to the ligand binding and catalytic sites, the PR dimer contains intermolecular features important for its quaternary structure. For dimer stability, the N and C termini of each subunit form an intercalating β-sheet of antiparallel strands. This highly conserved interface is critical for dimerization[106,107] as well as playing a role in autoprocessing[108] and has long been considered a potential drug-binding site[109–111]. Other important dimer interface regions are internal hydrophobic contacts, hydrogen bonds in the "fireman's grip" of the catalytic site[112], two salt bridges (Arg8-Asp29′ and Arg87-Asp29′)[113,114], and an interface at the tip of the flaps[115]. The latter interface is transient and only occur when the flexible flaps exist in a closed conformation, such as when PR is bound to substrate or inhibitor.

Each monomer contributes a mobile "flap" consisting of two antiparallel β-strands (residues 46-56) connected by a β-turn. The flap structures cover the active site cavity and act as a gate for entry of substrate or inhibitor. Proximal facing flap residues are important for substrate recognition and form direct interactions with inhibitors. The flexible flaps can move between an open and closed conformation[116,117], however NMR studies reveal that wild-type PR in solution tends to exist in the closed flap conformation in the absence of inhibitor[118]. Each flap pivots about the "hinge-loop" (residues 34-41) which interface with the rest of the protease. Mutations in the flap and hinge regions have major implications in drug resistance[119].

As an enzyme critical to the HIV life cycle, the structure of PR is linked to its substrate recognition and proteolytic functionality. As stated previously, PR must recognize a several

peptide substrates and cleave them at scissile bonds. Crystallography, both X-ray and neutron, has been a powerful tool to study the mechanisms of retroviral protease function by trapping different stages of the reaction[120,121] and revealing the location of critical hydrogen atoms[122–124]. The mechanism of HIV PR proteolysis was most recently reviewed by Brik and Wong in 2003[125]. Upon substrate binding to the active site, one of the catalytic Asp25 carboxylate sidechains is mono-protonated while the other is deprotonated[126]. The deprotonated aspartate acts as a general base to activate a lytic water molecule into attacking the amide-bond carbonyl-carbon of the scissile bond[127]. A metastable gem-diol reaction-intermediate forms which is stabilized the primary protonated Asp25[128–132]. The gem-diol dissociates, breaking the scissile bond, and leaving two peptide products.

Neutron structures of PR with inhibitors show that protonation of either the inner Oδ1 or the outer Oδ2 of Asp25 is dependent on the pH and chemical structure of the ligand[123,124]. Furthermore, the location of the hydrogen on Asp25 can be dependent on the protonation states of far-away residues[133]. Detailed knowledge of the PR structure, mechanism, and protonation states has provided valuable information for designing new PIs.



Wild-type HIV protease bound to protease inhibitor darunavir (PDB: 2IEN). Catalytic aspartates 25/25′ (sticks), flexible flap regions, and terminal intersubunit β-sheets are labeled. Subunits are represented as blue and salmon cartoons with protein surface. Darunavir is shown complexed in the active site cavity as green sticks.

Table 1.2: Gag and GagPol cleavage site sequences.
Peptide sequence of HIV-1 PR cleavage sites for HIV-1 PR reference genome HXB2. Sites are in genome order, not temporal order of cleavage. "|" represents the scissile bond. Adapted from Torrecilla *et al.* 2014.

Cleavage Site	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	T	S	A	I	M	|	M	Q	R	G	N
NC/p1	E	R	Q	A	N	|	F	L	G	K	I
p1/p6^{gag}	R	P	G	N	F	|	L	Q	S	R	P
NC/TFP	E	R	Q	A	N	|	F	L	R	E	N
TFP/P6^{pol}	E	D	L	A	F	|	L	Q	G	K	A
p6^{pol}/PR	V	S	F	N	F	|	P	Q	V	T	C
PR/RTp51	C	T	L	N	F	|	P	I	S	P	I
RT/RTp66	G	A	E	T	F	|	Y	V	D	G	A
RTp66/IN	I	R	K	V	L	|	F	L	D	G	I
Nef	A	A	C	A	W	|	L	E	A	Q	E

1.2.3 Structure guided-drug design of protease inhibitors

At present 9 drugs are approved by the FDA to target PR. Their chemical structures and year of introduction are provided in Figure 1.6. The first generation of inhibitors were influenced from the design of eukaryotic protease inhibitors and begun as transition-state mimics[134]. These designs were assisted by X-ray crystal structures of PR with different products[135] but have been superseded by improved designs since. Ritonavir is still used a PI pharmacokinetic booster as it also inhibits cytochrome P450.

The second generation of PIs were designed to target PRs with drug resistant mutations to the first generation of PIs. Lopinavir (LPV) was designed from Ritonavir (RTV) with the intention of minimizing influence of common drug resistance mutations at position 82[135]. Tipranavir (TPV) is the only non-peptidic PI and was discovered using high-throughput screening[136]. Darunavir (DRV) is the most recently introduced PI, reaching the market in 2006 under the tradename Prezista. DRV is nearly identical to APV but has a *bis*-tetrahydrofuran (TFH) ring in place of a single TFH. DRV boasts the best affinity for wild-type protease with a ligand dissociation constant of 5pM[137]. DRV is the only PI with a favorable enthalpy and entropy-driven binding activity[138].

PI treatments have side-effects including dyslipidemia, metabolic syndromes, and cardiovascular and cerebrovascular diseases[139]. New inhibitors are required to not only combat drug resistance but to minimize side effects. Current drug-design strategies seek to build on the successful DRV scaffold to minimize off-target effects and inhibit drug-resistant PR. Understanding the differences between toxic and benign chemical groups used in PIs aids in minimizing off-target effects in future designs[139]. Newer compounds aiming to tie down the flaps in mutant PRs by adding hydrogen bonds with the flap main chain have shown promise[140,141].

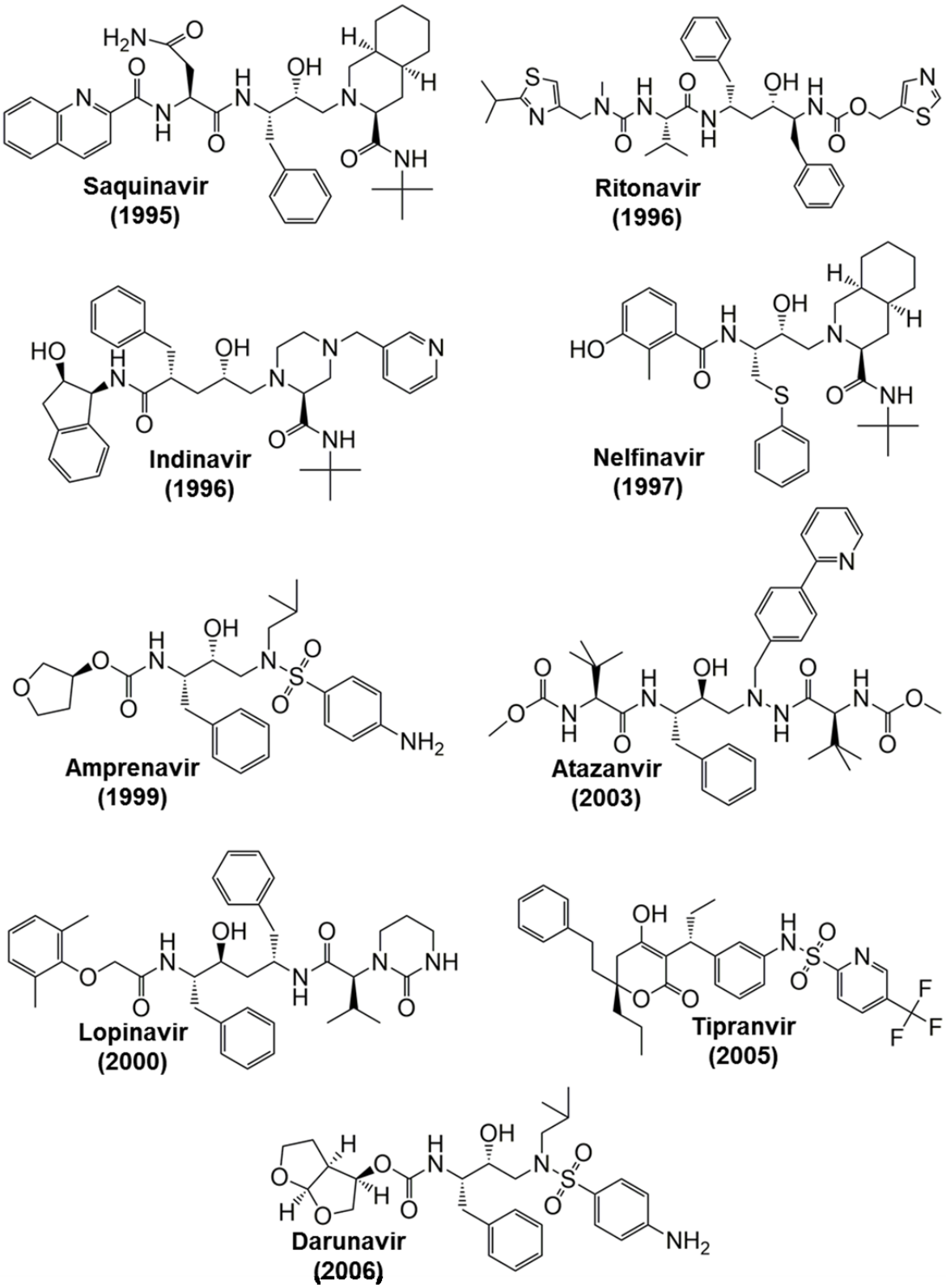


Figure 1.6: Clinical protease inhibitors of HIV-1 protease
All PIs approved for therapeutic use by the FDA to treat HIV-1 infections in descending order by year introduced. Chemdraw templates kindly provided by Yuan-Fang Wang.

1.2.4 Drug resistance in PR

HIV is highly heterogeneous in sequence thus is known as a quasispecies[142]. HIV-1 has low fidelity of replication that leads to spontaneous mutations capable of being selected upon when under drug pressure[47,143]. Low fidelity is caused by the viral reverse transcriptase (RT) lacking proofreading functionality[144]. In addition, RT naturally has low affinity for the for the RNA template causing template switches[145] leading to genetic recombination if a viral particle has non-identical RNA templates[146]. These factors contribute to a spontaneous mutation rate of $3x10^{-5}$ mutations per base pair per cycle of replication for HIV-1[147] and make RT a critical drug target[148]. Mutation rates for HIV-1 PR are 3 and 7-fold greater than for spleen necrosis virus and bovine leukemia virus respectively[147].

Drug resistance for HIV-1 PR is defined as the ability of PR to process Gag and Gag-Pol substrate in the presence of a PI *in vivo*. The high mutation rate resulting from the native features of RT are a system-wide mechanism which initiates genetic variation allowing for rapid selection under drug pressure. PR specific resistance associated mutations (RAM) for protease inhibitors undergo positive selection by additional mechanisms. PR is prone to resistance due its dynamic and flexible protein structure[149]. In addition, the range of substrate cleavage sites recognized by PR necessitates a large ligand binding site. Insertions have been observed to play a role in drug resistance[150–152]. As resistance evolves, mutations can co-evolve in the Gag and Gag-Pol precursors that can compensate for lower catalytic activity of resistant PR[153–155].

Cultural and socio-political factors also contribute to drug resistance. Complications due to adherence to the antiretroviral program has implications for drug resistance[156–158]. Adherence can be compromised by patient non-compliance due to toxicity, side-effects, and the pill-

burden[159,160]. Moreover, resistant strains can be directly transmitted from ART-treated patients to drug-naïve patients[161,162].

Drug resistance to protease inhibitors develops from mutations in the protease gene. Wensing *et al*.[119] classifies mutations as either major or minor mutations; the current report designates over 30% of the protease amino acid positions to be associated with resistance to all FDA approved PIs. Major and minor RAMs are distributed throughout the PR protein[163–165] (figure 1.7) with major mutations yielding significant loss of effectiveness to single inhibitors and minor mutations thought to provide ancillary roles to major mutations such as compensating for loss of catalytic efficiency or stabilizing a destabilized region. PRs with single substitutions or a combination of RAMs have been studied using X-ray crystallography and enzyme kinetic techniques. Selected examples from the following three mechanism categories are described below: active site residues altering drug interactions, mutations reducing dimer stability, and distal mutations with unique mechanisms.

Substitutions of residues in the ligand-binding site cavity can alter PR-PI interactions. This can be a straightforward mechanism by which fewer interactions between the target and the drug reduce binding affinity. Hydrophobic residues such as Val32, Ile47, Ile50, and Val82 can commonly mutate to different hydrophobic side chains reducing PI affinity[166]. Mutation I84V is the only mutation considered major for all clinical inhibitors[167]. Smaller side chains at position 84 are no longer able to make van der Waals interactions with inhibitors while maintaining affinity for substate[168–170]. The I47A major mutation for LPV produces two orders of magnitude worse inhibition for the inhibitor by reducing van der Waals contacts in the S2/S2′ subsites[171]. This effect was magnified by synergistic mutations V32I and I54V which alters the flap[171]. In double mutant with G48T and L89M mutations, altered side chains create an expanded ligand

binding cavity which confers resistance to SQV by a defective hydrophobic sliding mechanism[172].

Since active PR binds PIs as a dimer, mutations increasing dimer dissociation can reduce inhibition by PIs via dissociation of the dimer-PI complex. PR dimerization occurs in two steps where intermolecular interactions at the active site bring subunits together for more intermolecular forces at the dimer interface to occur[173]. DRV and TPV can inhibit this dimerization process by binding to the monomer[173,174]. Thus, mutations disrupting dimerization, such as L33I/F can alter inhibition by DRV and TPV[174,175]. Ile50 at the flap tip contributes interactions to both PIs and intermolecular attraction and thus RAMs at this position fit into both of the above categories[163].

Finally, mutations distal to the ligand binding site can induce loss of PI inhibition in diverse ways without directly contacting inhibitor. Major DRV and LPV RAM L76V has been well characterized in a series of crystallographic studies supported by molecular dynamics simulations[176–179]. A single carbon difference in the Leu76 residue shows resistance to inhibitors is conferred through lowered dimer stability, defects in auto-processing, and lost hydrophobic interactions at the flap-core interface. Distal mutation L90M is implicated in inducing small structural and dynamic changes that propagate throughout the PR to decrease PI binding[180–182]. Mutation G73S is a minor mutation associated with resistance to ATZ, APV, IDV, LPV, and SQV[119]. The serine substitution at position 73 sits along a β-strand away from the active site and initiates structural changes that propagate through residues Thr74 & Asn88 to Asp29 & Thr31 at the ligand binding site with implications for altered PI binding and substrate cleavage[183].

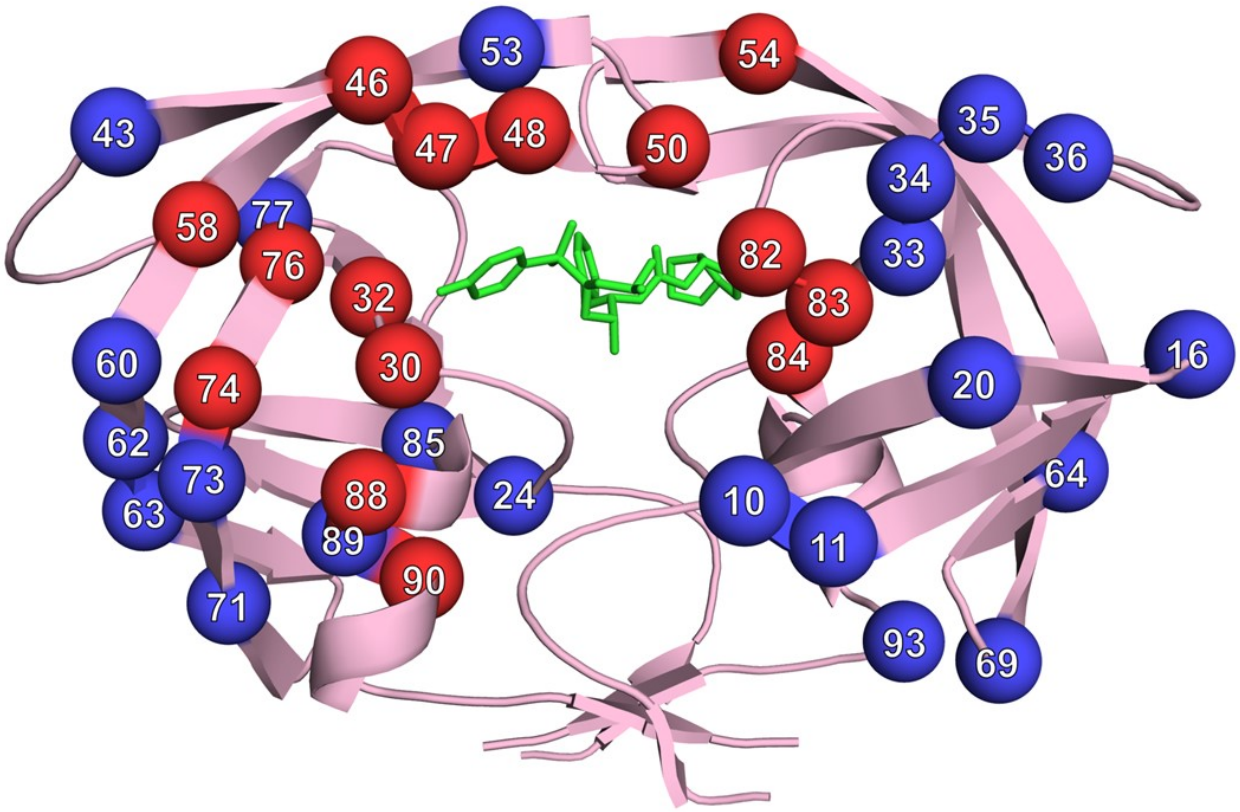


Figure 1.7: Resistance associated mutations locations in the HIV-1 PR dimer
Site of RAMs (spheres) on the PR dimer (pink cartoon) distributed across both subunits for clarity. Major and minor mutations are colored red and blue respectively. DRV is shown as green sticks. Adapted from Weber and Agniswamy (2009) and Weber, Kneller, and Wong-Sam (2015) using RAMs from Wensing *et al.*(2017).

1.2.5 Examples of highly drug resistant mutants

Years of research into PR variants with 1-5 RAMs have given valuable insight into the different mechanisms by which mutations can alter PR-PI interactions, dimer stability, and substrate recognition. More recently, this foundation of work has allowed study of PR mutants with different combinations of approximately 20 mutations. These extreme PR variants often display several orders of magnitude reduced potency to multiple clinical PIs. Such highly-resistant PR mutants share a PI-independent feature of altered flap conformations, but each employ unique and often synergistic mechanisms to evade inhibition by PIs. The following section will update and expound on an 2015 review of the subject[165]. Crystal structures of DRV-bound examples of the highly resistant mutants described here are shown in figure 1.8 and the inhibition measurements of DRV for the mutants and amino acid substitution list are in table 1.3.

An early look at multidrug resistant proteases was X-ray crystallographic studies on a clinical isolate dubbed MDR769. This PR mutant with 10 mutations, L10I, M36V, M46L, I54V, I62V, L63P, A71V, V82A, I84V, and L90M, has also been studied with an additional V82T mutant (MDR769 82T). Both MDR769 and the V82T variant display inhibition resistance to all clinical inhibitors[184]. The addition of the V82T mutation increases the resistance to DRV from 2.8 to 11-fold relative to PR[184]. MDR769 has also been proposed to be resistant to dimerization inhibitors[185]. X-ray crystal structures of MD769 show an open flap conformation, disrupted 80's loop, and perturbed S1/S1′ subsites[186,187]. The dual major RAMs V82A and I84V work in this structure to create a 3.0 Å expansion in the ligand-binding site[188]. Perturbations of the 80's loop and flaps of MDR769 observed in the crystal structures are supported by NMR data[189] and molecular dynamics simulations[190].

An exhaustive enzyme kinetics, biochemical, and X-ray crystallography study was performed on another clinical isolate with 22 mutations referred to here as PR22 (T4S, L10V, I13A, K14R, K20I, A22V, L33I, E35D, M36I, N37D, R41K, K43S, G48A, I54V, I66F, H69K, T74S, V82A, I84V, L89I, L90M and T91S)[153]. This variant shows extreme inhibition resistance (>3,600 worse than PR) for SQV, RTV, IDV, and NFV but only 11-fold relative resistance to DRV (K_i = ~0.6 nM). A crystal structure of PR22 in complex with DRV at 2.1 Å resolution (PDB: 3T3C) shows changes in the flaps, hinge region, and S1/S1′ subsites. PR22 was isolated in 2002 from a 10-year-old pediatric patient being treated with APV, LPV, and NFV yet failing ART with high viral titers and low $CD4^+$ T-cell count. Changes in treatment regimen did not reduce accumulated RAMs in PR suggesting efficient viral replication during drug selection pressure despite mutations which reduce catalytic efficiency. However, PR22 in a HIV-1 subtype B displayed 3-orders of magnitude worse infectivity in *in vivo* experiments. Analysis of the patient Gag cleavage site sequences elucidated that mutations in the patient derived cleavage sites (p2-NC, NC-p1, and p6pol-PR) rescued virus infectivity. The HIV genome encoding PR22 represents a coevolution of PR and Gag substrate to preserve viral fitness under drug-selection pressure. As of 2008, the patient was treated with DRV and showed improved $CD4^+$ T-cell count demonstrating the value of improvements to PI design to combat drug resistance mutants.

Several studies have been conducted on a clinical isolate with 20 mutations termed PR20. Isothermal Titration Calorimetry (ITC) show PR20 displays up to 8,000-fold worse binding affinity for DRV[137] and cross resistance to clinical PIs APV and SQV[191]. Reduced affinity for PIs occurs through multiple synergistic mechanisms. Similar to MDR769 described above, PR20 features an expanded ligand-binding pocket from the combination mutations I47V and I84V which reduce hydrophobic contacts with inhibitors[192]. PR20 also displays a 5-fold higher dimer

dissociation constant (K_d) compared to wild-type PR, which diminishes ability to retain inhibitors in the active site[193]. Finally, PR20 has altered flap dynamics relative to PR, as supported by several X-ray structures of PR20 with open and asymmetric conformations of the flaps[192,194] in combination with NMR studies showing open conformation in the absence of inhibitor[118], and multiple molecular dynamics simulations[194,195]. The extreme resistance exhibited by PR20 makes it an ideal prototype to assess how the next generation of investigational PIs perform against resistant protease. To date, three compounds have been discovered to inhibit PR20 proteolytic activity better than DRV (K_i for PR20 = ~40 nM). GRL-50-10A and GRL-44-10A achieve K_i values of 1.7 and 4.3 nM respectively for PR20 stemming from the formation of new hydrogen bonds with Gly48[140]. GRL-142-13A (**2**) has a K_i of 2.5 nM for PR20[196] and the details of its potency for PR20 relative to DRV are described in chapter 2.

A series of 6 multiple mutant PR variants with major RAMs for DRV (PR_{DRV1-6}) were characterized in a pair of studies[197,198]. ITC data suggests these mutants lose susceptibility to DRV due to less-favorable enthalpy of binding compared to PR. Crystal structures of three of the mutants in complex with DRV show fewer H-bonding interactions relative to PR indicative of diminished enthalpy of binding.

The previously described PR drug-resistant mutants were chosen from correlation studies[199] or broad enzyme kinetic experiments[138]. Alternative strategies for selecting mutant variants have produced mutants which have permitted new insights into PR drug resistance. Protease genotypes from databases are often from patients under selection pressure from multiple PIs during their lifetime which confound studies looking to study the accumulation of RAMs to specific PIs. In order to study the molecular mechanisms of resistance to a single inhibitor, laboratory selected PRs have been employed[200]. A laboratory-selected PR mutant with 4 RAMs

for DRV and 10 other substitutions (L10I, I15V, K20R, L24I, V32I, I33F, M36I, M46L, I54M, I63P, K70Q, V82I, I84V, L89M) was identified and termed P51. Ligand-free crystal structures of P51 show wider flap conformation compared to PR[137]. Inactive $P51_{D25N}$ crystallized in the presence of DRV displayed a remarkable mode of binding in which two inhibitor molecules are bound perpendicular to the typical orientation[201].

Another strategy for selecting PR mutants for further study is machine learning based on a unified encoding of genotype-phenotype and structural data[202–204]. This approach has predicted representative sequences for high PI-resistance for further characterization. PRS17, a clinical isolate[205] with 17 mutations (L10I, K20R, E35D, M36I, S37D, M46L, G48V, I54V, D60E, I62V, L63P, A71V, I72V, V77I, V82S, L90M, I93L) predicted to be resistant to 6 inhibitors is the best characterized. As predicted, all clinical PIs with the exception of TPV (which was not tested) show reduced potency for PRS17[206]. DRV has 10,000-fold worse affinity for PRS17 (K_L = 50 nM) despite the absence of any RAMs for DRV[167]. Like PR20, NMR experiments show PRS17 also exists with an open flap conformation in solution in the absence of inhibitor, in contrast to PR which exists in the closed formation suggesting altered flaps as a mechanism for resitance[118]. Two X-ray crystal structures of ligand-free PRS17 show the flaps form a curled phenotype relative to PR[118,207]. The PRS17/DRV crystal structure reveals weaker binding of DRV does not occur through lost active site contacts. Furthermore, reduced peptide-analogs of the p2-NC and CA-p2 cleavage sites were measured to inhibit PRS17 3-fold better than wild-type[207]. Crystal structures of PRS17 in complex with p2-NC and CA-p2 elucidating PRS17's enhanced substrate recognition are described in detail in chapter 3. Insights gleaned from studies on PRS17 demonstrate the potential of this machine learning strategy for predicting representative sequences. Another mutant produced from this approach, PRS5B, is characterized in chapter 4.

Highly-resistant HIV PRs often share altered flap conformational flexibility as a common mechanism for evading inhibition by PIs. The mutants described here use synergistic mechanisms to different degrees which lead to a range of magnitudes for reduced inhibition by PIs. All mutants in this list contained the common I84V mutation except PRS17. PR20, MDR769, and PR22 combine I84V with other smaller hydrophobic mutations to reduce van der Waals contacts with PIs through an expanded ligand-binding site. PRS17 and PR22 evolved greater recognition of substrates while PR20 possess a more stable monomer fold. Understanding of the molecular consequences of drug-resistance mutations is vital to designing new PIs against resistant PRs. Structural and biochemical characterizations of highly-resistant mutant exemplars provide valuable insight into the molecular mechanisms of resistance and prototypes to test investigational inhibiters.

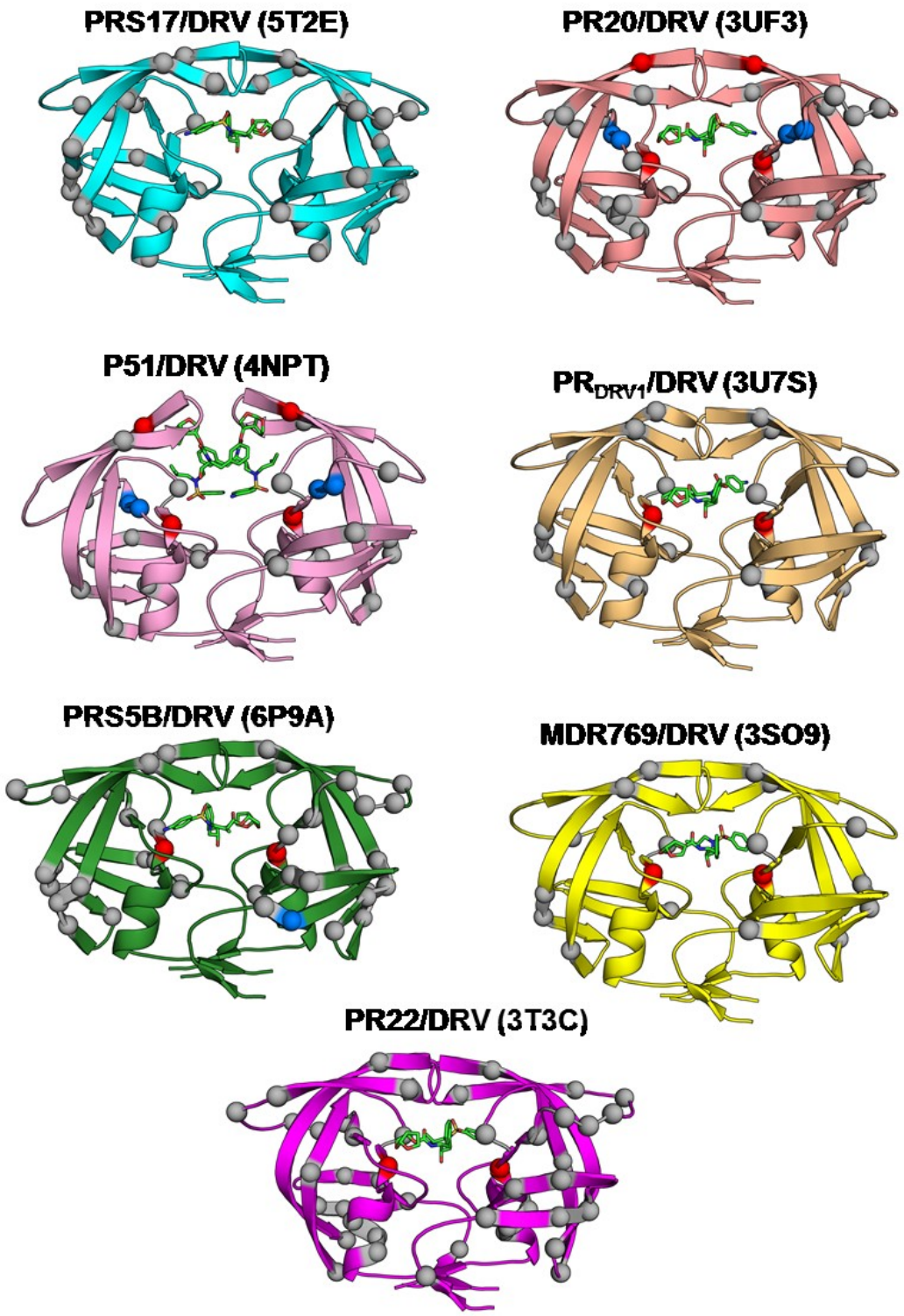


Figure 1.8: X-ray crystal structures of highly-resistant PR mutants in complex with DRV Cα of substituted amino acids are shown as grey spheres. Major and minor mutations for DRV are colored in red and blue respectively. DRV is shown as green sticks.

Table 1.3: Highly-resistant PR mutants and inhibition values for DRV.
Mutants are in descending order by relative resistance to DRV. Major and minor mutations for DRV are colored in red and blue respectively. Inhibition measurements are from ITC and kinetic measurements under different conditions and may not be directly comparable. As defined by Wensing *et al.* (2019), major mutations for DRV are I47V, I50V, I54M/L, L76V, & I84V. Minor mutations for DRV are defined as V11I, V32I, I33F, T74P, & I89V.

Protease (Color in fig. 1.8)	**K_i or Kd (nM)**	**K_d or K_i relative to wild-type**	**Mutations**
Wild-Type PR	0.005	1	
PRS17 (Cyan)	50	10,000	L10I, K20R, E35D, M36I, S37D, M46L, G48V, I54V, D60E, I62V, L63P, A71V, I72V, V77I, V82S, L90M, I93L
PR20 (Salmon)	41	8,200	L10F, I13V, I15V, D30N, **V32I**, **L33F**, E35D, M36I, S37N, I47V, **I54L**, Q58E, I62V, L63P, A71V, **I84V**, N88D, L89T, L90M
P51 (Light orange)	35	7,400	L10I, I15V, K20R, L24I, **V32I**, **I33F**, M36I, M46L, **I54M**, I63P, K70Q, V82I, **I84V**, L89M
PR_{DRV1} (Pink)	15	3,000	L10I, M36V, M46L, I54V, I62V, L63P, A71V, V82A, **I84V**, L90M
PRS5B (Green)	4	800	L10I, **V11I**, E21D, A22V, L24M, I33L, E35N, M36I, S37D, R41K, M46L, I54V, Q61H, I62V, I63P, I64V, I66V, A71V, I72T, G73T, N83D, **I84V**
MDR769 (Yellow)	1.1	220	L10I, M36V, M46L, I54V, I62V, L63P, A71V, V82A, **I84V**, and L90M
PR22 (Magenta)	0.059	11	T4S, L10V, I13A, K14R, K20I, A22V, L33I, E35D, M36I, N37D, R41K, K43S, G48A, I54V, I66F, H69K, T74S, V82A, **I84V**, L89I, L90M and T91S

1.3 X-ray crystallography and structural biology

1.3.1 Crystallography for biology – a brief history

German astronomer and mathematician Johannes Kepler first described crystals in a scientific context in 1611 in his Latin essay *A New Year's Gift of Hexagonal Snow, or On the six-cornered Snowflake*[208]. X-rays also known as Röntgen radiation were defined in 1896 when German physicist Wilhelm Röntgen published a radiograph of his wife's hand in *Science*[209]. It was not until the 1915 Nobel Prize in Physics awarded work by William Henry Bragg and Lawrence Bragg that the potential of X-ray diffraction from crystals was fully realized[210]. The field and size of molecules studied grew rapidly from solving the structure of NaCl[211] and diamond[212] in 1914 to the first small molecule, hexamethylenetetramine, in 1923[213]. Sperm whale myoglobin was the first protein structure solved by X-ray crystallography in 1958[214]. British chemist Dorothy Crowfoot Hodgkin pioneered X-ray crystallography for protein and biological molecules by solving the structure of cholesterol, penicillin, vitamin B_{12}[215], and insulin[216] earning her the Nobel Prize in Chemistry in 1964. Perhaps the most famous discovery from X-ray crystallography is the understanding of the double helical structure of DNA in 1953[217] when James Watson and Francis Crick obtained Rosalind Franklin's X-ray diffraction image of DNA from Maurice Wilkins.

At present, high resolution macromolecular X-ray crystallography is practiced at synchrotron radiation beamlines[218]. Approximately 140,000 structures of protein, nucleic acids, and protein-nucleic acid complexes have been determined by X-ray diffraction methods to date[219]. The frontier challenges of macromolecular crystallography research are being met by subatomic resolution with X-rays and direct hydrogen atom determination using neutron

diffraction[220]. To meet the need for high quality protein crystals, researchers and astronauts are even growing protein crystals in microgravity on the international space station[221].

1.3.2 X-Ray crystallography

An overview of a typical X-ray crystallography experiment is presented in figure 1.9. Protein crystals for X-ray crystallography are commonly grown using vapor diffusion techniques[222]. In a classic vapor diffusion crystallization experiment, purified protein is mixed with buffer containing precipitating salts and allowed to equilibrate in a system containing a reservoir of the same buffer at higher concentration[223]. As water molecules diffuse from the protein-buffer drop, the protein and precipitant concentrations in the drop gradually increase. Protein molecules may begin to aggregate in a crystalline form if the precipitant, temperature, pH, additive, and temporal conditions are correct[224]. Protein crystals are generally cryo-preserved in liquid nitrogen prior to shipping to synchrotron beamlines to reduce data noise, radiation damage, and transportation-related trauma.

For X-ray diffraction data collection, crystals are rotated about an axis using a goniometer while subjected to an X-ray beam wherein electrons from atoms in the molecule scatter the beam. Large ordered array of protein molecules from the crystalline form serve to amplify the signal intensity. Image sensors collect the X-ray diffraction pattern to be analyzed. The reflection pattern is a reciprocal lattice of the molecule crystal lattice and must be converted to structure factors, the amplitude and phase of the diffracted X-ray waves, in order to be interpreted visually as electron density maps. Amplitude information is gathered by the intensity of each spot in the pattern. Information on the phase of the X-ray wave for diffracted spots is determined by different means. The diffracted spots are related to the electron density of the crystal unit cell by a Fourier transform[225]. If the structure of a similar protein has already been

determined, the “phase problem” may be solved by an approach called molecular replacement[226]. Structure factors from the three-dimensional features of the known molecule are calculated for rotations and translations. Structure factors in agreement with the experimental diffraction data allow orientation of the molecule the unit cell to be estimated. This orientation is then tested for different positions in the unit cell. Orientation and translation models which agree with the experimental data are used to calculate phases. Other methods to determine phases for macromolecular diffraction patterns are multiple isomorphous replacement[227] and multiwavelength anomalous dispersion[228]. From structure factor information, electron density maps may be calculated and an atomic model interpreted[229].

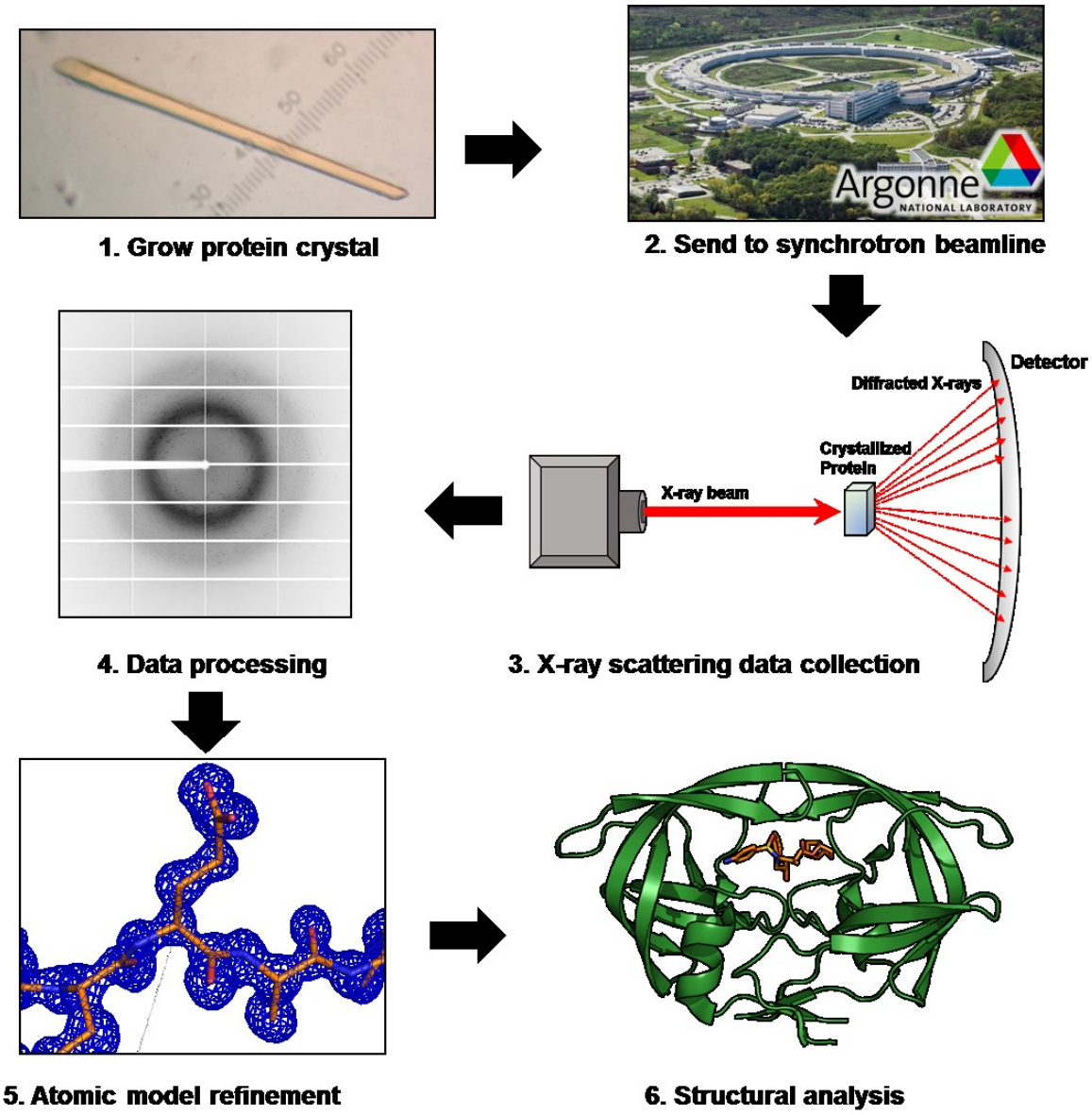


Figure 1.9: Overview of the protein X-ray crystallography process
In a typical X-ray crystallography diffraction study, protein crystals are grown, cryoprotected and sent to an X-ray source such as a synchrotron. X-ray scattering from the crystal is collected and processed to produce an electron density map (blue mesh in 5). The atomic model is built and refined to produce a final structure used for biological analysis.

1.3.3 Complimentary protein structural biology techniques

Single crystal X-ray diffraction is powerful and vetted technique to reveal the three-dimensional structure of proteins but possesses inherent limitations. The technique requires crystals suitable for diffraction, which are challenging to grow for intrinsically disordered proteins[230] and large proteins with transmembrane domains[231,232]. Protein preparation and the crystallization process can introduce artifacts that may not represent the physiologically relevant conformation of the native protein in solution[233]. Furthermore, low X-ray scattering power of hydrogens necessitates that hydrogen positions must be inferred or estimated. Luckily, these limitations can be countered using complimentary structural biology techniques.

Neutron crystallography is the forefront approach for determination of hydrogen position in protein structures. Neutron scattering is similar to X-ray crystallography and the complimentary results of the two diffraction experiments on the same crystal are often combined[234,235]. Neutrons scatter from atomic nuclei and permit room temperature data collection without destroying the crystal or producing artifacts associated with X-ray scattering. Hydrogen nuclei and heavy atoms have similar neutron scattering length, rendering hydrogens as visible as carbon, nitrogen, and oxygen. The caveat is that neutrons scatter with high inelastic scattering from protium nuclei giving high background. Perdeuteration or hydrogen-deuterium isotope exchange of the protein crystal sample can improve signal-noise ratios because deuterium offers the same scattering length with lower incoherent scattering than protium[236].

With clear advantages over X-ray crystallography, neutron diffraction has key limitations to overcome before becoming widespread. Chiefly, access to neutron source resources for crystallography is scarce with less than 25 neutron sources on Earth and even fewer single-crystal neutron diffractometers. In addition to lower resolution data sets compared to X-rays,

large crystals (>1-10 mm^3) and long data collection time requirements bottle-neck neutron structures, although higher flux neutron beams and improved detection instruments are addressing these concerns[237]. Combining the high-resolution heavy atom determination from X-ray crystallography with the hydrogen sensitivity of neutron diffraction offers a complete picture of atomic positions within a protein crystal.

Small angle X-ray scattering (SAXS) and small angle neutron scattering (SANS) are diffraction techniques which provide low resolution structural information for a protein in solution. Small angle scattering experiments record X-ray or neutrons scattered from protein in solution at angles less than 10º. These results allow interpretation of a protein envelope in solution without a crystallization[238]. SAXS and SANS yield information for protein size, shape, and surface-volume ratio to be determined[239,240] and are useful for validating complimentary structural biology technqiues[241]. Small angle scattering requires extremely pure protein samples at greater concentrations than for crystallization experiments[242].

Solution NMR spectroscopy is another valuable technique for studying protein conformations in solution[243]. Protein NMR takes advantage of changing atomic nuclei spin states for isotopes of biological important nuclei when a magnetic field is applied[244]. Information gathered in NMR spectroscopy experiments show an ensemble of conformations of proteins in solution and dynamic motion. NMR is useful for small proteins (<50 kDa) that resist aggregation at high concentration[244].

Structural techniques discussed here are extremely useful for single proteins such as enzymes, yet many of the most complex biological systems involve large multi-protein complexes. Where crystallography falls short in this regard, the rapidly advancing field of Cryogenic electron microscopy (Cryo-EM) permits structure information for large complexes in

solution. For example, cryo-EM tomographic information has provided never-before-seen details of the envelope of HIV[245], Zika[246], and dengue virus[247]. Improvement in image detection and dedicated Cryo-EM analysis packages will prove useful for determination of Cryo-EM structures in the future[248].

Structural biology techniques have been facilitated by advances in computer science. Molecular dynamics (MD) simulations of protein structures sit at the axis of the two disciplines to enable new understanding of protein structure data. MD is used to simulate the physical interactions of atoms in a system to predict the trajectories of those atoms over a time period[249]. Molecular mechanics are applied to all atoms a system to build molecules proteins, ligands, and solvent. From MD simulations, one can generate predictions of ligand binding energy, local and global protein dynamics, and electrostatic interactions[250,251].

The platform of complimentary techniques used in structural biology is employed for characterizing the three-dimensional properties of biomolecules. Evidence of the field's utility can be appreciated in contributions to areas such as drug-design[252], synthetic biotechnology[253], bioremediation[254], and oncology[255]. The protein structure-function information gathered by structural biology techniques are paramount to understanding how biomolecules are created, converted, and can be manipulated to human benefit.

1.4 Overview of Aims

The HIV/AIDS pandemic is nearing 40 million infected individuals. The viral PR, essential for virion maturation and infectivity, is a key target for competitive inhibitors approved by the FDA to treat HIV infection. PR evolves RAMs over time in PI-treated patients which threaten viral suppression status, necessitating continued development of treatment options. Understanding protease-inhibitor interactions in examples of HIV resistant mutants aid design of

the next generation of PIs. The overarching hypothesis is that the primary contributor to PI resistance is the structural and dynamic modifications that result from mutations in HIV-1 PR. In this body of work, X-ray crystallography and enzyme kinetic techniques were used to elucidate the molecular basis of PI-resistance for three highly resistant PR mutants (PR20, PRS17, and PRS5B) to assist rational design of new inhibitors.

1.4.1 Aim 1

To compare the effectiveness of investigational protease inhibitor GRL-142 with clinical inhibitor darunavir for inhibiting the drug-resistant HIV-1 protease mutant PR20. Enzyme kinetic inhibition measurements indicate GRL-142 inhibits PR20 better than does DRV. A high-resolution crystal structure of PR20 complexed with GRL-142 compared to PR20/DRV suggests the improved inhibition stems from greater favorable interactions with PR20. GRL-142 represents a successful case of structure-guided drug design to combat resistant PRs.

1.4.2 Aim 2

To investigate the molecular mechanisms of substrate binding to drug-resistant HIV-1 protease mutant PRS17. Reduced peptide analogs of the Gag p2-NC and CA-p2 cleavage sites show 3-fold better inhibition of resistant mutant PRS17 proteolytic activity than for wild-type PR. Crystal structures of PRS17 bound to p2-NC and CA-p2 reveal that enhanced substrate recognition is driven primarily by active site mutations G48V and V82S altering PRS17-Substrate interactions. This study highlights a dual-resistance mechanism of G48V and V82S to evade interactions with PIs and enhance substrate binding.

1.4.3 Aim 3

To characterize the enzyme kinetics, inhibition, and inhibitor-bound X-ray crystal structures of drug-resistant HIV-1 PR mutant PRS5B. Enzyme kinetics data confirm

resistance data in HIVdb by showing that a PR mutant, PRS5B, exhibits intermediate levels of inhibition by clinical PIs. A urea denaturation assay indicates PRS5B, PRS17, and PR20 have comparable catalytic activity to wild-type in the presence of a chaotropic agent. Crystal structures of PRS5B in complex with APV and DRV reveal clusters of mutations in distal regions act synergistically to decrease intra-subunit interactions with respect to equivalent wild-type PR structures. Molecular dynamics simulations of ligand-free enzymes show more dynamic flaps and conformational variability for PRS5B than for PR, consistent with increased dissociation of inhibitors. PRS5B shares common traits with other resistant mutants but also possesses mutations with opposing effects suggesting it has not evolved to produce the high levels of resistance seen for other mutants.

2 POTENT ANTIVIRAL HIV-1 PROTEASE INHIBITOR COMBATS HIGHLY DRUG RESISTANT MUTANT PR20

Kneller DW, Agniswamy J, Ghosh AK, Weber IT (2019) Potent antiviral HIV-1 protease inhibitor combats highly-drug resistant mutant PR20. *Biochemical and Biophysical Research Communications.* (519)61-66.

2.1 Abstract

Drug-resistance threatens effective treatment of HIV/AIDS. Clinical inhibitors are ineffective for highly resistant protease mutant PR20, however, antiviral compound **2** with fused tricyclic group at P2, extended amino-benzothiazole P2' ligand and two fluorine atoms on P1 shows 16-fold better inhibition than darunavir of PR20 enzyme activity. Crystal structures of PR20 and wild-type PR complexes reveal how the extra groups of **2** counteract the expanded ligand-binding pocket, dynamic flaps, and faster dimer dissociation of PR20.

2.2 Introduction

The HIV/AIDS pandemic comprises about 38 million infected individuals with almost 60% receiving antiretroviral therapy[56] and showing decreased mortality[5]. This success is undermined by treatment failure due to drug resistance and persistent reservoirs of latent virus[256].

Antiviral HIV-1 protease (PR) inhibitors (PIs) are effective for therapy, and darunavir (DRV, **1**) (Figure 1) is approved for first line treatments[257,258]. Its favorable properties include picomolar affinity for PR, high genetic barrier to resistance[259,260], inhibition of precursor autoprocessing[137,261] and inhibition of PR dimerization[262]. Nevertheless, DRV-resistant viral strains pose a problem[263]. Resistance to PIs evolves primarily by mutations in PR. Mutations of >35 of the 99 residues in PR are associated with resistance to one or more PIs[164]. Highly-

resistant PR variants in clinical isolates have different combinations of about 20 mutations and diminish the effectiveness of PIs by a variety of molecular mechanisms[118,153,165,187,206].

An HIV-1 PR variant with 20 mutations derived from a clinical isolate (PR20) has been characterized and used as a prototype to assess investigational PIs[138,192]. Compared to wild-type PR, PR20 shows drastically reduced inhibition by clinical PIs with over 8,000-fold worse binding affinity for DRV[137]. Moreover, autoprocessing of its precursor is unaffected by **1** unlike the wild-type precursor[137]. Structures of PR20 reveal an expanded ligand-binding cavity, which contributes to decreased affinity for PIs[192] and fewer interactions with substrate analogs. Crystallography and NMR spectroscopy experiments suggest a PI-independent mechanism for poor PI affinity due to increased mobility of the two flexible flap regions[192,194,195,264]. Finally, PR20 has a higher rate of dimer dissociation than wild-type enzyme, which further weakens drug binding[193]. PR20 contains four mutations in the ligand binding site (D30N, V32I, I47V, I84V) which act in the S1/S1′ and S2/S2′ pockets to decrease hydrophobic interactions with inhibitors[191,192]. The dimer structure of PR20/**1** and key mutations are shown in Figure 1B. PR20 provides an excellent prototype to assess the effectiveness of new inhibitors designed to combat highly resistant variants[140,191].

Potent new PIs are urgently required to target resistant variants such as PR20, but must also feature improved genetic barrier to resistance and better CNS penetration[265] for effective treatment. Recently, antiviral compound GRL-142 (**2**) was reported to show picomolar inhibition of PR activity, high barrier to resistance, and improved penetration of the blood-brain barrier compared to **1**[266,267]. Compound **2** was designed with modifications of three moieties of DRV (Figure 1A) to target drug-resistant variants[268]: P2 *bis*-tetrahydropyranofuran (*bis*-THF) is replaced by a "crown-like" *Crn*-THF; P2′-aniline is substituted by a cyclopropyl-amino-

benzothiazole (Cp-Abt) group, and two fluorine atoms modify the P1-phenyl group to create *bis*-fluoro-benzyl (*bis*-FBz). Fluorination of PIs has been shown to introduce halogen-bonds with protease[140,269] and increase lipophilicity[141,270].

Here we report that **2** is an order of magnitude better than **1** in inhibiting enzyme activity of drug-resistant PR20. The crystal structure of PR20/**2** complex reveals how the inhibitor effectively fills the expanded ligand binding site and stabilizes the dynamic flaps and dimer interface.

2.3 Methods

2.3.1 PR20 expression and purification

A synthetic 99 amino acid gene derived from the clinical isolate for PR20[138] was expressed in E. coli, purified, and folded as described in [192].

2.3.2 Kinetic inhibition assay

Compound **2** (>95% purity by HPLC) provided by Dr. Arun Ghosh at Purdue University was dissolved in 100% DMSO. The kinetic inhibition value (K_i) of **2** for PR20 was measured using a FRET-substrate (BACHEM H-2992) at 37°C and pH 5.6 based spectroscopic assay as described in [207].

2.3.3 X-ray crystallography

PR20 at 4 mg/mL was mixed with **2** at 1:5 molar ratio. Crystals were grown using hanging-drop vapor diffusion in 1.8 M sodium chloride, 0.1 M sodium acetate pH 5.5, 0.1 M yttrium chloride, and cryo-protected in mother liquor containing 30% glycerol prior to freezing with liquid nitrogen. Diffraction data were collected on the SER-CAT 22ID beamline at the Advanced Photon Source, Argonne National Laboratory (Argonne, IL). X-ray data were processed and scaled using HKL-2000[271] before solving the structure by molecular replacement

in phaser[272] from CCP4[273] with PR20/GRL-5010A (4YHQ)[140] as initial model. The structure was refined using COOT[274] and REFMAC5[275] applying anisotropic B-factors. Hydrogen bonds (2.4-3.5 Å) and hydrophobic contacts (3.6-4.2 Å) were inferred from interatomic distances and chemistry. Figures were generated using PyMOL[276]. Coordinates and structure factors for PR20/**2** have been deposited in the Protein Data Bank with accession code 6PRF.

2.4 Results

2.4.1 Compound 2 is more effective than DRV for PR20

Investigative compound **2** has an enzyme inhibition constant (K_i) of 14 pM for wild-type PR[266], comparable to the value of 5-10 pM for the best clinical inhibitor, DRV[277,278]. **2** exhibits a K_i of 2.5 ± 0.5 nM for PR20, approximately 1200-fold worse than for PR. However, this K_i for PR20 is 16-fold better than the 40 nM value observed for DRV[137].

2.4.2 Overall structure of PR20/2

The crystal structure of the PR20 complex with **2** was refined to a R-factor of 15.2% at 1.21 Å resolution, the highest resolution to date for PR20 (Table 1). The asymmetric unit has a homodimer of PR20 (residues 1-99 and 1′-99′). Inhibitor **2** binds in the active site in two orientations with 0.5/0.5 relative occupancy related by ~180° rotation. An Fo-Fc omit map for one conformation of **2** is shown in Figure 2A. Residues 45-47, 50-51, 44′-48′, and 50′-52′ in the flap region, Asn30/30′ and Arg8/8′ also show two alternate conformations, each associated with one inhibitor orientation. The two alternate conformations show no significant differences in PR20 interactions with **2**. Inhibitor-binding interactions in PR20/**2** were compared with those in PR/**2** and PR20/**1** complexes.

2.4.3 Mutant PR20/2 compared to wild-type PR/2

The PR20/**2** structure was compared to two reported structures of wild-type PR/**2**. The equivalent 198 Cα atoms of the two PR/**2** (6BZ2, 5TYS) dimers[266,267] superpose onto PR20/**2** with Root mean square deviations (RMSDs) of 0.63 and 0.68 Å, respectively. The two PR/**2** dimers are nearly identical (RMSD = 0.48 Å) with two conformations for **2** of almost equal occupancies. For clarity, the following analysis will describe significant differences in one **2** conformation in comparison to the major occupancy conformation of the higher resolution PR/**2** (6BZ2 at 1.67 Å) structure.

Polar interactions between PR20 and **2** are shown in Figure 2B. The wild-type PR/**2** complex has almost identical hydrogen bond and halogen interactions except for interactions with the mutated residue D30N. Polar and non-polar interactions are described separately for P2 and P2' groups.

The bulky P2-*Crn*-THF on **2** binds in the S2 pocket (Fig 3A), where the two oxygen atoms form three hydrogen bonds with main-chain amides of residues 29 and 30 in both PR and PR20. The D30N mutation in PR20 substitutes a carboxylic acid for a carboxamide side-chain which enables formation of a 3.5 Å hydrogen bond with **2** which cannot occur for Asp30 in the PR structures. The P2-*Crn*-THF packs into the hydrophobic pocket of the S2 site forming 3 van der Waals contacts with Ile47 in PR. The S2 pocket of PR20 contains the drug-resistance mutations I47V and V32I. Mutation I47V produces one less hydrophobic contact, while the longer isoleucine side-chain in V32I introduces 3 new van der Waals contacts with **2** that are absent in PR. This observation is consistent with molecular dynamics simulations showing improved van der Waals interaction energy between a V32I single mutant and **2**[267]. Therefore, the three drug-resistance mutations in the S2 pocket of PR20 result in a net increase of one

hydrogen bond and 2 hydrophobic interactions with the P2 group of **2** in comparison to those in the wild-type complex PR/**2**.

In the opposite protease subunit, the P2′-Cp-Abt group of **2** binds in the S2′ pocket (Fig 3B). In the PR/**2** structure, the side-chain carboxylate of Asp30′ forms a bifurcated hydrogen bond with the two nitrogen atoms of Cp-Abt. In PR20, the side-chain of the mutated D30′N is flipped 180° relative to its conformation in the other subunit. This orientation of Asn30′ forms a hydrogen bond with the P2′ amine of **2**, while also forming a hydrogen bond with the aspartate side-chain of N88′D in PR20. The B-value for Asn30′ side-chain in PR20/**2** is approximately half that of Asp30′ in PR/**2** and PR20/**1** (16 vs 33 and 28 $Å^2$), suggesting a more stable conformation. The main-chain of Val47′ in the PR20 flap is shifted by 0.5 Å relative to its position in PR, eliminating a hydrophobic interaction with the thiazole sulfur of **2**. However, the cyclopropyl moiety on **2** is shifted 1.5 Å towards the I47′V mutation, which introduces two hydrophobic contacts. Like in the S2 pocket, the larger V32′I mutation introduces a hydrophobic contact with **2**. Thus, PR20 shows favorable polar and hydrophobic interactions with the P2′ group of **2** despite multiple resistance mutations in the S2′ pocket.

In addition to the changes in the S2 and S2’ pockets of PR20, the shorter side-chain of I84V mutation in the S1/S1′ pockets accounts for a net loss of 4 hydrophobic contacts with **2** compared to those in the wild-type complex. These altered interactions due to mutations in the active site of PR20 are expected to enhance affinity for inhibitor **2**.

2.4.4 Comparison of PR20 interactions with inhibitors 1 and 2

The crystal structure of PR20/**2** was compared to PR20/**1** (PDB: 3UCB) to understand the different inhibition values. The bulky P2 *Crn*-THF of **2** forms 5 hydrophobic contacts with mutations I47V and V32I that cannot occur for *bis*-THF of **1** in the PR20/**1** structure, although

the hydrogen bonds are conserved (Figure 4A). The P2′-Cp-Abt group on **2** is also larger than the P2′-aniline of **1**. Both inhibitors show similar interactions with I47′V and V32′I side-chains, however, the extended P2′ group of **2** introduces 10 more hydrophobic contacts with the side-chains of Asp29′ and Asn30′ (Figure 4B). The thiazole ring of **2** forms a hydrogen bond with the Asn30′ amide, while the aniline amide of **1** interacts with the Asn30′ carbonyl. Moreover, the P2′ amine in **2** introduces a direct hydrogen bond with Asn30′ side-chain replacing a less favorable indirect water-mediated interaction in **1** (Figure 4C).

Difluorination of the P1-group of **2** enhances its penetration of the cell membrane and introduces halogen bonds with wild-type PR[267]. The P1 *bis*-FBz forms fluorine bonds with the Arg8′ side-chain and the main-chain amide of Ile50 at the flap tip. These halogen bonds bridge the S1 pocket and the dimer interface (Figure 4D). Arg8/8′ forms a critical intersubunit ion-pair with Asp29′/29 in most PR dimers[279]. In PR20/**1**, the larger side-chain of L10F mutation induces rotation of the side-chains of Arg8 and Arg8′, thus eliminating the ion-pair with Asp29′/29, which has been proposed to increase dissociation of PR20 dimers[192]. This rotated conformation of Arg8 and loss of its intersubunit ion-pair is observed in the S1′ pocket of both PR20/**1** and PR20/**2** structures where the identical P1′-isobutyl group is ~6 Å away from Arg8 (Figure 4E). However, in the S1 pocket of PR20/**2**, the fluorine bond between **2** and Arg8′ acts to overcome the effect of the L10′F mutation.

The P1 *bis*-FBz group of **2** introduces new hydrophobic contacts with PR20 that cannot occur for the smaller P1 of **1**. The fluorine atoms make additional contacts with residue Ile50 at the flap tip, which may decrease the mobility of the flaps and stabilize the binding of inhibitor **2** within the expanded active site cavity of PR20.

Overall, larger P2 and P2′ moieties for **2** relative to **1** introduce more van der Waals contacts with the expanded ligand binding pocket that contribute to better inhibition of PR20, although **2** is unable to make contacts with the shorter side-chain of I84V mutation. These observations in PR20 are consistent with molecular dynamic calculations comparing van der Waals interaction energies between **1** and **2** in PR[267]. Compared to DRV, compound **2** can form a larger number of direct polar interactions with PR20. A fluorine bond between the P1 of **2** and Arg8′ restores an inter-subunit ion-pair and may counteract the higher rate of dimer dissociation of PR20. The addition of a halogen bond bridge tethering one flap to the protease body stabilizes the dynamic flaps. These factors combine to deliver an order of magnitude better inhibition of compound **2** over DRV for PR20.

2.5 Discussion

Our studies of PR20 mutant address the key challenge of HIV drug resistance [137,191]. PR20 achieves decreased affinity for inhibitors by synergistic mechanisms of an expanded inhibitor-binding cavity, highly dynamic flaps, and accelerated dimer dissociation compared to wild-type enzyme. Clinical inhibitor **1**, the most potent for wild-type PR (5 pM), exhibits 8,000-fold worse binding affinity (40 nM) for PR20[137]. Compound **2**, an antiviral inhibitor derived from **1**, shows 16-fold better K_i than **1** for PR20 (2.5 nM). The larger *Crn*-THF at P2 and Cp-Abt at P2' groups of **2** introduce additional hydrophobic and polar interactions with the mutated side-chains of the expanded ligand binding site. The fluorines of P1-*bis*-FBz of **2** create halogen bonds bridging the flap and Arg8′ to stabilize the dynamic flaps and restore an intersubunit ion-pair that may help to decrease the rate of dimer dissociation.

Compound **2** successfully realizes the current structure-guided drug design strategy for resistant PR. By incorporating larger P2/P2′ groups that fit into the expanded binding pocket and

halogen flap interactions, **2** can inhibit PR20 better than the best current clinical PI. Future compounds might explore modifications expanding the P1′ group. Furthermore, although these larger moieties of **2** help fill the expanded ligand binding pocket of PR20, other mutants exhibit few changes in the active site and diminish the potency of PIs through the effects of distal clusters of mutations[118]. In conclusion, antiviral inhibitor **2** is a noteworthy advance in the pursuit of potent inhibitors for resistant HIV.

Table 2.1 Crystallographic statistics for PR20/GRL-142
(Values in parentheses are for the highest resolution shell.)

	PR20/2
Resolution (Å)	50 - 1.21
Cell dimensions	
a (Å)	60.57
b (Å)	60.57
c (Å)	85.14
Space group	P 6_1
Unique Reflections	46,059
Completeness (%)	99.7 (89.8)
Redundancy	7.5 (2.3)
I/σ(I)	21.0 (2.0)
R_{merge} (%)	9.2 (44.8)
R_{work} (%)	15.2
R_{free} (%)	19.0
Solvent atoms	226
Average B-factors (Å^2)	
Protein	20.7
Inhibitor	13.8
Solvent	31.4
RMSD from ideality	
Bond lengths (Å)	0.015
Bond angles (°)	2.20

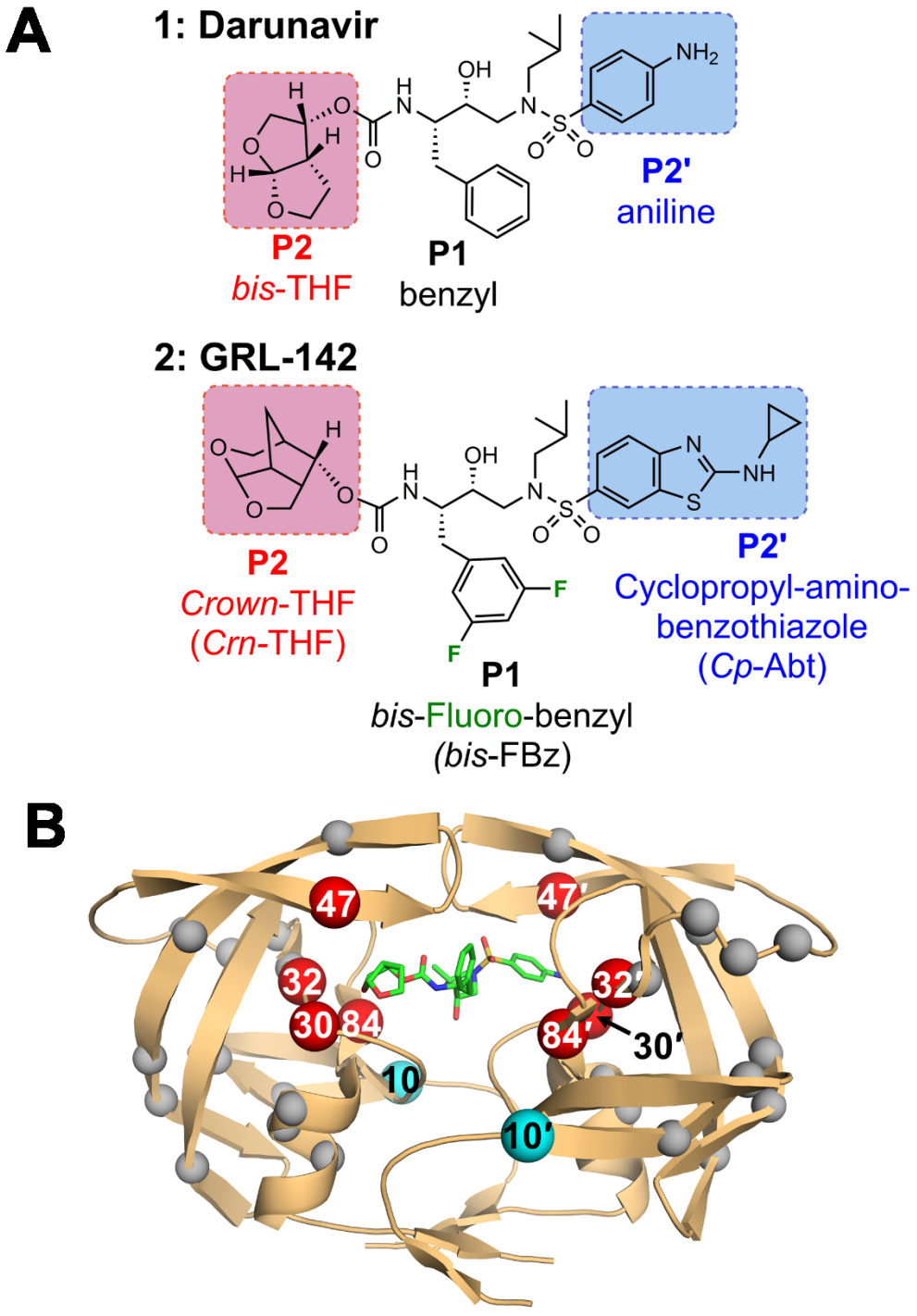


Figure 2.1: Chemical structures of compounds and PR20 dimer.
A. Inhibitor 1: darunavir and Inhibitor **2**: GRL-142 differ at the P2, P2′, and P1 positions. B. Inhibitor **1** (green sticks) bound to the active site of the PR20 dimer (orange cartoon). PR20 contains 4 mutations in the inhibitor-binding site D30N, V32I, I47V, and I84V (red spheres). The L10F mutation (cyan spheres) sits near the dimer interface. Grey spheres indicate other mutations.

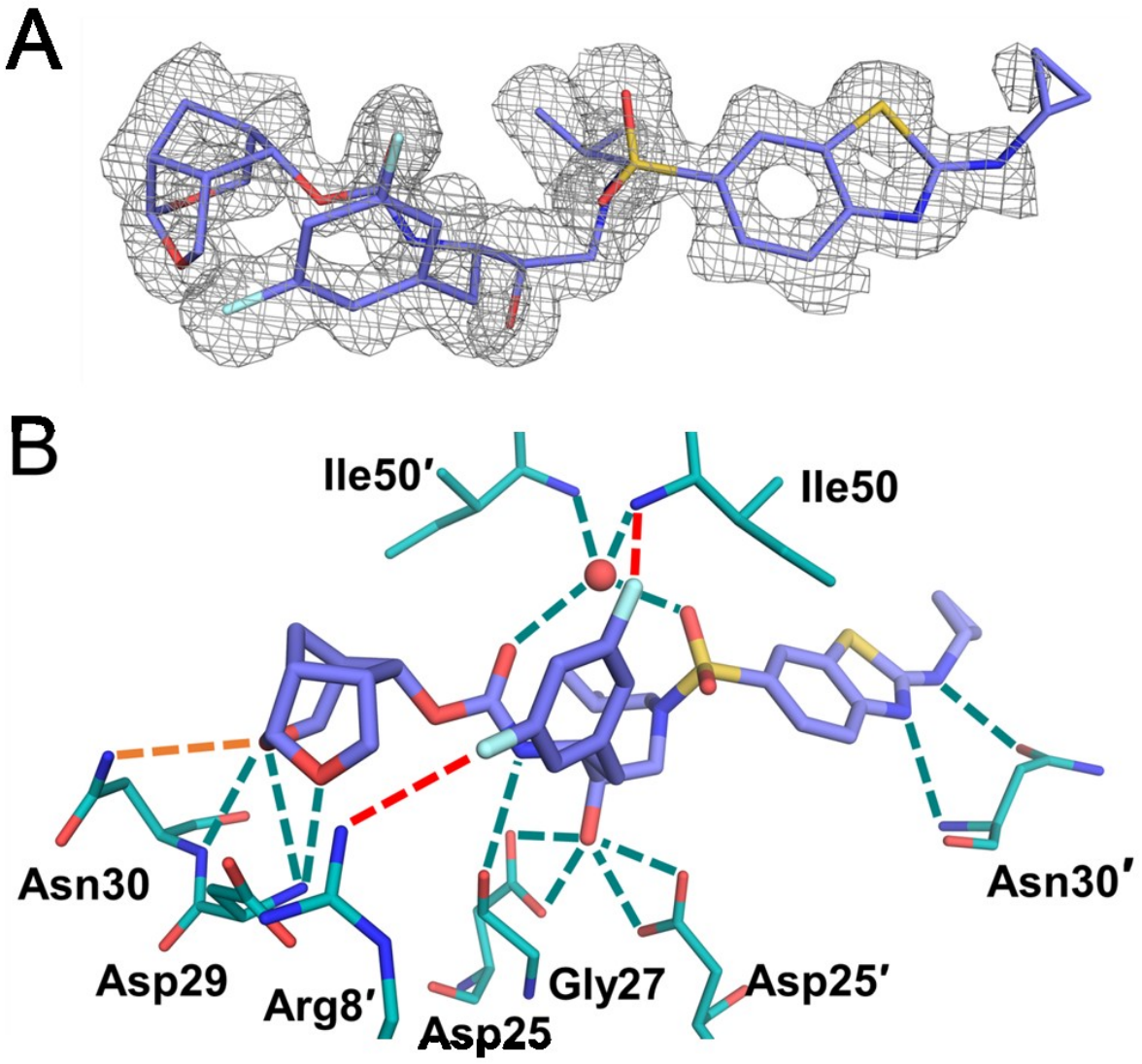


Figure 2.2: Inhibitor 2 bound in the active site of drug-resistant PR20.
A. Fo-Fc omit map (grey mesh) contoured at 2σ for the conformation of **2** (purple sticks) used for analysis. B. Polar interactions between **2** and PR20 residues (teal sticks). Dashed lines indicate hydrogen bonds (teal) and fluorine bonds (red dashes). New hydrogen bond in PR20/**2** is orange.

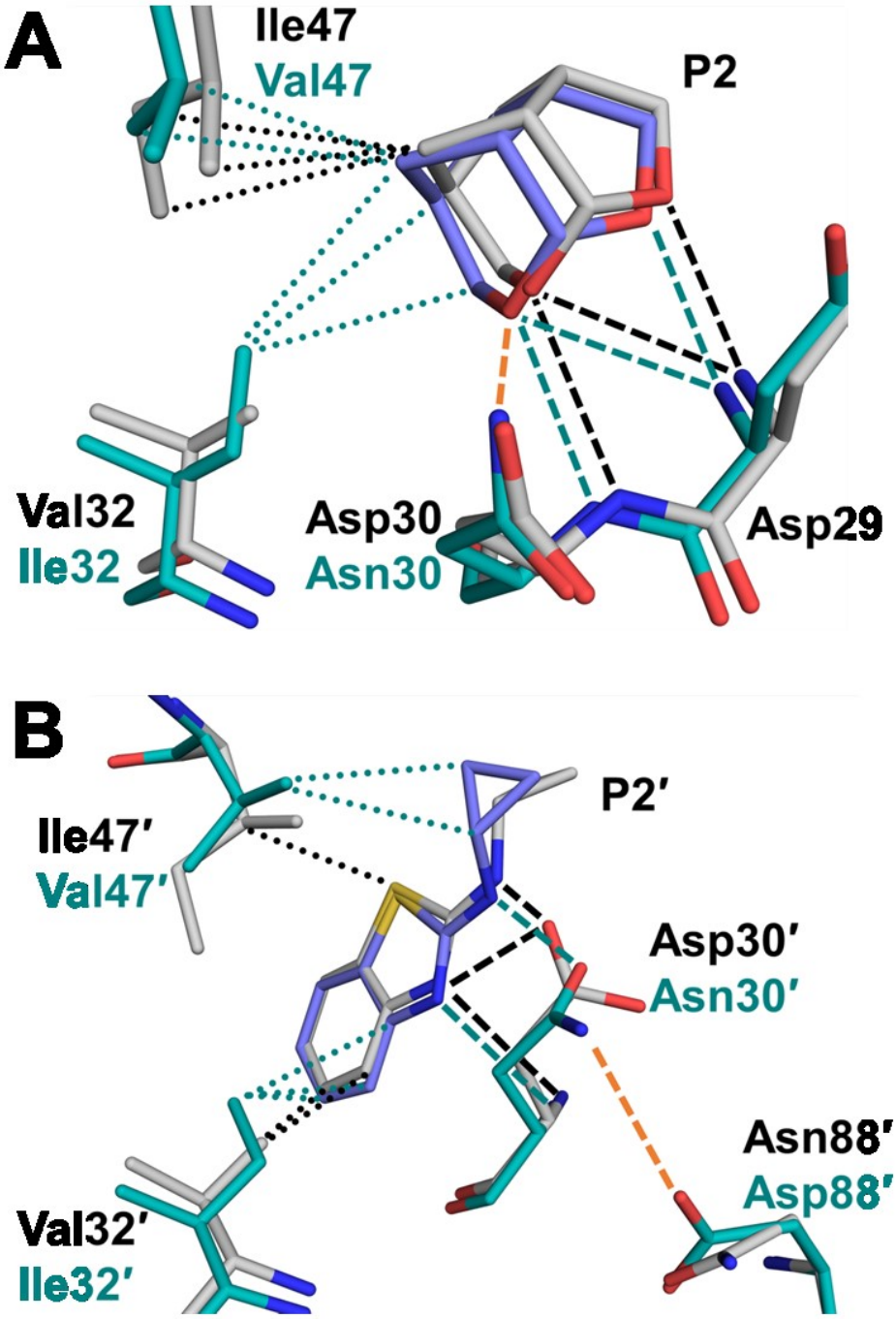


Figure 2.3: Interactions of the P2 and P2′ groups of **2** in PR20/2 compared to PR/**2**
A. Comparison of P2 *Crn*-THF group interactions with residues in the S2 pocket of PR20 and PR. PR20/**2** is colored purple/teal, PR/**2** is colored grey.
B. Comparison of P2′ Cp-Abt group interactions with residues in the S2′ pocket of PR20 and PR. Hydrogen bonds and van der Waals contacts are indicated as dashed and dotted lines, respectively, in teal for PR20/2 and black for PR/**2**. Orange dashes indicate hydrogen bonds introduced in the PR20/2 structure by mutations D30N and N88D.

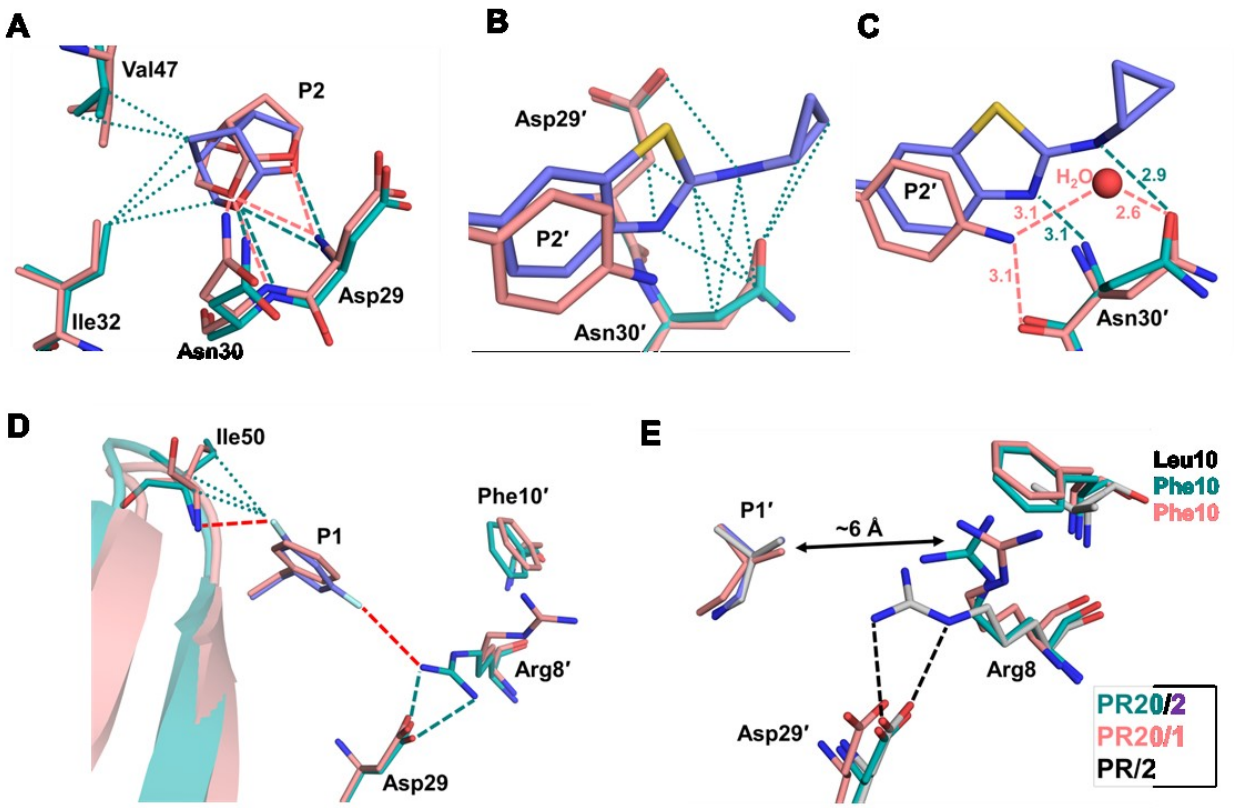


Figure 2.4: Comparison of interactions in PR20/**1** (salmon) and PR20/**2** (teal/purple).
A. Bulky P2 Crn-THF group of **2** creates 5 more van der Waals contacts with PR20 than does the *bis*-THF of 1.
B. Larger P2′ group on **2** adds van der Waals interactions with Asn30′ and Asp29′ side-chains that are absent in PR20/**1** structure.
C. The P2′ Cp-Abt group of 2 forms direct hydrogen bonds with the mutated side-chain of D30′N, whereas the aniline group of **1** makes weaker water-mediated hydrogen bonds. Distances in Å.
D. The *bis*-FBz of **2** makes fluoride bonds (red dashes) with Arg8′ and the tip of the flap at Ile50 that cannot be formed by the benzyl P1 of **1**. Intersubunit ion-pair between Arg8′ and Asp29 in PR20/2 is lost in PR20/**1** since Arg8′ shifts toward mutated L10F.
E. The short P1′-isobutyl in 1 and 2 cannot interact with Arg8 in PR20 or wild-type PR structures. Arg8 shifts toward mutated L10F in PR20 complexes instead of forming intersubunit ion pair observed in PR/**2** (grey).

3 HIGHLY DRUG RESISTANT HIV-1 PROTEASE MUTANT PRS17 SHOWS ENHANCED BINDING TO SUBSTRATE ANALOGS

Agniswamy J, Kneller DW, Brothers R, Wang Y-F, Harrison RW, Weber IT. (2019) Highly drug resistant HIV-1 protease mutant PRS17 shows enhanced binding to substrate analogues. *ACS Omega.* 4(5)8707-8719.

3.1 Abstract

We report the structural analysis of highly drug-resistant HIV protease variant PR^{S17}, rationally selected by machine learning, in complex with substrate analogs. Crystal structures were solved of inhibitor-free inactive PR^{S17}-D25N, wild-type PR/CA-p2 complex and PR^{S17} in complex with substrate analogs, CA-p2 and p2-NC. Peptide analogs p2-NC and CA-p2 exhibit inhibition constants of 514 and 22 nM, respectively, for PR^{S17} or approximately three-fold better than for PR. CA-p2 is a better inhibitor of PR^{S17} than are clinical inhibitors (K_i = 50 – 8390 nM) except for (K_i = 11 nM). G48V resistance mutation induces curled flap tips in PR^{S17}-D25N structure. The inner P2 to P2 ' residues of substrate analogs in PR^{S17} complexes maintain similar conformations to those of wild-type complex, while significant conformational changes are observed in the peripheral residues P3, P4′ of CA-p2 and P3, P4 and P3' of p2-NC. The loss of β-branched side-chain by V82S mutation initiates a shift in 80's loop and reshapes the S3/S3′ subsite, which enhances substrate binding with new hydrogen bonds and van der Waals interactions that are absent in wild-type structures. The steric hindrance caused by G48V mutation in the flap of PR^{S17} contributes to altered binding interactions of P3 Arg, P4′ Nle of

CA-p2 and P4 and P3' of p2-NC with the addition of new hydrogen bonds and van der Waals contacts. The enhanced interaction of PR^{S17} with substrate analogs agrees with their relative inhibition, suggesting this mutant improves substrate binding while decreasing affinity for clinical inhibitors.

3.2 Introduction

Human Immunodeficiency Virus (HIV) currently infects approximately 36.7 million people worldwide. At present, there is no cure for HIV infection or AIDS[280]. However, the advent of combined antiretroviral therapy (cART) in the late 1990's using a combination of reverse transcriptase and protease inhibitors dramatically suppressed the viral loads and increased CD4+ T-cell counts among infected patients[281,282]. In addition to viral fusion to host cell, reverse transcriptase and integrase, HIV protease (PR) is a critical anti-retroviral drug target due to its pivotal role in processing of Gag and Gag-Pol polyproteins during viral maturation. HIV PR is a homodimeric aspartic protease in which each 99-amino acid monomer contributes one of the two catalytic triads (Asp25-Thr26-Gly27) required for catalytic function[163]. The Gag and Gag-pol polyproteins are processed by HIV PR at 10 sites by recognizing specific sequences of 7-8 residues at the cleavage site. The substrate residues at the cleavage site are numbered P4-P3-P2-P1-P1′-P2′-P3′-P4′ with the scissile peptide bond between P1 and P1′ residues. 18O exchange mass spectrometry studies have shown that the peptide hydrolysis by HIV PR occurs through the formation of a reversible and metastable gem-diol reaction intermediate[128]. Several intermediates in the reaction pathway have been identified in X-ray crystal structures[283]. Inhibition of HIV PR leads to disruption of the virus life cycle at a vital stage[125]. Currently, there are 9 FDA-approved PR inhibitors (PI) that are available for cART. However, due to the rapid turnover rate of HIV and the low fidelity of reverse transcriptase (~2 x 10-5 errors/nucleotide per

replication cycle), mutations in PR associated with resistance to each of the 9 FDA-approved PIs have decreased the efficacy of current treatment options[164,284]. Accumulation of 10-20 mutations in PR may be required to evolve high levels of resistance[165]. Patients with multidrug-resistant HIV PR were shown to have convergent evolved patterns of accumulated resistance mutations[285].

Recently studied HIV PR variant PRS17 has 17 mutations compared to wild-type PR and was selected by mean-shift clustering on genotype-phenotype data using a new algorithm for predicting drug resistance with a unified encoding of the sequence and 3-D structure[202,203]. PRS17 was predicted to exhibit high-level resistance to six inhibitors: atazanavir, nelfinavir, ritonavir, indinavir, lopinavir, and saquinavir, and was subsequently confirmed to have resistance to 8 of the 9 currently available clinical inhibitors except for tipranavir which could not be measured due to solubility limitations[204,206]. Structural studies using X-ray crystallography and NMR spectroscopy have shown that PRS17 exhibits an open flap conformation in the absence of inhibitor in both crystal structure and in solution unlike wild-type PR which is present in the closed-flap conformation without the inhibitor in NMR studies[118]. In spite of different mutational profiles, correlated clusters of mutations in other multidrug-resistant variants of HIV PR (PR20 and MDR 769) cause them to also favor an open flap conformation like PRS17 that lowers the binding affinity of inhibitors[186,192]. Altered dynamics of the flap due to mutations is likely to increase dissociation rates, and hence, decrease potency of inhibitors[286]. Clusters of mutations at the hinge loop in PRS17 and PR20 are associated with increased flap dynamics and enhanced inhibitor dissociation of the two multidrug-resistant PRs[118,192]. These mutational clusters that function synergistically to produce conformational changes in the PR suggest convergent evolution by combinations of single mutations. Variant PRS17 is about 10-fold less

efficient than wild-type PR in processing the Gag polyprotein while maintaining the same order of substrate cleavage[206]. PRS17 is also 2-fold less efficient than PR20 in Gag polyprotein processing and forms an excellent prototype for structural studies on substrate binding to multidrug-resistant PR.

HIV PR hydrolyzes several sites with markedly different sequences on the nascent polyprotein substrate. In order for PR to evolve resistance to multiple drugs while maintaining catalytic efficiency for viral replication, mutations that destabilize inhibitor binding must combine with mutations that contribute to a new or altered form of substrate binding. This perspective shifts the focus of structure-guided drug design from point-based mutations to mutation clusters. Studying the mechanism by which multidrug-resistant variants like PRS17 bind and cleave substrates may elucidate conserved methods of binding across other mutant PRs with similar mutational clusters and give insights for improved structure-guided drug design. In the current study, we use structural and kinetic studies on PR, PRS17-D25N and PRS17 with substrate analogs CA-p2 (R-V-L-r-F-E-A-Nle) and p2-NC (Ace-T-I-Nle-r-Nle-Q-R) to evaluate differences in substrate recognition between PRS17 and wild-type PR. The reduced peptide analogs are suitable for structural studies with active PR variants since they cannot be hydrolyzed. Ace and the isosteric amino-acid norleucine (Nle) in which the sulfur of the Met side-chain is replaced by a methylene group were used to simplify synthesis.

3.3 Materials and methods

3.3.1 Construction, Expression and purification of PRS17

Synthetic genes encoding the 99 amino acid protein named PRS17 and its active site mutant PRS17-D25N (DNA2.0, Menlo Park, CA) were cloned in pJ414 vector flanked by Nde1 and BamH1 sites, and transformed into E.coli BL-21 (DE3; Stratagene). DNA sequencing and

electrospray ionization mass spectrometry were used to verify the recombinant construct. Protein expression, purification and refolding were carried out as described previously[287,288]. In brief, cells were expressed in Luria-Bertani medium and induced with 2mM IPTG. The harvested cells were lysed by addition of lysozyme (100 µg mL-1) and sonication at 4°C. The pelleted inclusion bodies were washed and solubilized in 50 mM Tris-HCL, pH 8.0, 8 M guanidine hydrochloride, 5 mM EDTA and 10 mM DTT. The protein was purified over a Superdex-75 column (HiLoad 2.6 cm x 60 cm, GE HealthCare) under denaturing conditions. The peak fractions were further purified by reverse-phase HPLC on SOURCE 15RPC ST 4.6/100 column using AKTA pure chromatography system (GE HealthCare). The purified protein was refolded by extensive dialysis against 30 mM formic acid and concentrated to the desired level using Amicon Ultra concentrators.

3.3.2 Kinetic inhibition measurements

Kinetic inhibition values (K_i) of p2-NC and CA-p2 for PRS17 were measured using a spectroscopic FRET substrate analog of the p2-NC site (H-2992 – BACHEM, Bubendorf, Switzerland). The peptide sequences of the two substrate analogs are R-V-L-r-F-E-A-Nle for CA-p2 and Ace-T-I-Nle-r-Nle-Q-R p2-NC. Enzyme kinetic parameters k_{cat} and K_m were determined for PRS17 using Michaelis-Menton analysis at 14-180 µM substrate concentration [S]. The k_{cat} and K_m are 69.7 ±14.4 min^{-1} and 143 ± 33.0 µM, respectively. The k_{cat}/K_m for PRS17 of 0.49 ± 0.05 min^{-1} μM^{-1} is comparable to previously reported values using a different method (0.55 min-1 µM-1)[206]. Reactions were performed in a total volume of 100 µL at 37°C in 50 mM MES pH 5.6, 200 mM sodium chloride, 0.5 mM EDTA, and 2.5% glycerol. Enzyme concentration [E] was 180-370 nM as measured by active site titration with APV. Inhibition assays used 60 µM [S]. The rate of substrate cleavage under a range of substrate analog

concentrations was measured continuously for 5 minutes by excitation at 340 nm and emission at 420 nm using a POLARstar OPTIMA microplate reader. The K_i value was determined using the equation $K_i = (IC50 - [E]/2)/(1 + [S]/K_m)$. IC50 was determined from dose-response curves. K_m values of FRET substrate for PRS17 were determined from Michaelis-Menton analysis at 14-180 μM [S]. All calculations were fitted using Sigmaplot 12.0 (Systat Software Inc., San Jose, CA, USA).

3.3.3 Crystallization

PRS17 was complexed with substrate analogs CA-p2 and p2-NC dissolved in DMSO at 1:5 molar ratio and incubated on ice for 30 minutes. The complex was centrifuged at 10,000 rpm for 5 minutes and the supernatant was used for crystallization. The crystallization trials were performed at room temperature with hanging drop vapor diffusion technique. The hanging drop in all crystallization trials was set up with 1 μL of protein solution (5mg/mL) and 1 μL of reservoir solution. Crystals of inhibitor-free PRS17-D25N were obtained from 1.95M sodium chloride and 0.1M Bis-Tris pH 7.5. Crystals of PRS17/CA-p2 were grown from reservoir solution containing 2.1M sodium chloride and 0.1M HEPES *(4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid)* at pH 7.6. The PRS17 crystals in complex with substrate analog p2-NC were obtained in two different space groups $P4_1$ and $P6_1$. The well solution used for growing PRS17-P41/p2-NC crystals was 29.5% PEG 4000, 0.2M ammonium acetate and 0.1M sodium acetate buffer at pH 4.6. The PRS17-P61/p2-NC crystals were grown from 35% Tacsimate™, pH 7.0 (Hampton Research Corp., Aliso Viejo, CA). Tacsimate™ contains 1.83 M malonic acid, 0.25 M ammonium citrate tribasic, 0.12 M succinic acid, 0.3 M DL-malic acid, 0.4 M sodium acetate trihydrate, 0.5 M sodium formate, and 0.16 M ammonium tartrate dibasic. Crystals of wild-type PR/CA-p2 were obtained from 1M sodium chloride and 0.1M sodium

acetate at pH 4.8. The crystals were cryo-cooled with cryoprotectant containing the respective mother liquor together with 30% glycerol.

3.3.4 X-ray data collection and structure determination

X-ray diffraction data were collected at 100 K on beamline 22-ID of the Southeast Regional Collaborative Access Team (SER-CAT) at the Advanced Photon Source, Argonne National Laboratory. The data were integrated and scaled with HKL2000[271]. The crystal structures of PRS17 active site mutant PRS17-D25N and the substrate analog complexes, PR/CA-p2, PRS17/CA-p2, PRS17-P41/p2-NC and PRS17-P61/p2-NC were solved by molecular replacement using PHASER[289,290]. The previously solved crystal structure of active PRS17 (5T2E)[118] was used as the starting model for PRS17 active-site mutant PRS17-D25N. For PR/CA-p2 substrate analog complex, the crystal structure of PR/p2-NC24 (2AOD) was used as the starting model. For PRS17/CA-p2 substrate analog complex, the crystal structure of PRS17 complex with darunavir (DRV)[118] (5T2Z) was used as initial model. Crystal structure of multiple mutants V6 in complex with indinavir (1SGU)[291] was used to solve both PRS17/p2-NC complexes. The PR/CA-p2 complex was refined with SHELX-1997[292]. The PRS17/CA-p2 complex was refined using REFMAC[293] while the PRS17-D25N and PRS17/p2-NC complexes were initially refined by SHELX-2014[294] and switched to REFMAC during later cycles of refinement. COOT[274] was used for most model building except O[295] was applied for PR/CA-p2. In PRS17/p2-NC complexes, mutations were added to the model based on the primary sequence and difference density maps. The substrate analogs were fitted into unambiguous electron density in the four complex structures. Solvent molecules were inserted at stereo chemically reasonable positions using 2Fo-Fc and Fo-Fc maps at 1 and 3 sigma levels, respectively. Molecular figures were prepared with PyMOL (http://www.pymol.org).

3.3.5 Accession codes

Coordinates and structure factors for the crystal structures PR/CA-p2 (6O48), PRS17-D25N (6O54), PRS17/CA-p2 (6O5X), PRS17-P41/p2-NC (6O57) and PRS17-P61/p2-NC (6O5A) have been deposited in the Protein Data Bank.

3.4 Results

3.4.1 Substrate analog inhibition of PRS17

Gag cleavage sites p2-NC and CA-p2 are the first and last proteolytic sites to be processed in the Gag precursor[83]. Mutations accumulating near the p2-NC cleavage site have been shown to improve processing fitness of a highly-mutated HIV PR sequence isolated from a patient failing antiviral therapy[153]. Reduced peptide analogs corresponding to the amino acid sequence of Gag cleavage sites act as competitive inhibitors of HIV PR catalytic activity[170,296,297]. In the present study, reduced peptide analogs for p2-NC and CA-p2 cleavage sites inhibited proteolytic activity of the drug resistant HIV-1 PR mutant PRS17. The k_{cat} and K_m values measured for PRS17 are 69.7 ±14.4 min^{-1} and 143 ± 33.0 μM, respectively. The k_{cat}/K_m of PRS17 is 0.49 ± 0.05 min^{-1} μM^{-1}. Inhibition constants (K_i) measured for PRS17 were 514 nM for p2-NC and 22 nM for CA-p2. The inhibition constants (K_i) of substrate analogs for PRS17 are compared with values for PR and PR_{V82A} determined previously in Table 3.1. Interestingly, reduced peptide analog CA-p2 displays better inhibition values for PRS17 than those of clinical inhibitors (K_i = 50 – 8390 nM) except for amprenavir (APV) (K_i = 11 nM) as measured by isothermal titration calorimetry[206]. CA-p2 and p2-NC substrate analogs are 3.4 and 4.2-fold better inhibitors for PRS17, respectively, when compared to PR whose K_i values are 75 nM and 2.17 μM. The major effect appears to be mutation to a smaller side-chain at Val82, since PR bearing single mutation V82A (PRV82A) displays almost identical inhibition kinetics for CA-p2

(K_i = 24 nM) and p2-NC (K_i = 530 nM) as those measured for PRS17[170]. PRS17 has a serine at this position, which is the only mutation in the active site cavity apart from those in flaps.

3.4.2 Overall structure

Crystal structures of inhibitor-free PRS17-D25N and wild-type PR in complex with CA-p2 substrate analog were determined in the $P322_1$ and $P2_12_12$ space groups at 1.21 and 1.46 Å resolution, respectively. PRS17 complex with CA-p2 crystallized in the monoclinic space group P21 and was refined to 1.7 Å, while PRS17 in complex with p2-NC was determined in two different space groups $P4_1$ and $P6_1$ at 1.71 and 1.67 Å resolution, respectively. The five structures were refined to R-factors of 14.8 – 19.8% as shown in the crystallographic statistics of Table 3.2. The inhibitor-free PRS17-D25N structure has one monomer in the asymmetric unit with residues numbered 1-99, while PR/CA-p2, PRS17-P41/p2-NC and PRS17-P61/p2-NC complexes contain a dimer of subunits numbered 1-99 and 1′-99′ in the asymmetric unit. PRS17/CA-p2 complex has two dimers per asymmetric unit. The substrate analogs bound at the active site in a single conformation in all PR and PRS17 complexes. All 17 mutations in the PRS17 (Fig 3.1A) and substrate analogs were unambiguously modeled in the electron density maps (Fig 3.1B and C). The reorganization of the hinge loop by mutations E35D, M36I and S37D that breaks the ion pair anchoring the flaps and thereby increases the flap flexibility as described in the previously determined complex of PRS17 with darunavir (DRV)[118] is also observed in the current substrate analog complexes.

3.4.3 Curling of flaps in PRS17-D25N

The current crystal structure of PRS17-D25N at 1.21 Å is the highest resolution HIV PR structure in the open conformation available to date. Comparison of inhibitor-free PRS17-D25N with the wild-type PR open form without inhibitor (2PC0) shows that the two structures

superpose with RMSD of 1.4 Å for the 99 topologically equivalent Cα atoms. The maximum deviation of 6.9 Å between the two structures occurs at Ile50 at the tip of the flap. Similar to wild-type PR, the flaps of PRS17-D25N are in an open conformation. The three mutations M46L, G48V and I54V in PRS17 flaps act to curl the tip of flaps (Figure 3.2A). The twist in the flap starts at G48V of PRS17-D25N with a drastic change in φ angle by ~172° in comparison with Gly48 of wild-type PR. The introduction of Val48 side-chain in PRS17-D25N results in ~1.1 Å shift of Phe53 side-chain's benzene ring when compared to that of wild-type PR. This shift results in loss of van der Waals contacts observed in wild-type PR between the side-chain of Phe53 and Gly48. In addition to the shift in Phe53, mutation of Met46 to shorter Leu46 in PRS17-D25N also eliminates van der Waals contacts observed in wild-type PR between Phe53 and Met46. This curling of flap tips propagates from G48V up to Gly52 in PRS17-D25N. As a result, the tip of the flap of PRS17-D25N moves ~ 4.4 Å closer to 80's loop as measured by the distance between the Cα atoms of Ile50 and Thr80. This distance is 14.5 Å for the wild-type PR, while it is 10.1 Å for the PRS17-D25N structure. The residues in 80's loop play an important role in the binding of both substrate and inhibitors to PR. Conformational changes and shifts in the 80's loop due to resistant mutations can eliminate interactions with inhibitors in both single mutants of PR and variants with multiple mutations[163,187,192]. On the other hand, the flap curling also increases by ~1.4 Å the distance between the flap tip of PRS17-D25N and active site, since the distance between the Cα atoms of Asp25 and Ile50 increased from 17.6 Å in the wild-type structure to 19.0 Å in PRS17-D25N. Nearly identical curled-flap conformation was seen in the recently reported inhibitor-free PRS17 structure[118]. The overall structure of PRS17-D25N is very similar to the inhibitor-free PRS17 structure with a RMSD of 0.18 Å for 99 equivalent Cα atoms. The maximum deviation between the two structures occurs for residue Ile50 at the flap tip with a

RMSD of 1.1 Å. Similar to PRS17-D25N, the flap curling in PRS17 starts with ~176° change in φ angle of G48V and ends at Gly52. However, the peptide bond between Ile50 and Gly51 is flipped in PRS17 structure in comparison to PRS17-D25N (Figure 3.2B). This flipped peptide bond enables residue Ile50 at the flap tip of PRS17 to move closer to the 80's loop residue Thr81 (~8.8 Å) than PRS17-D25N (10.1 Å), while the distance between catalytic Asp25/Asn25 and the flap tip Ile50 remains more or less same between the two structures (18.7 Å for PRS17 and 19.0 Å for PRS17-D25N). Both conformations of peptide bond between Ile50 and Gly51 are observed in other PR structures, which suggest it can exist in either conformation. Also, PR with individual mutations of M46L, G48V and I54V in the flaps showed worse inhibition constants (K_i) for DRV and saquinavir relative to wild-type PR[298,299]. Thus, identical curling of flaps in two independently determined structures of PRS17-D25N and PRS17 strongly implies that the conformational changes are due to the three mutations M46L, G48V and I54V in the flaps of PRS17 and contribute to the altered kinetic characteristics of PRS17.

3.4.4 PRS17 recognition of the P2, P1' - P3' groups of CA-p2 is unaffected by the resistance mutations

The PRS17/CA-p2 structure has 2 PR dimers in the asymmetric unit. The two dimers agree well and can be superposed with RMSD of 0.78 Å for the 198 canonically equivalent Cα atoms. Comparison of the two dimers in PRS17/CA-p2 structure with the wild-type PR/CA-p2 complex with one dimer in the asymmetric unit shows RMSD of 1.05 and 0.71 Å, respectively, for each dimer. The maximum deviation of 4.4 Å occurs between residues Glu35′ of PR/CA-p2 and PRS17/CA-p2 in dimer 1, while 4.3 Å deviation is observed between Glu35 positions in dimer 2. PR forms a series of hydrogen bond interactions with main-chain amide and carbonyl of the substrate that are important for recognition of peptides[170]. Similar hydrogen bonds between

the main-chains of PR and substrate analog were observed in both PR/CA-p2 and PRS17/CA-p2 complexes (Figure 3.3A), except for a new longer hydrogen bond between the main-chain amide of P4′ Nle and the side-chain of Asp29′ or Asp30′ in PRS17/CA-p2 complex which will be discussed in a later section. The conformations of P2 Val, P1′ Phe, P2′ Glu and P3′ Ala of CA-p2 and their interactions with the wild-type PR and PRS17 are similar in both complexes. However, conformational differences in P1 Leu, P3 Arg and P4′ Nle of CA-p2 and variations in their interaction with PR residues are observed between PR/CA-p2 and PRS17/CA-p2 complexes.

The P1 Leu of PR/CA-p2 forms van der Waals interactions with □-branched aliphatic side-chain of Val82′. In PRS17/CA-p2 complex, Val82 is mutated to polar Ser82 which alters the size of the S1 pocket where Pl Leu binds. The Pl Leu side-chain in PRS17/CA-p2 is rotated by ~46° about the χ2 angle in comparison to its conformation in the wild-type complex thus enabling P1 Leu to retain its van der Waals interaction with the side-chain of V82′S in PRS17 (Figure 3.3B). The conformational change in the side-chain of P1 Leu also facilitates retention of van der Waals contacts with Leu23′, Gly27, Ile50, Pro81′ and Ile84′ observed in PR/CA-p2 complex (Figure 3.3B). The P1 Leu of CA-p2 in the second dimer of PRS17/CA-p2 complex exhibits an identical conformation change with ~46° rotation about χ 2 angle and retention of van der Waals contact with V82′S side-chain, which further confirms that the observed change in P1 Leu of CA-p2 occurs in response to V82′S mutation in PRS17.

3.4.5 Altered recognition of CA-p2 P3 Arg modulated by V82S mutation

The P3 Arg main-chain of PR/CA-p2 is fixed in position by the hydrogen bond between the N-terminal group of P3 and the side-chain of Asp29 and a second hydrogen bond between the main-chain carbonyl of P3 Arg and the main-chain amide of Asp29. These two hydrogen bonds between Asp29 and the main-chain of P3 Arg are also present in PRS17/CA-p2 complex

(Figure 3.4A). The side-chain of P3 Arg in the wild-type complex extends over the Gly48 in the first strand of flap and forms van der Waals contacts with the Phe53 side-chain in the second strand of the flap. In PRS17, the Val side-chain of G48V mutation produces steric hindrance and prevents formation of the interaction between Phe53 and P3 Arg observed in wild-type PR complex. The tip of Phe53 benzene ring in PRS17/CA-p2 moves away by more than 2 Å in comparison to wild-type complex. In PRS17/CA-p2 complex, P3 Arg is observed in three different conformations. Conformation 1 is in dimer 1 of PRS17/CA-p2 complex, while conformations 2 and 3 of P3 Arg are present as alternate conformations of 0.5 occupancy each in the dimer 2 of PRS17/CA-p2 complex. Due to steric hindrance caused by G48V mutation, all three conformations of the side-chain of P3 Arg in PRS17/CA-p2 dimers swing away from Val48 and interact with Arg8′. The inter-subunit ion pair formed by Arg8′ and Asp29 is critical for the dimer stability of HIV PR[92]. PRS17 utilizes both residues involved in this conserved ion pair to recognize and position P3 residue of CA-p2 substrate.

The new conformation of P3 Arg forms additional van der Waals contacts with Arg8′, G48V, Pro81′ and Ser82′ in PRS17 relative to those observed in the wild-type complex (Figure 3.4A). Also, the new orientation of P3 Arg guanidine head forms water-mediated hydrogen bonds with the hydroxyl side-chain of Ser82′, which stabilizes the conformation of P3 Arg. Further, the guanidine head of P3 Arg in PRS17/CA-p2 forms a double water-mediated hydrogen bond with the carbonyl oxygen of catalytic-triad residue Gly27 and Asp29. These internal structural waters are highly conserved in most HIV PR structures including the wild-type PR/CA-p2 complex. Similar interactions are also observed for the conformation 2 and 3 of P3 Arg in the second dimer of PRS17/CA-p2 (Figure 3.4B). In conformation 2, the guanidine of P3 Arg forms stacking interaction with Arg8′ (Figure 3.4C) while the guanidine of P3 Arg in

conformation 3 forms van der Waals interaction with Arg8′ and hydrogen bond interaction with the main-chain carbonyl oxygen of Pro81′ (Figure 3.4D). The P3 Arg conformation 3 in dimer 2 forms fewer van der Waals contacts to Arg8′ than seen for conformation 1 in dimer 1. The guanidine head of P3 Arg in conformation 2 is buried deep in the pocket and forms additional van der Waals contacts with Leu23′ that are not present in conformations 1 or 3. This conformation is further held in place by a single water-mediated hydrogen bond between P3 Arg guanidine head and Asp29 as well as another one to the main-chain carbonyl of catalytic triad residue Gly27 of PRS17. In contrast, conformation 3 forms double water-mediated hydrogen bonds with Gly27 and Asp29. The guanidine head of both conformation 2 and 3 of P3 Arg in dimer 2 has a water-mediated hydrogen bond to the hydroxyl of Ser82′ which further anchors P3 Arg of substrate analog. The V82′S mutation in PRS17 facilitates the new conformation of P3 Arg side-chain, since Val82′ in the wild-type complex would sterically interfere with P3 Arg in this location. Mutations of Val82 to Ala, Thr, Phe, Ile, Ser and Leu are associated with resistance to clinical HIV PR inhibitors[164,300]. Similar to V82S mutation in PRS17, V82′A mutation is also expected to expand the binding site for P3 Arg, but other larger and β-branched mutations may block the access to this site. The co-occurrence of flap mutations like G48V that block P3 substrate-binding site and large substitutions at residue 82 like V82F, V82L and V82I that block the alternate P3 binding site are expected to be detrimental to PR activity. In the previously solved crystal structure of PRV82A/CA-p2, the substrate analog is in two alternate conformations related by 180° with 0.65/0.35 occupancies as opposed to a single conformation in the current PR and PRS17 complexes. The major conformation of P3 Arg in the PRV82A/CA-p2 complex forms van der Waals contact with Phe53 in the second strand of the flaps similar to PR/CA-p2 complex while the carbonyl of P4 Nle of the minor conformation related by 180°

occupies the P3 Arg site observed in the PRS17/CA-p2 complex (Figure 3.4E). Similar to V82S, V82A mutation opens up the S3 site for substrate recognition. Thus, in PRS17/CA-p2 complex, the G48V flap mutation in conjunction with the V82S mutation facilitates binding of the side-chain of P3 Arg in the new site. The effect of Val82 mutation appears to be two-fold; it alters the resistance profile of PR inhibitors and regulates the access to Arg8/8′ in the recognition of substrate P3 residue in response to resistance flap mutations like G48V that hinder substrate recognition.

3.4.6 Effects of G48'V mutation on binding of CA-p2 P4' Nle to Leu76' hydrophobic pocket

Comparison of PR and PRS17 complexes with CA-p2 reveals significant differences in the binding of P4′ Nle. In wild-type PR/CA-p2 complex, the terminal P5′ NH2 group forms a hydrogen bond interaction with the main-chain carbonyl of Leu46 and van der Waals interactions with Phe53 in the second strand of the flap (Figure 3.4F). However, in the mutant PRS17/CA-p2 complex, the introduction of G48V mutation causes the Phe53 side-chain to rotate out by more than 140° in comparison to wild-type complex. The G48V mutation further contributes to more than 1.5 Å shift in the position of terminal P5′ NH2 which losses van der Waals contact with Phe53 but retains the hydrogen bond with the carbonyl of Leu46. The main-chain amide of P4′ Nle is moved by ~0.8 Å in the PRS17/CA-p2 complex in comparison to the wild-type complex, which results in a new long hydrogen bond (~3.5 Å vs ~4.1 Å in the PR/CA-p2 complex) between the main-chain amide of P4′ Nle and the side-chain of Asp29′ in dimer 1. P4′ Nle amide in dimer 2 forms a water-mediated hydrogen bond with Asp29′ and a new long hydrogen bond with the side-chain of Asp30′ (~3.5 Å vs ~3.8 Å). The Cα atom of P4′ Nle is also shifted by ~1.2 and 0.5 Å in comparison to its position in PR/CA-p2 in dimer 1 and 2 respectively. Along with the shift in the main-chain atoms, the side-chain of P4′ Nle in dimer1

swings into the hydrophobic pocket around Leu76′ and forms a hydrophobic interaction with the side-chain of Leu76′. The side-chain of Lys45′, which occupies this pocket and has van der Waals contacts with Asp30′ in the wild-type PR/CA-p2 structure, swings away to form a van der Waals interaction with the Gln58′ side-chain in PRS17/CA-p2 complex, thereby enabling P4′ Nle of CA-p2 to interact with Leu76′. P4′ Nle bound in this pocket also forms van der Waals interactions with Asp30′, Lys45′ and Gln58′ in the mutant. In contrast, the P4′ Nle side-chain in wild-type complex shows van der Waals contact with only one residue, Lys45′. The P4′ Nle side chain in the second dimer of PRS17/CA-p2 also binds in the same pocket with similar interactions due to the G48′V flap mutation in PRS17. Thus, the G48′V mutation in the flap and the altered conformation of Phe53′ in PRS17/CA-p2 complex induce a shift in P4′ Nle side-chain away from the flaps to form a new hydrogen bond with Asp29′ or Asp30′. Further, the side-chain of P4′ Nle in the new orientation binds in the hydrophobic pocket near Leu76′ with added van der Waals contacts. Overall, this new arrangement in the mutant confers favorable interactions with P4′ Nle.

3.4.7 Mutation V82S mediates enhanced binding of P3′ Arg in p2-NC

PRS17 in complex with p2-NC was crystallized in two different space groups $P4_1$ and $P6_1$ and refined to resolutions of 1.71 and 1.67 Å, respectively. Both structures have one PRS17 dimer per asymmetric unit and a single conformation of substrate analog bound at the active site cavity. The two structures are very similar with RMSD of 0.5 Å for 198 equivalent Cα atoms. The previously solved PR/p2-NC complex can be superposed with the PRS17-P41/p2-NC and PRS17-P61/p2-NC complexes with RMSD values of 0.99 and 0.96 Å, respectively, for the 198 equivalent Cα atoms. Despite their different space groups, both structures exhibit a maximum deviation of more than 4.3 Å at the same residue, Asp35, in comparison to wild-type structure.

The conformation of P2, P1, P1′, P2′ and P4′ residues of p2-NC and their interactions with PRS17 residues are similar for PRS17/p2-NC complexes in the two space groups and PR/p2-NC. However, noticeable differences occur at residues P3′, P3 and P4 of PRS17-P41/p2-NC and PRS17-P61/p2-NC in comparison to PR/p2-NC.

Unlike the P3 Arg of PR/CA-p2 complex, the P3′ Arg of PR/p2-NC binds in the space between Arg8 and Val82 (Figure 3.5A). The side-chain of P3′ Arg forms van der Waals contacts with Arg8 and the CD2 side-chain atom of Val82. In both PRS17-P41/p2-NC and PRS17-P61/p2-NC complexes, the P3′ Arg binds in the same pocket albeit the guanidine head of P3′ Arg is tilted towards V82S mutation due to the lack of a β-branched Cδ atom in residue 82. The Nε2 atom of the guanidine head of P3′ Arg in PRS17 complexes is shifted by 1.0 Å (PRS17-P41/p2-NC) and 1.2 Å (PRS17-P61/p2-NC) towards V82S in comparison to the PR/p2-NC complex. The loss of β-branched side-chain in V82S mutation results in a significant shift of 80's loop residues towards Arg8 in PRS17 complexes. The Cα atoms of residues Thr80, Pro81 and V82S are shifted by 1.1, 1.3, 0.7 Å, respectively, for the PRS17-P41/p2-NC complex and 1.3, 1.4, 0.6 Å for the PRS17-P61/p2-NC complex from the corresponding Cα atoms in PR/p2-NC complex (Figure 3.5A). This shift contracts the binding pocket for P3′ Arg of p2-NC by 1.3 Å and 1.4 Å, as measured by the distance between the guanidine head Nη1 of Arg8 and the Cα atom of Pro81, in PRS17-P41/p2-NC and PRS17-P61/p2-NC in comparison to the wild-type complex. This narrowing of the pocket results in enhanced binding of P3′ Arg as it is able to form a new hydrogen bond interaction with the main-chain carbonyl of Pro81 in addition to van der Waals interactions with the side-chain of Pro81 in PRS17/p2-NC complexes. In contrast, P3′ Arg in the wild-type PR/p2-NC complex has no interaction with Pro81. Both mutant PRS17/p2-NC complexes retain the extensive van der Waals contacts between the Arg8 of PRS17 and P3′ Arg

of substrate analog seen in wild-type complex. Similar shifting of 80's loop residues is also observed in P3 pocket of both dimers of PRS17/CA-p2 complex in comparison to wild-type structure, although the shifts are reduced by half ranging from 0.3 – 0.5 Å, likely due to the two alternate conformations of P3 Arg bound in this pocket. Thus, analysis of the three PRS17 substrate analog complexes indicates that the loss of β-branched side-chain by V82S mutation initiates shifting of 80's loop and reshapes the S3/S3' subsite for enhanced binding and interactions with P3/P3'.

3.4.8 G48V mutation of PRS17 results in altered binding position of P3 and P4 residues of p2-NC and addition of two new hydrogen bonds

The P3 Thr in both PRS17-P41/p2-NC and PRS17-P61/p2-NC complexes exhibits significant structural changes in comparison to wild-type complex. The p2-NC main-chain carbonyl oxygen of P3 Thr occupies identical positions in all three structures. The hydrogen bond observed between the P3 carbonyl and the main-chain amide of Asp29 in PR/p2-NC complex as well as the conserved water-mediated hydrogen bonds between the P3 carbonyl and Gly27 and Asp29 are preserved in both PRS17/p2-NC complexes. However, the larger G48V mutation in PRS17 complexes induces shifts in P3 Thr Cα and Cβ atoms by ~ 0.7 Å and 1.8 Å in comparison to their corresponding positions in PR/p2-NC complex (Figure 3.5B). In this new conformation, the P3 Thr side-chain forms several van der Waals contacts with the side-chains of G48V mutation and Arg8′ in PRS17 complexes. It also forms a hydrogen bond interaction with Arg8′ side-chain in both the complexes. In contrast, P3 Thr side-chain in wild-type PR/p2-NC complex forms a lone van der Waals contact with Arg8′ side-chain, but no hydrogen bonds. In comparison to the wild-type complex, the main-chain amide nitrogen of P3 Thr in both PRS17/p2-NC complexes is rotated about ~160° along the main-chain Cα-C bond. This results in

a shift of the amide nitrogen of P3 Thr in both mutant complexes by ~2.8 Å from the corresponding position in PR/p2-NC complex. The amide nitrogen of P3 Thr forms a hydrogen bond interaction with the carbonyl oxygen of Gly48 in wild-type PR/p2-NC complex, while it interacts with the side-chain of Asp29 in both PRS17/p2-NC complexes.

The terminal P4 Ace residue of p2-NC exhibits the maximum deviations among the substrate analog residues in both PRS17 complexes in comparison to PR/p2-NC. The P4 residue in the PRS17 cannot occupy the same pocket as in the wild-type structure due to steric hindrance from G48V mutation and the loss of interaction with Val82′ as a result of V82′S mutation. The P4 Cα is shifted by ~4.6 Å in both complexes from its position in the wild-type complex. Due to G48V mutation inducing ~160° of rotation along the Cα-C bond of P3 residue, the P4 Ace residue occupies a different binding pocket in PRS17 complexes to that observed in the wild-type structure. The P4 residue in wild-type PR/p2-NC complex forms van der Waals contacts with Gly48, Gly49, Pro81 and Val82. In contrast, the P4 residue in PRS17/p2-NC complexes has van der Waals contacts with a different set of residues: Val48, Asp29, Asp30, Lys45, and Ile47. In addition, the main-chain carbonyl oxygen of the P4 residue in PRS17 complexes forms new hydrogen bond interaction with the main-chain amide of Val48. Thus, despite large shifts in P3 and P4 residues induced by the G48V mutation, PRS17 accommodates the substrate residues in alternate sites and gains two hydrogen bond interactions. The new hydrogen bonds observed in the PRS17/p2-NC complexes likely contribute to the improved p2-NC substrate analog inhibition constant observed for PRS17 compared to PR.

3.4.9 PRS17 binds substrate analogs CA-p2, p2-NC and inhibitor DRV without drastic change in PR conformation around the active site

The two dimers of the PRS17/CA-p2 complex together with the PRS17-P41/p2-NC and PRS17-P61/p2-NC structures can be superposed with PRS17/DRV complex with an RMSD of 0.4 to 0.7 Å. No dramatic changes are observed in the conformation of residues around the active site catalytic triad and the binding of substrate analogs and inhibitor (Figure 3.5C). Most of the conformational changes are seen in the flap residues. The serine side-chain of V82S mutation in both monomers of PRS17/DRV has two alternate conformations mimicking the two β-branches of valine, while in the substrate analog complexes it has a single conformation that points away from Arg8 to accommodate the substrate residues. Side-chain rotamer changes are observed in Phe53 and in mutated side-chains of Leu46 and Val48 in complexes with DRV and the two substrate analogs. In addition, shifts are observed for flap residues. A shift of 0.8 Å in M46L is observed in PRS17/CA-p2 in comparison to complex with DRV. Similarly, Val48 of PRS17 shifts by 0.6 Å when bound to p2-NC analog in comparison to its position in the complex with DRV. Flap mutations M46L, G48V and I54V in PRS17 are implicated in the curling of flaps observed in the inhibitor-free structure[118]. In PRS17/DRV complex, M46L and G48V mutations are shown to synergistically alter the conformation of Phe53 by steric hindrance underling the relationship between the three residues[118]. Through changes in the flap strands including the position of M46L or G48V in combination with the change in rotamer for side-chains of Leu46, Val48 and Phe53, PRS17 binds substrate analogs as well as it binds DRV (K_i = 50 nM). The larger side chain of the G48V mutation in the S2 pocket may play an important role in the better inhibition constant (K_i = 11 nM) of APV for PRS17. Thermodynamic integration and MM-PBSA studies have shown that the larger side-chain introduced by G48V mutation strengthens van der

Waals interactions with APV30. However, G48V together with M46L also alters the conformation of other flap residues like Phe53 as seen in PRS17/DRV complex[118] which may affect the inhibitor binding. Thus, minimal structural changes enable highly drug-resistant PRS17 to bind substrate analogs with similar affinity as for DRV, and better than for the other clinical inhibitors, contributing to both viral fitness and drug resistance.

3.5 Discussion

Our studies on PRS17 demonstrate the value of using machine learning on genotype-phenotype data to select informative mutants for further analysis. Among the 17 mutations in PRS17, G48V and V82S play critical roles in binding of CA-p2 and p2-NC substrate analogs. The active site cavity around Val82 is important for the binding of PR substrates as well as clinical inhibitors. Drug-resistant mutations of Val82 are reported for all clinical PR inhibitors except for DRV[164]. The side-chains of mutations vary in size from the smaller V82A to bulkier V82F in addition to V82S, V82T, V82L and V82I that are in between[164]. V82A is one of the earlier PR mutations to occur at the active site among patients undergoing antiviral therapy[301] and has been the best studied mutation at position 82. Kinetic studies have shown that PR bearing the single V82A mutation has significantly worse affinity for PR inhibitors indinavir, nelfinavir, ritonavir and DRV, while smaller effects were seen for cleavage of PR substrate[180,302–304]. Structural studies on V82A mutant in complexes with nelfinavir, indinavir, ritonavir, DRV and saquinavir all reveal a small shift in residues to accommodate the inhibitor, explaining the cross-resistance profile of this mutation to all PR inhibitors[163,305]. Resistance mutations like V82A at the active site of HIV PR interfere with the binding of drugs, however, replication of the virus requires successful binding and hydrolysis of Gag and Gag-Pol substrates by resistant PR variants. Crystal structures of inactivated PR with active site D25N and V82A mutations in

complex with substrates CA-p2, p1-p6 and MA-CA show high structural similarity with inactivated PR-D25N/substrate complexes[306] which indicates that V82A mutation in conjunction with active site D25N does not affect substrate binding. However, mutations in the Gag substrate cleavage sites can co-evolve with resistance-associated mutations in HIV PR or mutations that alter the fitness of HIV PR[154,307,308]. NC-p1 cleavage site, which is the rate limiting step in Gag polyprotein processing[309], has been reported to co-evolve with Ala to Val mutation at the P2 position of the substrate in response to V82A resistant mutation[277,307,310].

Mutations in the flap region also alter the PR susceptibility to inhibitors. In particular, flap mutation G48V confers a high level of resistance to saquinavir[205]. G48V mutation is also selected by PR inhibitors atazanavir, indinavir, lopinavir and nelfinavir[205,311,312]. G48M has a similar resistance profile as G48V[205]. Less prevalent resistance mutations G48A/S/T/Q are less well studied and their clinical significance is not fully understood[313]. Most of the Gly48 mutations occur in viruses with multiple PI-resistant sequences[313] like PRS17. G48V is associated with V82A in saquinavir resistance[314,315] and this pair occurs as a mutational cluster together with non-active site mutant C95F, which emphasizes the importance of co-evolution of mutations at 48 and 82 in drug resistance. Mutations at 48 and 82 together with those at 30, 84 and 90 are considered primary mutations that are necessary for inhibitor resistance[316]. The present structural study reveals that PRS17, a multidrug-resistant PR harboring primary mutations at 48 and 82 in addition to others, recognizes substrate analogs CA-p2 and p2-NC in an altered fashion in comparison to wild-type structures. The crystal structure of unliganded PRS17-D25N exhibits curling of flap tips in comparison to unliganded wild-type structure. This flap curling is initiated by G48V mutation with a ~172° change in ϕ angle in comparison to PR wild-type and extends up to Gly52'. In addition to G48V other flap mutations M46L and I54V

may also contribute to curling in PRS17. Identical flap curling was observed in the recent unliganded PRS17 crystal structure[118]. Also, solution studies of PRS17-D25N by NMR demonstrated that the inhibitor-free PRS17-D25N adopts a flap conformation very similar to that of inhibitor-free active PRS17 crystal structure. The current crystal structure of PRS17-D25N further confirms that the conformation changes observed in the flaps of the two unliganded structures of PRS17 are due to the mutations. Curling or twisting of flaps in the PR structures due to mutations in the flaps is implicated in the altered resistance profile of the mutant PR. F53L single mutant which exhibits curled flaps was shown to poses 15% lower catalytic efficiency and 20-fold weaker inhibition by clinical drug indinavir[317]. Twisting of flaps observed in the recently reported PR with 22 mutations in addition to inactivating D25N mutation has >175-fold resistance to DRV in addition to APV[318]. The flap twisting initiated by G48V mutation in PRS17 is implicated in the cross resistance of PRS17 to multiple clinical drugs including DRV even though it lacks any major mutations associated with DRV resistance[118].

Although the binding site for P3 Arg of CA-p2 in the current PR/CA-p2 complex is partly exposed to solvent, P3 Arg interacts with flap residue Phe53′ in a manner similar to that seen in CA-p2 complexes with PR (4EP3), PRV82A (2AOE) and inactive PR-D25NV82A (1MT8). With P3 Arg bound at the preferred wild-type flap binding site, mutations at 82 with large side-chains like V82F, V82I or V82L hinder binding of PR inhibitors thereby altering PR susceptibility to PI's while retaining binding to substrates. The G48V mutation in PRS17 also confers resistance to PR inhibitors16, and it blocks the flap binding site for P3 Arg of CA-p2 observed in the various wild-type PR complexes. However, V82/82′S mutation in PRS17 results in ~1 Å shift in the residues of the 80's loop and rearrangement of the S3/S3' pocket between V82/82′S and the conserved Arg8′/8. Both P3 Arg of CA-p2 and P3′ Arg of p2-NC bind in this

subsite of PRS17 with enhanced interactions including a new direct hydrogen bond, water-mediated hydrogen bonds and van der Waals contacts to the 80's loop that are absent in corresponding wild-type structures. This alternate binding site for P3/P3′ substrate residues is particularly critical for PR fitness when resistance mutations like G48/48′V are present at the flaps. The G48/48′V mutation also alters the binding of P4′ Nle of CA-p2 as well as P3 Thr and Ace residues of p2-NC substrate analogs.

The enhanced binding of CA-p2 and p2-NC substrate analogs with new hydrogen bond interactions observed in PRS17 complexes in comparison to PR complexes agrees well with 3.4 and 4.2-fold better K_i values for PRS17 relative to those for wild-type enzyme. In addition, comparison of PRS17 substrate analog complexes with that of DRV reveals that highly multiple drug resistant PRS17 efficiently binds substrate analogs and DRV with minimal conformational changes, thereby contributing to both viral fitness and drug resistance. The current structural studies of PRS17 strongly suggest that G48V mutation plays an important role in the enhanced binding of substrate analogs CA-p2 and p2-NC. Further, the size of the mutated side-chain selected at residue 82 may be influenced by the presence of mutation at the flap residue Gly48. In the absence of G48V mutation, Val82 mutations enhance resistance profile of PR through steric hindrance to inhibitors at the binding site. However, in the presence of G48V flap mutation, mutations of residue 82 to bigger side-chains like V82F, V82L, and V82I may interfere with binding of substrates, while smaller V82S and V82A mutations might play vital role in both enhanced binding of substrate and increased resistance to PI's through structural shift of 80's loop. The structural features identified in the current studies reveal valuable information for the future design of drugs targeting multidrug-resistant HIV PRs. Clearly, further

studies on the substrate recognition by multidrug-resistant PR variants will bring out new ways of targeting viral fitness and activity.

Table 3.1: Inhibition constants (K_i) in nM for PR, PRV82A, and PRS17

protein	P2-NC	CA-p2
PR	2170	75
PR^{V82A}	530 (4.1)	24 (3.2)
PR^{S17}	514 (4.2)	22 (3.4)

[a]Values in parentheses are K_i values relative to the wild-type PR.

Table 3.2: Crystallographic data and refinement statistics

	PR^{S17}-D25N	PR/CA-p2	PR^{S17}/CA-p2	PR^{S17}/p2-NC	PR^{S17}/p2-NC
space group	$P3_221$	$P2_12_12$	$P2_1$	$P4_1$	$P6_1$
cell dimensions					
a (Å)	49.65	58.41	53.70	54.88	63.83
b (Å)	49.65	86.24	60.36	54.88	63.83
c (Å)	86.70	46.58	66.48	82.13	83.04
α (deg)	90	90	90	90	90
β (deg)	90	90	90.93	90	90
γ (deg)	120	90	90	90	120
resolution range (Å)	50.0−1.21 (1.25−1.21)	50.0−1.46 (1.51−1.46)	50.0−1.7 (1.76−1.70)	50.0−1.71 (1.77−1.71)	50−1.67 (1.73−1.67)
unique reflections	35 459 (1956)	38 064 (2696)	46 434 (4633)	26 117 (2508)	21 945 (1989)
redundancy	5.0 (2.1)	4.1 (1.7)	3.9 (3.8)	6.1 (3.4)	4.4 (2.3)
completeness	91.9 (52.1)	91.9 (66.5)	99.5 (99.6)	99.1 (95.2)	98.4 (89.2)
$\langle I/\sigma(I)\rangle$	18.7 (2.0)	11.0 (2.0)	17.1 (6.1)	21.2 (2.6)	16.9 (2.0)
R_{sym} (%)	7.1 (42.5)	9.0 (43.8)	7.6 (41.8)	7.1 (47.1)	5.7 (44.4)
refinement resolution range (Å)	50.0−1.21	50.0−1.46	34.53−1.7	27.46−1.71	33.22−1.67
R (%)	14.8	18.8	17.6	19.8	17.6
R_{free} (%)	16.4	24.5	21.2	23.1	21.6
number of atoms in asymmetric unit	920	1762	3416	1744	1804
number of water molecules	105	115	170	78	125
average *B*-factor (Å^2)					
main chain	18.6	15.7	18.0	39.1	28.6
side chain	23.3	21.1	23.8	44.9	35.1
inhibitor		26.2	19.6	37.9	25.4
waters	29.3	27.3	24.6	44.3	34.0
RMS deviations from ideality					
bond length (Å)	0.02	0.01	0.02	0.02	0.02
angles	1.8°	0.03 Å	2.2°	2.2°	2.2°

[a]Values in parentheses are for the highest shell.

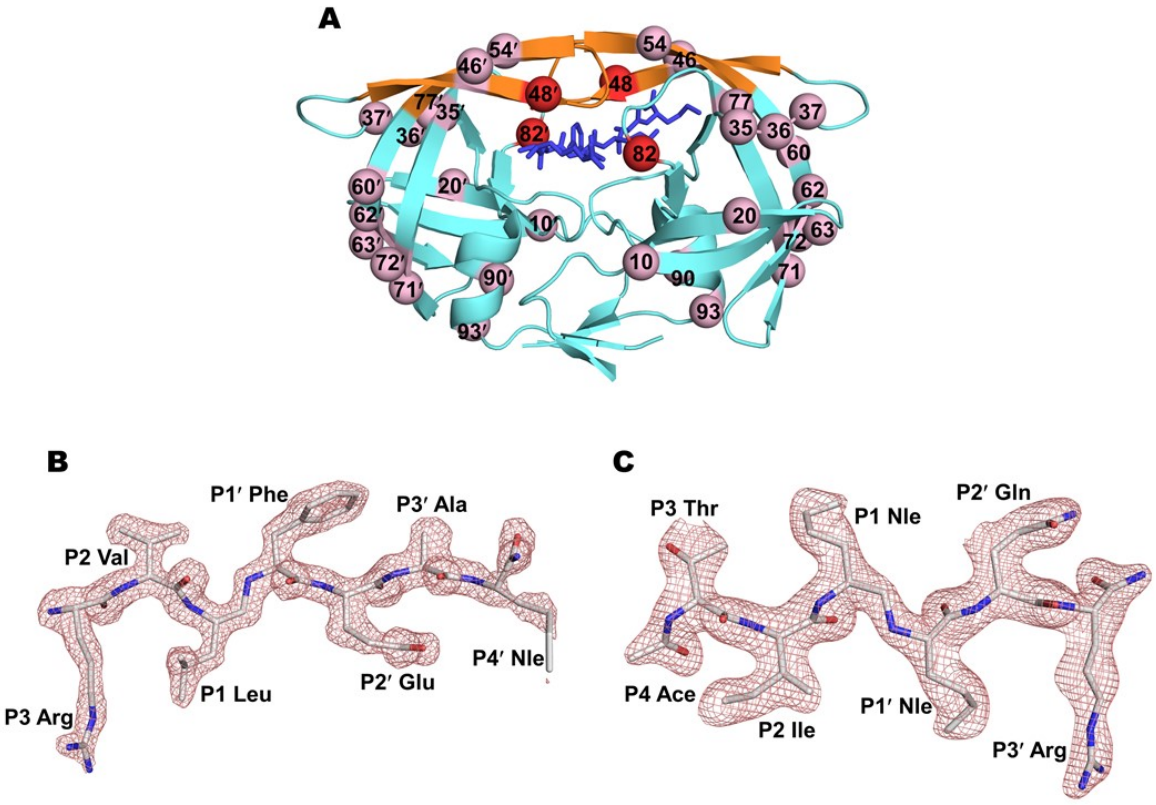


Figure 3.1 Sites of mutation in PR^{S17} dimer and substrate analogs.
A. Sites of 17 mutations are mapped on PR^{S17} dimer (cyan cartoon representation) with flap region in orange and bound substrate analog CA-p2 shown as blue sticks. The critical G48V and V82S mutations involved in substrate analog binding are shown as red spheres in each subunit.
B. F_o-F_c omit map of CA-p2 colored by element type contoured at 2.8σ level.
C. F_o-F_c omit map of p2-NC colored by element type contoured at 3σ level.

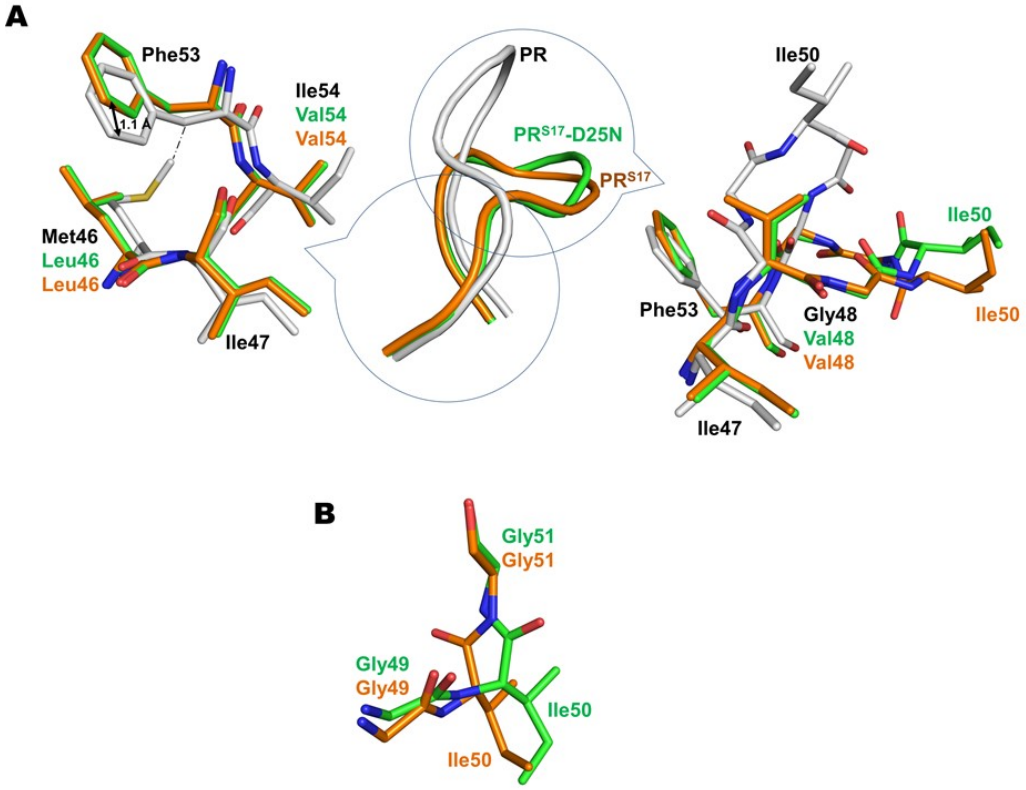


Figure 3.2: Altered flap of PR^{S17}-D25N.
A. Curling of flap due to flap mutations M46L, G48V and I54V observed in inhibitor-free structures of PR^{S17}-D25N (green ribbon) and PR^{S17} (orange ribbon) in comparison to PR (grey ribbon). The circled portions of flap are magnified in sticks for inhibitor-free PR^{S17}-D25N (green carbon), PR^{S17} (orange carbon) and PR (grey carbon).
B. Flipped peptide bond between Ile50 and Gly51 of PR^{S17}-D25N (green carbon) and PR^{S17} (orange carbon).

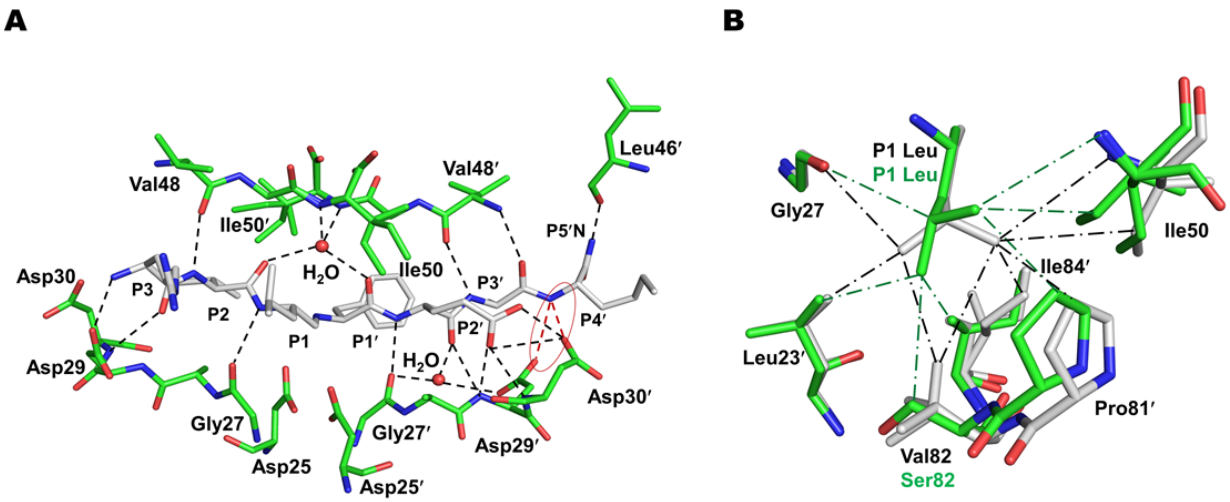


Figure 3.3: Overall interaction of substrate analog CA-p2 with PR^{S17}.
A. Hydrogen bond interactions between PR^{S17} (green carbon) and substrate analog CA-p2 (grey carbon). Water molecules are shown as red spheres in this and subsequent figures. Hydrogen bonds are shown in black dotted lines and the new hydrogen bond observed between PR^{S17} and CA-p2 is in red dotted line.
B. Comparison of CA-p2 P1 Leu interaction in PR^{S17}/CA-p2 (green carbon) and PR/CA-p2 (grey carbon). The van der Waals contacts are represented by (– · – · –) lines in green in PR^{S17}/CA-p2 and in black in PR/CA-p2 complex.

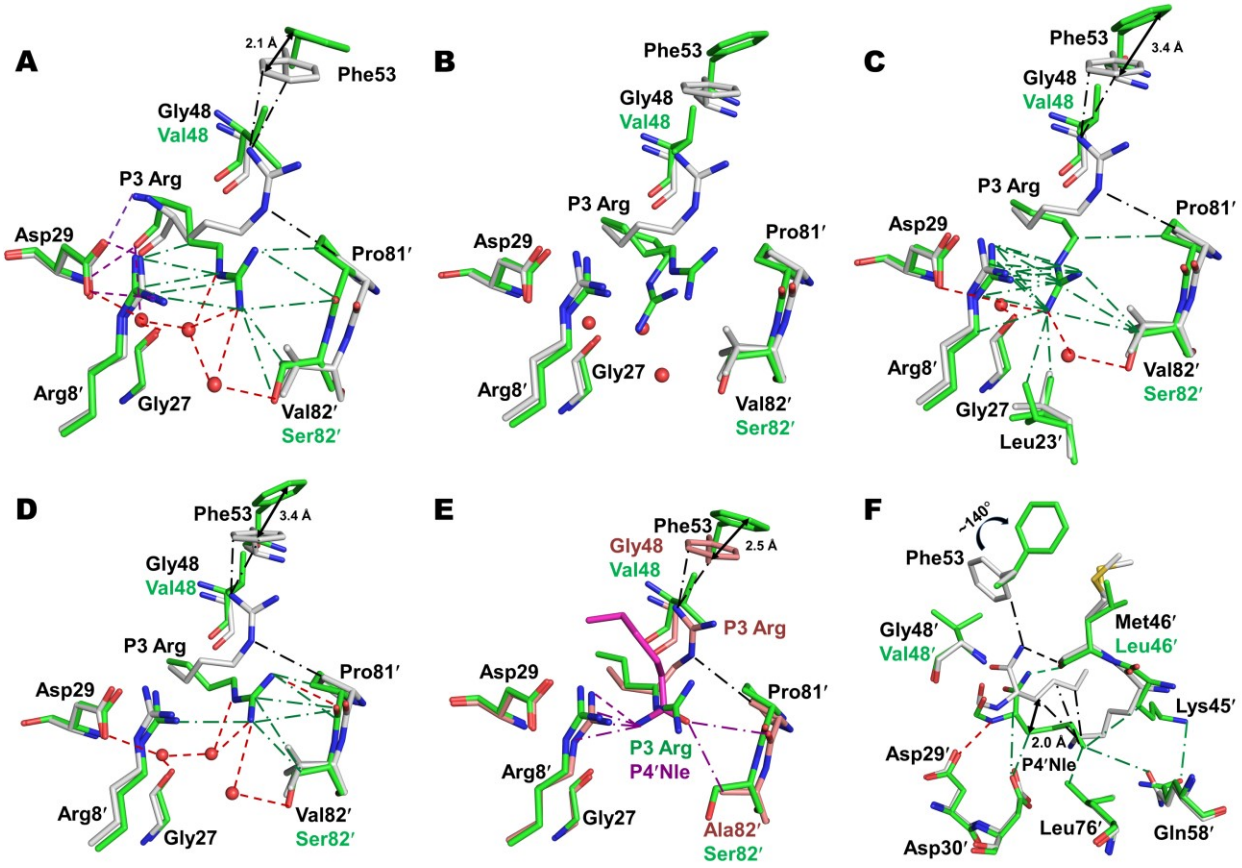


Figure 3.4: Interactions of P3 and P4′ of substrate analog CA-p2 with PR^{S17}.
A. Interaction of P3 Arg of CA-p2 (conformation1) in dimer 1 of PR^{S17}/CA-p2 (green carbon) and PR/CA-p2 (grey carbon). The hydrogen bonds common to both the complexes are shown in purple dotted lines. In all sub panels, new hydrogen bonds observed in PR^{S17}/CA-p2 complex are shown in red dotted lines. The van der Waals contacts are represented by (– · – · –) lines in green in PR^{S17}/CA-p2 and in black in PR/CA-p2 complex.
B. P3 Arg of CA-p2 (conformation 2 and 3) in dimer 2 of PR^{S17}/CA-p2 (green carbon). Interactions of conformations 2 and 3 are shown in panels C and D respectively.
C. Interactions of P3 Arg conformation 2 in dimer 2 of PR^{S17}/CA-p2 (green carbon) in comparison to PR/CA-p2 (grey carbon).
D. Interactions of P3 Arg conformation 3 in dimer 2 of PR^{S17}/CA-p2 (green carbon) in comparison to PR/CA-p2 (grey carbon).
E. Interactions of major conformation P3 Arg (salmon carbon) and minor conformation P4' Nle (magenta carbon) of V82A single mutant complex PR^{V82A}/CA-p2 in comparison to PR^{S17}/CA-p2 (green carbon). The van der Waals contacts of P3 Arg of PR^{V82A}/CA-p2 are shown in black while those of P4' Nle are shown in magenta.
F. Difference in the binding of P4′ Nle between PR/CA-p2 dimer1 (grey carbon) and PR^{S17}/CA-p2 (green carbon). The hydrogen bond present only in PR^{S17}/CA-p2 is colored in green, and the hydrogen bond present in PR/CA-p2 complex alone is shown in black dotted lines.

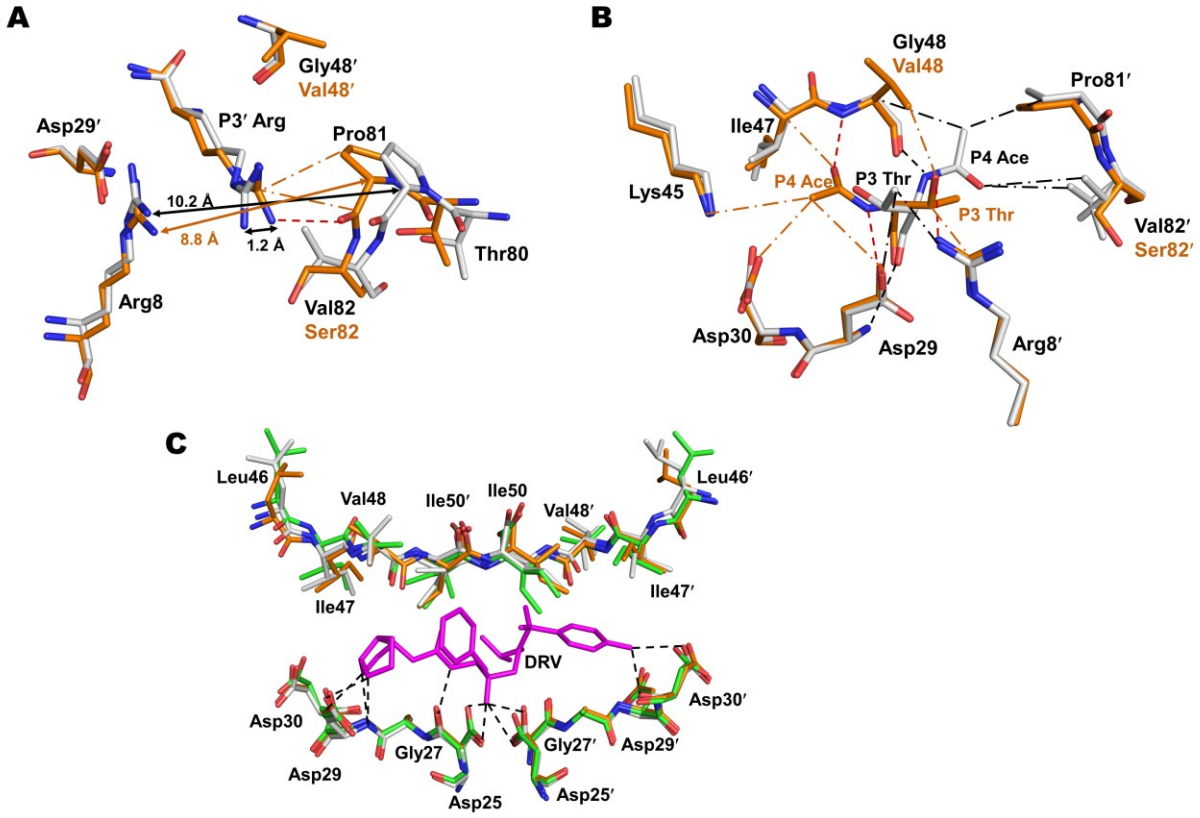


Figure 3.5: Binding of P3′, P3 and P4 residues of p2-NC and DRV to PR^{S17}.
A. The P3' Arg forms new hydrogen bond and van der Waals contacts colored in red with Pro81 of PR^{S17}/p2-NC (orange carbon) in comparison to PR/p2-NC (grey carbon, 2AOD).
B. The difference in P3 Thr interactions between PR^{S17}/p2-N (orange carbon) and PR/p2-NC (grey carbon). The three new hydrogen bonds observed in PR^{S17}/p2-NC are colored in red. The minor conformation of Arg8' observed in PR/p2-NC was omitted for clarity. The van der Waals lines in this panel show residues in contact and may have more than one interaction between the residues.
C. Superposition of active site cavity of PR^{S17}/CA-p2 (green carbon), $PR^{S17\text{-}P41}$/p2-NC (orange carbon) and PR^{S17}/DRV (grey carbon, 5T2Z). DRV bound at the active site is shown in magenta sticks. The hydrogen bond interactions between DRV and PR^{S17} are shown in dotted lines. For the sake of clarity, water-mediated hydrogen bonds between DRV and the flaps are omitted.

4 HIGHLY DRUG-RESISTANT HIV-1 PROTEASE REVEALS DECREASED INTRA-SUBUNIT INTERACTIONS DUE TO COORDINATED STRUCTURAL CHANGES IN CLUSTERS OF MUTATIONS

Kneller, DW, Agniswamy, J, Harrison, RW, Weber, IT. (2019) Highly drug-resistant HIV-1 protease reveals decreased intra-subunit interactions due to coordinated structural changes in clusters of mutations. In revision, *The FEBS Journal*.

4.1 Abstract

Drug resistance is a serious problem for treatment of the HIV/AIDS pandemic. Potent clinical inhibitors of HIV-1 protease show several orders of magnitude worse inhibition of highly drug resistant variants. Hence, the structure and enzyme activity were analyzed for HIV protease mutant PRS5B from a clinical isolate that was selected by machine learning to represent high level drug resistance. PRS5B has 22 mutations including only one (I84V) in the inhibitor-binding site, however, clinical inhibitors had poor inhibition of PRS5B activity with K_i values of 4-1000 nM or 18-8000-fold worse than for wild-type PR. High resolution crystal structures of PRS5B complexes with the best inhibitors, amprenavir and darunavir (K_i ~4 nM), revealed only minor changes in protease-inhibitor interactions. Instead, two distinct clusters of mutations in distal regions induce coordinated conformational changes that decrease favorable internal interactions across the entire protein subunit. The largest structural rearrangements are described and compared to other characterized resistant mutants. In the protease hinge region, the N83D mutation eliminates a hydrogen bond connecting the hinge and core of the protease and increases disorder compared to highly resistant mutants PR^{S17} and PR20 with similar hinge mutations. In a

distal β-sheet, mutations G73T and A71V coordinate with accessory mutations to bring about shifts that propagate throughout the subunit. Molecular dynamics simulations of ligand-free dimers show differences consistent with loss of interactions in mutant compared to wild-type PR. Clusters of mutations exhibit both coordinated and antagonistic effects suggesting PRS5B may represent an intermediate stage in the evolution of more highly resistant variants.

4.2 Introduction

Globally, human immunodeficiency virus (HIV) is estimated to infect 36.7 million people[5]. Treatments using a combination of inhibitors targeting HIV entry/fusion steps and the viral enzymes, reverse transcriptase (RT), integrase, and protease (PR), have greatly improved patient outcomes. However, HIV resistance to all current drugs is a serious problem due to the presence of numerous genomic polymorphisms and rapid evolution of virus[319]. Moreover, expansion of antiretroviral therapy has resulted in infection with virus resistant to RT inhibitors in greater than 10% of untreated individuals in some countries[33]. Thus, recent WHO guidelines promote second-line PR inhibitor (PI) treatments.

HIV evolves resistance to PIs primarily by mutations in the PR gene[320]. HIV-1 PR is required for proteolytic cleavage of Gag and Gag-Pol precursor polyproteins into individual proteins during viral maturation, and is essential for producing infectious virus[86]. The enzyme forms a dimer of 99-residue subunits. Mutations in over 50% of the PR residues are associated with resistance to all current clinical PIs and lead to decreased effectiveness and treatment failure [164]. Major mutations produce drug resistance by three general molecular mechanisms: mutations of residues in the inhibitor binding site can directly alter PR-PI interactions; some distal mutations confer drug resistance by altering dimerization via changes in inter-subunit

interactions; and other distal mutations, such as L76V, can destabilize the dimer and decrease PI potency by creating structural rearrangements in key locations of the protease[163,172,176–178].

During drug treatment, the virus can evolve increasingly higher levels of resistance by combining major mutations with additional accessory or compensatory mutations[321] which can persist even after termination of treatment[322]. Several well characterized PR mutants include about 20 different mutations and show several orders of magnitude decreased susceptibility to clinical inhibitors [165]. Two such variants, PR20 and PR^{S17}, form stable dimers with 2-4-fold less efficient Gag processing compared to wild-type PR while maintaining the same order of cleavage [137,193,206]. Mutations in PR20 coordinate to expand the binding cavity for inhibitors, consistent with dramatically decreased binding affinity. Analysis of ligand-free and darunavir (DRV)-bound X-ray crystal structures of PR20 and PR^{S17} showed clusters of mutations in the hinge and flap regions conferring changes in conformation of the flaps[118,192]. Nuclear Magnetic Resonance (NMR) spectroscopic analysis of ligand-free PR20 and PR^{S17} demonstrated that these mutants tend to exist in an open-flap conformation in the absence of inhibitor, while wild-type enzyme preferentially occurs in the closed conformation, even in the absence of inhibitor[118,323]. These changes in PR structure and dynamics are proposed to be important for acquiring high level resistance to PIs[118].

Here, we have investigated the enzyme kinetics and inhibitor-bound structures of an HIV protease mutant with 22 substitutions from a resistant clinical isolate, dubbed PRS5B. Like PR^{S17}, the mutant was selected as a representative of highly drug-resistant PR by mean-shift clustering on genotype-phenotype data in Stanford HIVdb [300,324] using a unified encoding of the sequence and 3-D structure[202–204]. In this procedure, PR^{S17} was the sole mutant common to the clusters with highest level resistance for each of six PIs, whereas PRS5B mutant was the second

most common mutant predicted to represent high-level resistance to only five PIs: atazanavir (ATV), nelfinavir (NFV), indinavir (IDV), tipranavir (TPV), and saquinavir (SQV). Two mutations in PRS5B have a major association with drug-resistance (M46L and I84V), while nine are classified as minor drug resistance mutations (L10I, V11I, M36I, I54V, I62V, I63P, I64V, A71V, and G73T)[164]. Mutation N83D is classified solely as a major resistance mutation for TPV[164]. Here, we confirmed by enzyme kinetics experiments that these drugs, in addition to lopinavir (LPV), amprenavir (APV), and DRV, have significantly worse inhibition for PRS5B relative to wild-type PR. Analysis of X-ray crystal structures of PRS5B in complexes with DRV and APV revealed how mutation clusters in distal regions of the protease may contribute to drug resistance via long-range indirect effects. The PRS5B structures are compared to PR^{S17} and PR20 complexes with the same inhibitors. Conformational variation of PRS5B and wild-type PR in the absence and presence of inhibitors is explored by molecular dynamics (MD) simulations. These studies illuminate how HIV PR can evolve drug resistance through synergistic effects of clusters of mutations in regions far from the inhibitor binding site. This information will be advantageous for improving inhibitor design against highly drug resistant virus.

4.3 Methods

4.3.1 Selection of the PRS5B sequence

The protein sequence for PRS5B was obtained by the same machine learning procedure described for PR^{S17}. Mutants representing broad classes of high level resistance to each inhibitor were selected by mean shift cluster with regression analysis of genotype-phenotype data in the Stanford HIVdb (https://hivdb.stanford.edu) [204]. The procedure used a unified encoding of the sequence and 3D structure as described in [204]. Only a single mutant, PR^{S17}, was observed in the

highest resistance bin for 6 inhibitors. PR^{S17} and PRS5B mutants were identified as representing high resistance to different combinations of 5 inhibitors.

4.3.2 Expression and purification of PRS5B

The gene coding for PRS5B was synthesized and cloned on a pJ414 plasmid with a T7 promoter [ATUM] and transformed into E. coli BL21 (DE3) (Invitrogen) cells[138]. PRS5B contains additional mutations Q7K to reduce autoproteolysis and C67A and C95A to eliminate cysteine thiol crosslinking[325,326]. Protein expression, purification, and folding were performed as described[287,288].

4.3.3 Enzyme kinetic assays

Kinetic experiments and enzyme dilutions were carried out in reaction buffer (50 mM MES pH 5.6, 200 mM sodium chloride, 0.5 mM EDTA, and 2.5% glycerol final reaction concentration). PIs obtained from the NIH AIDS Reagent program were dissolved in 100% DMSO. A spectroscopic fluorescence resonance energy-transfer (FRET) substrate analog derived from the HIV-1 Gag p2/NC cleavage site (H-2992 – BACHEM, Bubendorf, Switzerland) was used to monitor the initial rate of substrate hydrolysis by PRS5B at 37°C. The enzyme kinetic parameters (k_{cat}, K_m of FRET substrate, and k_{cat}/K_m) and inhibition constants (K_i) were determined as described in [207]. Enzyme concentration [E] was determined using active site titration with tight-binding inhibitor, APV. K_i was obtained from the equation $K_i = (IC_{50} - [E]/2)/(1 + [S]/K_m)$. Values are reported as mean ± SD (n=3-5).

UC_{50} for PRS5B, PR^{S17}, and PR20 were measured using the same reaction buffer, substrate, enzyme, and temperature conditions as K_i measurements with the addition of 0-1.2 M urea as described in [327]. Values are reported as mean UC_{50} ± 3 standard deviations (n=3-4) where UC_{50} = urea concentration at 0.5 velocity at 0 M urea.

4.3.4 Crystallization and X-ray structure determination

Inhibitor dissolved in 100% DMSO was mixed at an 8:1 molar ratio with 4.2 mg/mL PRS5B solutions and incubated on ice for 30 minutes. Crystallization of the complexes were performed using hanging-drop vapor diffusion at room temperature using 1 μL protein/PI complex and 1 μL reservoir solution. Crystallization conditions for PRS5B/DRV were 2.2 M Ammonium Phosphate and 100 mM Tris buffer at pH 7.9. PRS5B/APV crystallized in 1.8 M Ammonium Phosphate and 100 mM Tris buffer at pH 7.9. Crystals grew after approximately 14 days and were cryo-protected stepwise in reservoir solution with 30% glycerol. Diffraction data were collected at 100°K on SER-CAT beamline 22-ID at the Advanced Photon Source, Argonne National Laboratory (Argonne, IL, USA).

Diffraction data were integrated and scaled using HKL2000[271]. The crystal structure of PRS5B/DRV was solved by molecular replacement in Phaser using PR/DRV (2IEN) as a search model[272,289]. The refined PRS5B/DRV structure was used as a search model for PRS5B/APV. Integrative rounds of refinement using COOT[274] and REFMAC5[293] were used to construct the atomic models. In each structure, the inhibitor was fitted during refinement into unambiguous electron density in a single conformation. Statistical information was gathered using baverage and superpose[328]. Hydrogen bond interactions and van der Waals contacts were inferred from interatomic distances in the ranges of 2.4-3.4 Å and 3.0-4.2 Å, respectively. Structural figures were prepared using PyMOL[276].

4.3.5 Molecular dynamics simulations

Molecular dynamics (MD) simulations were performed for two ligand-free (PRS5B and PR) and four inhibitor-bound structures (PRS5B/APV, PRS5B/DRV, PR/APV and PR/DRV). The starting models were the wild-type PR complexes with DRV (PDB: 2IEN[304]) and APV

(PDB: 3NU3[166]) and the corresponding complexes of the PRS5B mutant (PDB: 6P9A and 6P9B). Alternate conformations were deleted from the models, keeping the majority conformation. Inhibitors were removed for MD simulations of the ligand-free proteases. Solvent molecules from the crystal structures were included in the simulations and a mobile proton (H+) was introduced between the carboxylates of the catalytic Asp25 and 25' to enable the protein to assume the variable protonation states observed in different neutron structures of HIV PR[124,133,329]. Each dimer was solvated with about five thousand water molecules generated randomly to fill the free space within a 30Å shell of the protein. Energy minimization and MD simulations used the program, AMMP[330] with the Tuna potential set[331], a modification of the SP4 set described in [332] with charges updated to those of Amber ff14SB[333]. In tests on small molecule benchmarks, the potential sets compare well to the CHARMM and AMBER sets[332]. AMMP is the molecular mechanics and dynamics engine of VegaZZ[334] and has undergone extensive professional code review as part of the SPEC2000 benchmark[335]. It is designed to be a computational backend that is plugged into other programs.

Positions were calculated for all hydrogen atoms and the geometry of the solvated protein system was minimized by conjugate gradients prior to starting the MD simulations. All-atom simulations were performed as described in [194,336]. The aperiodic amortized fast multipole algorithm was used for the electrostatic and non-bonded terms with parameters (mxdq = 0.75 Å; mmbox = 10.0 Å), 300K temperature constrained with Nose constraints, and a bounding volume constraint of 35 Å. Frames were saved for every 1 pc during the 10-20 ns simulations. The RMSD values and variance from the mean were calculated after superimposing each frame on the starting model. Selected interatomic distances were calculated using Python programs. The structural variation during the simulations was examined in a gif of the consecutive frames.

4.4 Results

4.4.1 Enzyme kinetic parameters

Kinetic parameters for PRS5B were determined using a FRET-based assay that employed a substrate analog of the p2/NC cleavage site. The measured k_{cat} of 109 ± 20 min^{-1} for PRS5B is ~2-fold lower than for wild-type PR (194 min^{-1}), and PRS5B exhibits 5-fold higher K_m of 160 ± 45 μM relative to PR (30 μM)[298]. The catalytic efficiency of PRS5B (k_{cat}/K_m = 0.69 ± 0.07 $\mu M^{-1}min^{-1}$) is comparable to that of PR^{S17} (0.49 $\mu M^{-1}min^{-1}$) under identical conditions[207] and close to that of PR20 (0.35 $\mu M^{-1}min^{-1}$) measured under different conditions with chromogenic substrate[137]. All three of these highly resistant mutants show about 10 to 20-fold reduced catalytic efficiency compared to PR (6.5 $\mu M^{-1}min^{-1}$ for FRET-substrate at 26 °C).

Inhibition constants of 8 PIs measured for PRS5B are compared with previously reported values for wild-type enzyme [206,337] in Table 1. The PIs showed K_i values of 4-1000 nM for PRS5B, or 18-8,000-fold worse than for wild-type enzyme. Machine-learning predictions suggesting resistance to inhibition by TPV, ATV, NFV, IDV, and SQV for PRS5B were confirmed by these enzyme inhibition results. PRS5B was poorly inhibited by SQV (2,500-fold relative to wild-type PR) and LPV (8,000-fold), while significantly worse inhibition of 40 to 800-fold was measured for DRV, ATV, TPV, NFV, and IDV. The lowest K_i value (3.6 nM) and relative inhibition (18-fold) for PRS5B was shown for APV, however, DRV also had a K_i of 4.0 nM. LPV showed significant resistance, although not the highest level, in the Stanford HIV database PhenoSense® assay for the isolate with PRS5B, which explains why resistance to this inhibitor was not identified in the mean-shift clustering analysis. DRV and APV PhenoSense® data were absent from the Stanford database at the time of the machine learning training. The

lower resistance reported from the cell-based PhenoSense® assay compared to that observed by kinetics with the purified enzyme reflects differences between the two assays. Such discrepancies reinforce the importance of not over-training machine learning models for drug resistance.

4.4.2 Drug-resistance mutations do not significantly affect protease sensitivity to urea

Proteolytic activity of PRS5B, PR^{S17}, and PR20 was measured for a range of urea concentrations to assess sensitivity to denaturation. Loss of proteolytic activity in the presence of urea has been shown to reflect the unfolding process[287] and provides a chemical-based assay of enzyme stability. UC_{50} values for PRS5B, PR^{S17}, and PR20 were measured to be 0.50 ± 0.21, 0.64 ± 0.18, and 0.73 ± 0.24 M, respectively, or very similar to the value measured for wild-type PR (0.7 M)[327].

For some mutants, increased sensitivity to urea denaturation is observed together with increased K_d for dimer dissociation relative to values for wild-type enzyme. For PR with single substitution of L76V (PR_{L76V}), two-fold increased susceptibility to denaturation by urea was accompanied by a 7-fold higher dimer dissociation constant (K_d) relative to wild-type enzyme, although the catalytic efficiency was unchanged [177]. PR20 also showed 5-fold higher K_d relative to wild-type enzyme[193], although PR20 and PR have similar UC_{50} values indicating no association between the different measures of stability. PR^{S17} exhibited similar K_d[177] and UC_{50} values as wild-type enzyme, despite structural evidence for fewer intersubunit contacts[118]. Variants with multiple mutations can harbor substitutions conferring compensatory effects to restore dimer stability. Therefore, unlike PR_{L76V}, the three drug resistant mutants with multiple substitutions and lower catalytic efficiency do not exhibit altered protein stability in the presence of urea.

4.4.3 Overall Structure and designation of clusters

X-ray crystal structures of PRS5B in complexes with DRV and APV were solved at 1.66 and 1.75 Å resolution, respectively. Data collection and refinement statistics are listed in Table 2. Both structures were solved in the tetragonal space group $P4_32_12$ with one dimer per asymmetric unit. Structures of PRS5B/DRV and PRS5B/APV were refined to R_{work}/R_{free} of 0.18/0.21 and 0.18/0.22, respectively. Each dimer is numbered 1-99 for subunit A and 1′-99′ for subunit B corresponding to the orientation of the single conformation for the two peptidomimetic inhibitors. In both complexes, side chains of mutations were unambiguously modeled with three exceptions. Local side chain disorder was observed near the mutations E35N and S37D in subunit A, likely due to lack of crystal contacts as observed for subunit B and the clusters of mutations leading to loss of interactions in this region. The side chain atoms of mutated residue R41K show no electron density in both subunits. The basic side chain of residue 41 occurs at a polymorphic site on the protein surface. In the absence of crystal contacts, this side chain is typically disordered in PR structures. Alternate conformations were observed for main chain residues 34 and 35 in the hinge region as described in a later section.

The dimer of PRS5B/DRV is shown in Figure 4.1 with key structural regions and mutation clusters labeled. Each protease monomer contains three major elements. The flaps (residues 42-56) control access to the active site and are important for binding of substrates and inhibitors. The dynamic flaps pivot about the hinge-loop (residues 34-41) which interacts with the main body of the protease at the 80's loop (residues 78-85) and two other β-strands. The β-strand of the protease flap continues into two-stranded β-sheet referred to here as Loop 2 (residues 61-76). Loop 2 adjoins another two-strand antiparallel β-sheet, Loop 1 (residues 10-

21), to form a 4-strand β-sheet that wraps around the sole α-helix (residues 87-93). Hydrophobic residues facing the center of this assembly create the hydrophobic core of the protein.

PRS5B contains 22 mutations, including only a single mutation (I84V) in the inhibitor binding site and two mutations (M46L and I54V) in the flaps. The other mutations fall into two main clusters, designated after the major protease regions they affect. Hinge and Loop 2 mutation clusters show coordinated effects on the hinge and Loop 2 regions of the protease respectively. Two additional mutations in Loop 1 bring about minor effects. The mutation clusters in PRS5B and in previously characterized drug resistant mutants PR^{S17} and PR20 are shown in Figure 4.1. The three PR mutants share a common cluster of mutations in the hinge region. Within each hinge mutation cluster, E35N/D and M36I are shared, although each mutant contains different accessory mutations. PR^{S17} contains a cluster of three flap mutations, M46L, G48V, and I54V, which primarily affect the conformation of the flaps[118,207]. PR20 shows an expanded inhibitor binding cavity due to a cluster of mutations (D30N, V32I, I47V, and I84V)[191,192]. Mutation L10I in Loop 1 of PRS5B was excluded from the designated clusters due to absence of significant structural changes as also seen for the same mutation in the PR^{S17}/DRV structure. In PR20, however, a phenylalanine substitution at this position is associated with loss of an intersubunit ion pair between the nearby Arg8 and Asp29′ [137,191].

The 198 Cα atoms of PRS5B/DRV dimer superimpose on equivalent atoms of wild-type PR/DRV (PDB:2IEN[304]) with a root mean square deviation (RMSD) of 0.94 Å. The hinge and Loop 2 regions of PRS5B/DRV show the largest structural changes relative to PR/DRV suggesting an effect from the local mutation clusters in those regions. Likewise, the Cα atoms of PRS5B/APV superimpose onto PR/APV dimer (PDB: 3NU3[166]) with a RMSD of 1.02 Å. PRS5B/APV exhibits larger shifts in the hinge and Loop 2 relative to PR/APV compared to

the corresponding DRV complexes. The two structures of PRS5B with DRV and APV are essentially identical and superimpose with a low RMSD of 0.16 Å.

4.4.4 PRS5B inhibitor-binding site and flaps

The inhibitors and their interactions with PRS5B are shown in Figure 4.2. The two inhibitors are chemically similar, apart from a *bis*-tetrahydrofuran (THF) moiety at P2 in DRV in place of the single THF ring of APV (Figures 4.2A and D). Both inhibitors bind to the active site of PRS5B in a single conformation, possibly due to crystallizing in the $P4_32_12$ space group, unlike many crystal structures of PR-PI complexes that show inhibitor bound in two orientations related by 180°. The hydrogen bond interactions with inhibitors (Figures 4.2B and E) are conserved in PRS5B and wild-type PR complexes. The P2 THF group of APV accepts a hydrogen bond from the backbone amine of Asp30. In DRV, the P2 *bis*-THF group conserves this interaction in addition to forming hydrogen bonds with the amine of Asp 29. However, these additional interactions of DRV relative to APV do not alter the inhibition of PRS5B (Table 4.1).

I84V is the sole mutation in PRS5B with a side chain that faces the active site cavity. This mutation contributes resistance to all PIs[164]. Substitution of the smaller valine side chain at residue 84 increases the volume of the S2/S2′ pockets and results in fewer van der Waals contacts with APV compared to the PR complex (Figures 4.2C and F).

The side chains of flap mutations M46I and I54V do not form van der Waals contacts with these inhibitors. The single substitution of I54V in PR has been described previously to alter flap conformation and effect inhibitor binding [298]. This mutation is associated with resistance for all PIs except DRV [164]. In PRS5B structures, the shorter sidechain of I54V decreases hydrophobic interactions with the Ile50 side chain on the opposite subunit. PRS5B contains a second flap mutation M46I, which is a drug resistance mutation for all clinical PIs except DRV,

SQV and TPV[164]. The side chain of flap residue 46 is adjacent to Phe53 at the protein surface, however, the shorter hydrophobic side chain of M46I shows no significant effects on the flap structure of PRS5B inhibitor complexes. M46I and I54V mutations co-occur frequently in drug resistance [337–339]. Adding M46I and I54V mutations to a RTV-resistant mutant was shown to improve catalytic efficiency while sustaining resistance [291]. Molecular dynamics simulation suggests these mutations may act by altering the flexibility of the flaps [340].

4.4.5 Hinge cluster mutations E35N and M36I coordinate with N83D to disengage flaps

Comparison of the new PRS5B structures with equivalent PR, PR^{S17}, and PR20 complexes illustrates how the addition of N83D induces more extensive loss of interactions at the hinge-body interfaces than seen for previously characterized highly-resistant mutants lacking this mutation. The PDB codes used for comparison are as follows: PR/APV=3NU3, PR/DRV=2IEN, PR^{S17}/DRV=5T2Z, PR20/DRV=3UCB, PR20/APV=4J5J. Consistent changes were observed in both PRS5B/DRV and PRS5B/APV structures. Structurally, the hinge (residues 34-41) is an anchoring point for the base of the flexible flap (residues 42-56). Mutations in the hinge region are known to increase flap flexibility and contribute to drug resistance in PR20, PR^{S17}, and other mutants[118,186,191,192,337,341].

PRS5B harbors 4 mutations in the hinge, E35N, M36I, S37D, and R41K. Only M36I is associated with resistance to ATV, IDV, NFV and TPV, while the other hinge mutations have no significant association with resistance [164]. The side chain of residue 34 shows altered interactions with mutated residue N83D in the 80's loop in the core of the dimer. With the exception of subunit A of PRS5B/DRV, hinge residues 34 and 35 display alternate conformations for the main chain, while the side chains of Glu34 and E35N show alternate conformations or disorder. Intermolecular crystal contacts partially mitigate the disorder in

subunit A. Therefore, the extent of changes arising from hinge-loop mutations is described for subunit B and compared to PR (Figure 4.3). The loss of charge due to the E35N mutation in PRS5B eliminates the ion pair in wild-type enzyme between the side chains of Glu35′ and flap residue Arg57′ (Figure 4.3A). In PRS5B/APV, loss of this ionic interaction leads to a 2.8 Å shift in positions of Cα atoms for residue 35 and 7 Å shift in side chain position compared to wild-type PR. Furthermore, this change allows the Arg57′ side chain to move to a conformation which eliminates its hinge-flap hydrogen bond interaction with the carbonyl oxygen of Met36′. The new conformation of Arg57′ also retains a hydrogen bond with Tyr59′ and introduces new intra-flap π-interaction with the side chain of Trp42′. In the absence of stabilizing crystal contacts, the side chain of E35′N shows little electron density suggesting the mutated side chain does not form any significant new interactions compared to PR.

Glu34′ is the first residue in the hinge of PR. Its carboxylate side chain faces in the opposite direction to Glu35′ to form a bifurcated hydrogen bond with the side chains of Asn83′ in the body of the protease and Lys20′ in Loop 1 (Figure 4.3B). In this way, Glu34′ and Glu35′ of the hinge link the flap to the rest of the protease. These interactions are disrupted in PRS5B due to the N83D mutation. The pH for the crystallization and enzyme assay conditions (7.9 and 5.6) likely mean Asp83 is monoprotonated. N83D introduces a negatively charged carboxylate group and eliminates the hydrogen bonds with Glu34′ side chain and with the backbone carbonyl oxygen of Glu21′. Instead, the carboxylate side chain of Glu34′ shifts by ~6 Å away from the hinge-body interface to the protein surface of PRS5B/APV structure. The void left by the shifted Glu34′ side chain is filled by a single partial-occupancy water molecule that bridges the N83′D and Lys20′ side chains.

Met36′ in PR/DRV and PR/APV structures interacts with nearby side chains of Leu33′, Leu38′, and with Ile15′ in Loop 1. The hinge mutation M36I substitutes a shorter hydrophobic side chain. As shown in Figure 4.3C, the smaller M36I residue in PRS5B structures retains significant van der Waals contacts with Ile15′ and Leu38′ and the mutated I33L sidechain due to a shift in its main chain. The Cα of Ile36 is closer to Loop 1 by 1.3 and 1.9 Å in mutant complexes with DRV and APV structures, respectively, compared to PR structures. Mutations E35N and N83D drive two alternate conformations for the main chain of the hinge-loop at residues 34′ and 35′ through decreased interactions with 80's loop and flap. By allowing a continuation of van der Waals contacts at the hinge-Loop 1 interface using a shorter side chain, M36I counterbalances the effects of E35N and N83D and stabilizing the hinge-loop. Coordinated substitutions of M36I and I33L maintain van der Waals contacts between the hinge and Loop 1, thus counteracting the loss of interactions of hinge with main chain of residue 21′ and flaps in mutant.

Comparison of the PRS5B/DRV hinge-region mutations with those in the DRV complexes with PR^{S17} and PR20 (Figures 4.3D and E) highlights the influence of the N83D mutation. Despite the presence of several mutations, introduction of a partially occupied (0.5) water molecule in PRS5B preserves many interactions of residues 20, 21, 34, 35 and 83 seen in the other two resistant mutants. All three resistant proteases have mutations for Glu35 (E35N or E35D) that break the ion pair with flap residue Arg57′ and a compensating M36I. However, PR20/DRV and PR^{S17}/DRV lack the alternate main chain conformation of hinge-loop and side chain disorder exhibited by PRS5B/DRV. In PR^{S17}/DRV, the E35D side chain forms an ion pair with K20R and a hydrogen bond with the side chain of Asn83′. PR20 does not contain a substitution for Asn83′ and thus retains the stabilizing interactions between the protease body

and the hinge-loop. PR20 utilizes a full-occupancy water molecule interacting between Asn83′ and the main chain amide of residue 35′. Unlike PRS5B, both PR^{S17} and PR20 preserve the hydrogen bond between Asn83′ and the main chain carbonyl of Glu21′ that helps stabilize the body of the protease and Loop 2 with the hinge. PR^{S17} and PR20 also have additional compensating mutations that act synergistically with M36I (K20R in PR^{S17} and I33F, I13V and I15V in PR20) to further stabilize the hinge.

4.4.6 Mutations A71V and G73T coordinate with accessory mutations L11I, A22V, L24M, and I62/64/66V to induce structural shifts in Loop 2

Mutation A71V is associated with resistance to all PIs except for DRV, APV and TPV, while G73T only shows significant resistance to ATV. Comparison of both PRS5B/PI structures with analogous wild-type complexes reveals how G73T and A71V mutations act in synergy with other accessory mutations to drive structural shifts of almost 3 Å in Loop 2 in both subunits. This section describes the PRS5B/APV complex specifically unless noted. Subunit B of PRS5B/APV is compared with wild-type structure in Figure 4.4.

Gly73 is situated at the top of Loop 2 where the Cα forms van der Waals contacts with the side chain of Leu89 (Figure 4.4A). In the wild-type PR, the large Leu89 side chain fits snugly into the hydrophobic core making contacts with Ile64, Thr31, Ile66, and Ala71. In PRS5B, Gly73 and Ala71 are replaced with larger residues, threonine and valine, respectively. The polar side chain of G73T is stabilized by hydrogen bonds with the side chain of Glu92 in a shifted conformation. In addition, subunit B of both PRS5B structures features a water molecule amidst this formerly non-polar core. This new water molecule forms a set of hydrogen bond interactions with the carbonyl oxygen of Thr74, the side chain hydroxyl groups of Thr31 and mutated Thr73, and the side chain of Asn88 (Figure 4.4B). More importantly, the larger side

chain of G73T induces a 180° flip in the bulky Leu89 side chain where the Cγ atom is ~2.5 Å from the position in PR. The flipped Leu89 side chain is directed toward the new bulkier side chain of the A71V mutation and retains hydrophobic contacts with the larger side chains of mutated Val71 and Thr73. Together, these substituted residues combine with other compensatory mutations to initiate a 1.2-3.0 Å outward shift in all Cα atoms of Loop 2 from mutated I64V through G73T.

The coordinated effects of G73T, A71V, and flipped Leu89 conformation combine with clustered accessory mutations to bring about rearrangements in the Loop 2. Larger hydrophobic residues from mutations L11I, V22I, and L24M are located deep within the hydrophobic core of the protease and near the main chain of catalytic Asp25/25′. These three mutations help to fill voids in the hydrophobic core created by the shift of Loop 2 and propagate changes from Loop 2 to the catalytic site.

The bulge in Loop 2 induced by G73T and A71V is offset by ~1 Å due to smaller side chains in the I62V, I64V and I66V mutations, which have a compensatory effect (Figure 4.4C). Three other mutations in this cluster, Q61H, I63P, and I72V, project to the protein surface and do not appear to induce changes in the interior. Thus, the large distal movements for Loop 2 stemming from A71V and G73T mutations are compensated by repacking of the hydrophobic core from other mutations in the cluster.

Consequences of Loop 2 movement are best observed at the β-turn (residues 67-69) shown in Figure 4.5A. Large deviations in Cα atoms between PRS5B and PR structures (2.6-3.1 Å) occur at Ala67. In PR structures, the β-turn of Loop 2 is secured to the rest of the protease by two interactions: an anti-parallel β-sheet hydrogen bond from the backbone amine of Ala67 to the backbone carbonyl oxygen of Thr12 (Loop 1) and an intersubunit ion pair from the side chain

of His69 to the C-terminal carboxylate of Phe99′. Both interactions are absent in PRS5B structures. The phenylalanine side chain of the C-terminal Phe99 interacts with Leu24, the amino acid adjacent to the catalytic aspartates 25/25′. The larger Met24 side chain in PRS5B displaces the side chain of Phe99′ by ~1.5 Å. The loss of interactions is significant as these two interactions contribute half of all polar contacts between Loop 2 from residues 62-73 and the rest of the protease present in the wild-type structures. It should be noted that the PRS5B structures possess no significant crystal contacts along the Loop 2 region. PRS5B structures feature mutations producing significant alternations of main chain conformation and repacking of hydrophobic side chains compared to wild-type PR structures. These changes likely result in altered dynamics that propagate to the active site and contribute to the poorer inhibition measured for clinical inhibitors.

PR^{S17} and PR20 also share the A71V mutation. However, the addition of G73T combined with the other mutations in the cluster produces a more dramatic expansion of Loop 2 in PRS5B than seen in the two other drug-resistant mutants. The tips of Loop 2 and adjacent residues are compared for DRV complexes with PRS5B, PR^{S17}, PR20 and wild-type PR in Figure 4.5B. In PRS5B, the Cα atom of Ala67′ is 2.4 and 1.8 Å farther from Loop 2 at Cα atoms of Thr12′ and residue 11′ than in PR, whereas there is no significant difference in these distances for PR^{S17} and PR20. Similarly, the distance between the C-terminus of Phe99 to the His69′ imidazole ring is ~0.5 Å farther from the PR location in PRS5B compared to PR^{S17} and PR20.

4.4.7 Mutation clusters decrease polar interactions across the entire monomer

Mutation clusters at the hinge and Loop 2 are linked via mutation E21D in Loop 1. In subunit B of PR/APV structure, the large side chain of Glu21′ forms a hydrogen bond with the hydroxyl side chain of Thr12′ linking the two β-strands of Loop 1. The shorter E21D side chain

in PRS5B mutant cannot form this hydrogen bond with Thr12′ side chain. Loop 1 residues 21 and 12 provide an interface between the corresponding mutation clusters in the hinge and Loop 2 (Figure 4.6). PRS5B has a total of six fewer polar interactions than the wild-type complex in this region spanning nearly the entire monomer from the C-terminus and tip of Loop 2 through the protease body (Loop 1 and residue 83′) to the hinge-flap interface (residues 34′-35′ to 57′).

4.4.8 Molecular dynamics simulations explore conformational variation in PRS5B and wild-type PR

Since the PRS5B structures showed only minor changes in the interactions with inhibitors compared to wild-type complexes, MD simulations were performed to evaluate the conformational variation and calculate the protease-inhibitor interaction energies. Simulations of the ligand-free enzymes were used to assess the variation in the absence of inhibitor. Inhibitor APV was removed from the starting structures prior to 20 ns MD simulations of the ligand-free enzymes. As shown in the trajectories (Figure 4.7A), the ligand-free structures equilibrate rapidly to about 2.2 Å RMSD on Cα atoms. The mutant has slightly higher RMSD values than the wild-type PR. In comparison, earlier MD simulations starting from open conformation crystal structures showed RMSDs rising to 4.5 Å for PR20 mutant and about 2 Å for wild-type PR compared to about 1.5 Å for simulations on closed conformation dimers after removing inhibitor[194,195,342].

The variation in flap conformation was assessed by calculating the distances between Cα atoms of Ile50 and 50’ at the flap tips and the catalytic Asp 25 and 25’ (Figure 4.7B). These distances fluctuate during the first 15 ns and then stabilize at about 13-14 Å for mutant and wild-type PR. The two flaps in each dimer remain closer together with the separation between Cα

atoms of Ile50 and 50' at the flap tips varying around 5.8 Å for both mutant and wild-type protease.

The variation in Cα positions of ligand-free dimers during the simulations is illustrated by the superimposed structures in Figure 4.7C. PRS5B mutant and wild-type dimers exhibit larger displacements from the crystal structures of 2 to 5 Å in the flaps, both strands of loop 1, the tip of loop 2 and the C-terminus. The mutant and wild-type PR converge into separate conformations in these regions. Large variations in loop 2 and the C-terminus were also described in a recent MD simulation by another group [343]. The flaps of the wild-type dimer shift into the active site cavity in this orientation, whereas the mutant flaps remain closer to the conformation in the crystal structure. Loop 2 varies about two separate positions observed in the two crystal structures. Several shifts, especially those encompassing the two strands of loop 1 and the C-terminus, are consistent with the loss of polar interactions across the subunit of PRS5B relative to wild-type PR observed in the crystal structures (Figure 4.6).

MD simulations of the inhibitor-bound complexes of PRS5B/DRV, PRS5B/APV and the equivalent complexes of wild-type protease were run for 10 ns to assess the effect of inhibitor. The simulations had similar trajectories reaching about 2.2 Å RMSD from the starting structures, except for a slower divergence of PR/DRV to about 2.6 Å as shown in Figure 4.8A.

The structural variation during the simulations is illustrated in Figure 4.8B. Similar to the ligand-free dimers, the inhibitor complexes show higher variation in the flaps, loops 1 and 2, and termini. In addition, the hinge shows conformational changes, especially in the APV complexes, while the two DRV complexes have larger differences in the strand from residues 6-10. The flaps of the mutant move away from the catalytic site compared to their conformation in the wild-type PR. Moreover, the inhibitors show greater variation in the simulations for mutant (RMSD on

non-hydrogen atoms of 2.3 and 1.7 Å for APV and DRV) than for wild-type PR (RMSD of 0.6 and 1.0 Å for APV and DRV). The greater fluctuation of inhibitors in the mutant is consistent with greater flexibility of residues in the binding cavity and thereby poorer inhibition. Increased fluctuation of inhibitor was also reported in a recent MD simulation by another group[344].

The non-bonded protease-inhibitor interaction energies were calculated over the 10 ns simulations (Figure 4.8C). For wild-type PR, the calculated interaction energy had a mean value of -323.0 ± 21.6 kcal/mol for DRV and -260.4 ± 15.4 for APV. The differences are in good qualitative agreement with the experimental values of 40-fold higher affinity for DRV relative to APV. Previously, another group reported close agreement with experimental values in more extensive 500 ns simulations of wild-type PR with DRV and APV[345]. In contrast, the interaction energies calculated for the mutant gave a mean of -285.3 ± 16.4 for DRV and -277.9 ± 18.0 kcal/mol for APV consistent with similar experimental inhibition values for the two inhibitors (Table 4.1). The calculations for mutant and wild-type complexes cannot be compared directly, however, since they do not account for contribution of the entropic differences in the two ligand-free enzymes and free inhibitors in solution. These differences are likely increased due to the large number of mutations (22) in PRS5B, including four changes that alter the charge of the side chain.

4.5 Discussion

PRS5B is an HIV protease mutant from a clinical isolate that is poorly inhibited by all FDA-approved PIs as shown by enzyme kinetics experiments. Extremely poor inhibition was measured for LPV (K_i=160 nM) and SQV (K_i=1 μM). The best inhibitors, APV and DRV, displayed K_i values of approximately 4 nM for PRS5B. Structural analysis of PRS5B in complexes with APV and DRV demonstrates minor alteration in the inhibitor binding site,

consistent with the presence of only I84V mutation. The distance between closest atoms of Ile47 and Val84 on opposite sides of the cavity increases by about 1 Å compared to the corresponding separation of Ile47 and Ile84 in the wild type complexes. PR20 also contains the active site mutation I84V, which works synergistically with 3 other mutations to substantially increase the size of the inhibitor binding cavity as shown by 2 Å bigger distance between the closest atoms of Val47 and Val84 and thus decrease the affinity for PIs. Like PRS5B, PR^{S17} contains only one active site mutation, V82S, which does not significantly alter the size of the inhibitor binding site or interactions with DRV relative to those of wild-type PR/DRV but may contribute to substrate recognition when combined with the G48V flap mutation[207]. These three HIV protease variants, PR20, PR^{S17} and PRS5B, display decreased affinity for DRV despite retaining all hydrogen bond interactions with inhibitor observed in wild-type PR crystal structures.

In the case of PRS5B, mutations in the hinge and Loop 2 regions induce major conformational changes in regions distal to the active site that eliminate interactions observed in wild-type PR structures. Perturbations in PRS5B did not alter interactions with its two most effective inhibitors. Instead, significant conformational changes occurred far from the active site suggesting hinge and Loop 2 rearrangements lead to loss of inhibitor affinity through PI-independent mechanisms. MD simulations also show distinct conformations for these regions in the mutant and wild-type PR.

Unlike other studied examples of highly-resistant mutants, PRS5B contains a cluster of mutations protruding into and extending along two strands of Loop 2 (Figure 4.1). Non-active-site mutations in this region, such as A71V and G73S, have been shown to propagate changes to the catalytic site and provide cross-resistance for inhibitors[118,183,197,291,339,346,347]. Another study on single mutants V32I, L33F, L76V, and L90M showed how remote mutations can impact

inhibitor binding by altering protein ensemble dynamics through a rearranged network of residues circulating to the active site[179]. Several molecular dynamics studies of proteins have shown that the effects of distal mutations can propagate to the active site and influence protein function[348,349]. These mechanisms combined with hydrophobic sliding described for HIV protease[166,172,350] are likely occurring in PRS5B. The dramatic perturbations resulting from the Loop 2 mutation cluster in PRS5B work synergistically with the hinge cluster and other mutations to decrease inhibitor affinity and catalytic activity.

N83D has long been identified as a major resistance associated mutation for TPV [351–353]. G73T is a rare minor drug resistance mutation for ATV [164,354–356]. This study is the first to describe the structural impact of these two mutations. L10I and L63P did not show structural changes in PRS5B but have been shown to help maintain thermal stability when combined with I84V[357].

NMR measurements of inhibitor-free PR^{S17} and PR20, supported by molecular dynamics simulations and crystal structures, demonstrate that these drug resistant proteases possess dynamic flaps that exist preferentially in an open conformation in the absence of inhibitor. This is consistent with Electronic Spin Resonance experiments on PR with D30N, M36I, and A71V mutations[358]. In contrast, wild-type PR tends to adopt a closed flap conformation even in the absence of ligand[118,193,359]. PRS5B shares similar hinge-loop mutations with PR^{S17} and PR20. In addition, PRS5B bears the unique N83D mutation which leads to fewer interactions between flap and hinge than observed in both PR^{S17} and PR20. MD simulations on PRS5B show larger variations in the surface features of the flaps, hinge, loop 1, the tip of loop 2 and the C-terminal residues. Overall, the conformational differences in the MD simulations are consistent with the loss of hydrogen bond interactions observed across the entire subunit in the crystal structures of

the mutant compared to the wild-type PR. These conformational changes are expected to contribute to the changes in affinity for inhibitor. Protease-inhibitor interaction energies calculated in MD simulations agree qualitatively with the experimental difference in binding affinity of wild-type enzyme and mutant for DRV compared to APV.

Comparison of the three well-characterized examples of highly resistant mutants PR20, PR^{S17} and PRS5B reveals a trajectory of how clusters of mutations might evolve to promote drug-resistance as illustrated in Figure 4.9. All three mutants utilize mutations in the hinge region at 35 and compensating M36I mutation to drive rearrangements of the hinge and alter its separation from the flap. PRS5B, with its unique N83D mutation, features two backbone conformations for a section of the hinge-loop, whereas PR^{S17} and PR20 contain additional compensating mutations in Loop 1 to stabilize the hinge. Compared to PRS5B, PR^{S17} and PR20 have increasingly more effective mechanisms to compensate for loss of hinge-flap anchor point (Glu35-Arg57 ion pair) and show increasingly worse inhibition by PIs. PRS5B's Loop 2 cluster facilitates more dramatic shifts compared to those in other mutants containing A71V, although three mutations serve to diminish these changes. Drug selection pressure may drive accumulation of mutations in protease initially towards the resistance-inducing changes in intermediate mutants like PRS5B, but over time, these viruses may be outcompeted by mutants with more synergistic mutations, such as PR^{S17}. Other highly evolved mutants might resemble PR20 by introducing mutations that enhance monomer stability and expand the active site cavity or possess co-evolved mutations in Gag substrate cleavage sites [153].

The PRS5B mutant provides a prime example of how mutation clusters in distal regions can function synergistically to decrease inhibitor effectiveness while retaining viable catalytic

efficiency and enzyme stability. Together, PRS5B, PR^{S17}, and PR20 provide prototypes for the design of novel inhibitors to combat variations of cross-resistant mechanisms.

Table 4.1: Table 1: K_i values (nM) of clinical PIs for PRS5B
PIs are listed in order of increasing K_i value for PRS5B. K_i fold change over PR calculated as K_i for mutant/K_i for wild-type.
[a]Values from Park *et al.* 2016. Biochemistry, except where noted.
[b]K_i is from Muzammil *et al.* 2007. Journal of Virology [28]
[c]K_i from Agniswamy *et al.* 2013. J. Med. Chem. and Louis et al. 2011. PNAS
Dash (-) indicates K_i is not currently published

Inhibitor	K_i for PRS5B	K_i for PR[a]	PRS5B K_i fold change	PR^{S17} K_i fold change[a]	PR20 K_i fold change[c]
APV	3.6 ± 0.8	0.2	18	55	890
DRV	4.0 ± 1.3	0.005	800	10,000	8,200
ATV	8.4 ± 0.9	0.035	240	2,000	-
TPV	8.6 ± 1.2	0.019[b]	453	-	-
NFV	15 ± 1.9	0.36	42	1,630	-
IDV	46 ± 7	0.25	184	3,200	-
LPV	160 ± 10	0.02	8,000	3,650	-
SQV	1000 ± 70	0.39	2,600	22,000	2,400

Table 4.2: Table 2: Crystallographic data collection and refinement statistics for PRS5B/ DRV and PRS5B/APV
Parenthesis indicate highest resolution shell. The lowest resolution shells were from 50 Å to 3.58 and 3.77 for PRS5B/DRV and PRS5B/APV respectively.

	Structure	PRS5B/DRV	PRS5B/APV
	PDB entry	6P9A	6P9B
Data Collection	Space group	P 4_3 2_1 2	P 4_3 2_1 2
	a, b, c (Å)	74.00, 74.00, 93.98	73.66, 73.66, 94.33
	α, β, γ (°)	90, 90, 90	90, 90, 90
	Space group	P 4_3 2_1 2	P 4_3 2_1 2
	Resolution range (Å)	50 – 1.66 (1.72-1.66)	50 – 1.75 (1.81-1.75)
	Unique Reflections	31,010 (3033)	26,189 (2497)
	Completeness (%)	99.9 (97.0)	100 (95.0)
	Redundancy	4.0 (4.1)	4.1 (3.7)
	I/σ(I)	18.1 (3.2)	20.3 (2.8)
	R_{merge} (%)	6.6 (49.1)	5.9 (48.9)
	$CC_{1/2}$ (%)	99.5 (88.8)	100 (82.1)
Refinement Statistics	R_{work} (%)	18.1	18.2
	R_{free} (%)	21.5	21.7
	Solvent atoms	157	141
	Average B-factors (Å^2)		
	Protein	23.7	26.3
	Inhibitor	16.8	20.3
	Solvent	25.8	29.8
	RMSD from ideality		
	Bond lengths (Å)	0.013	0.011
		1.98	1.90

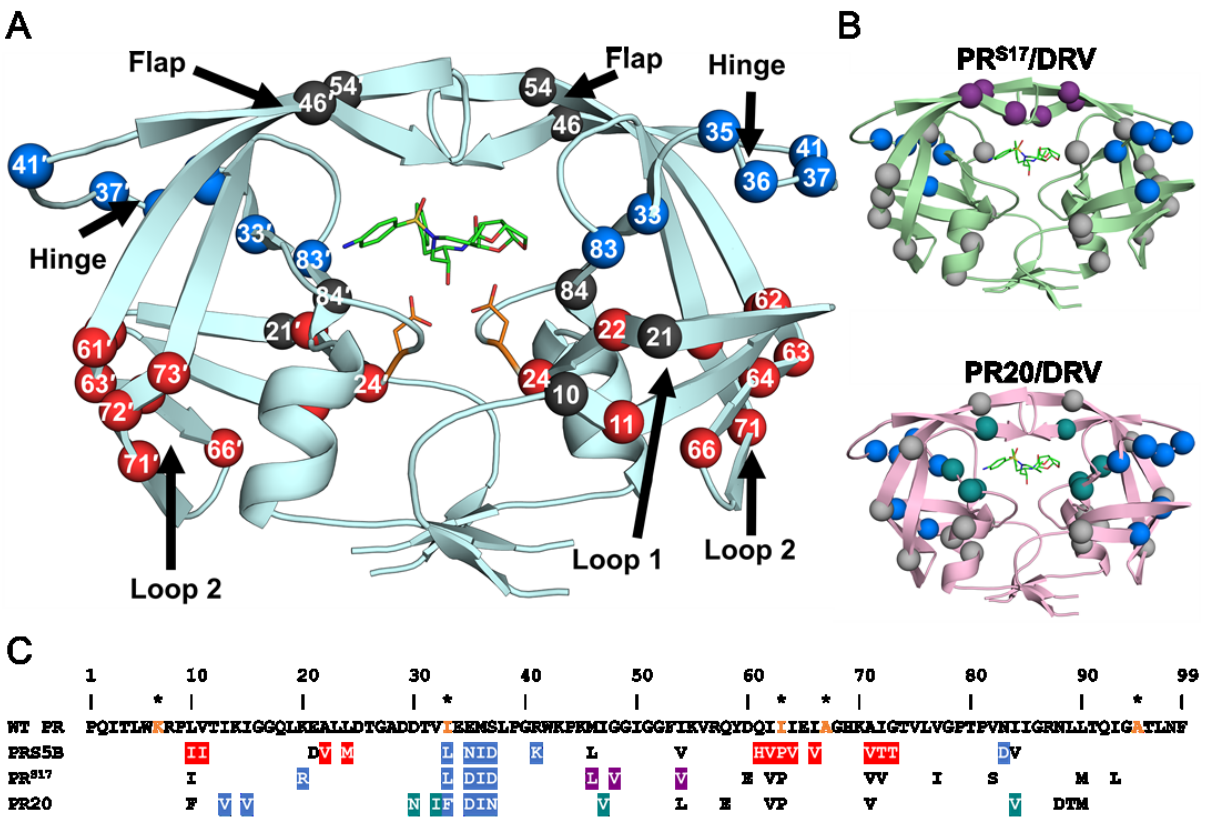


Figure 4.1: Distribution of mutations in PRS5B
A. PRS5B dimer structure in ribbon representation**.** Major regions of the protease are indicated by arrows. Catalytic Asp25/25′ and protease inhibitor DRV are shown as orange and green sticks, respectively. Sites of mutation are shown as spheres at the Cα position. Mutation clusters affecting the hinge and Loop 2 regions are colored as blue and red spheres, respectively. Active site mutation I84V, flap mutations M46L and I54V, and Loop 1 mutations L10I and E21D are shown as black spheres. Spheres hidden from view in one subunit are labeled on the opposite subunit. B. Structures of drug resistant mutants PR^{S17} (light green) and PR20 (pink) complexed with DRV showing mutation clusters. A mutation cluster affecting the hinge is shared by all three mutants as shown by blue spheres. The flap cluster in PR^{S17} is indicated by purple spheres. Mutations in PR20 altering direct inhibitor interactions are shown in teal. All other mutations are shown in grey. C. Amino acid sequence alignment of wild-type HIV-1 PR with PRS5B, PR^{S17}, and PR20. Mutated residues are colored as in figure 1A and 1B. Identical residues are omitted. HIV is a pseudo-species with polymorphic populations within an infected individual. This consensus pseudo-wild-type PR contains mutations (*) to restrict autoproteolysis (Q7K, L33I, and L63I) and cysteine-thiol oxidation (C67A and C95A) and has similar catalytic properties as HIV PR without. For this study, substitutions from the reference wild-type are considered mutations in PRS5B.

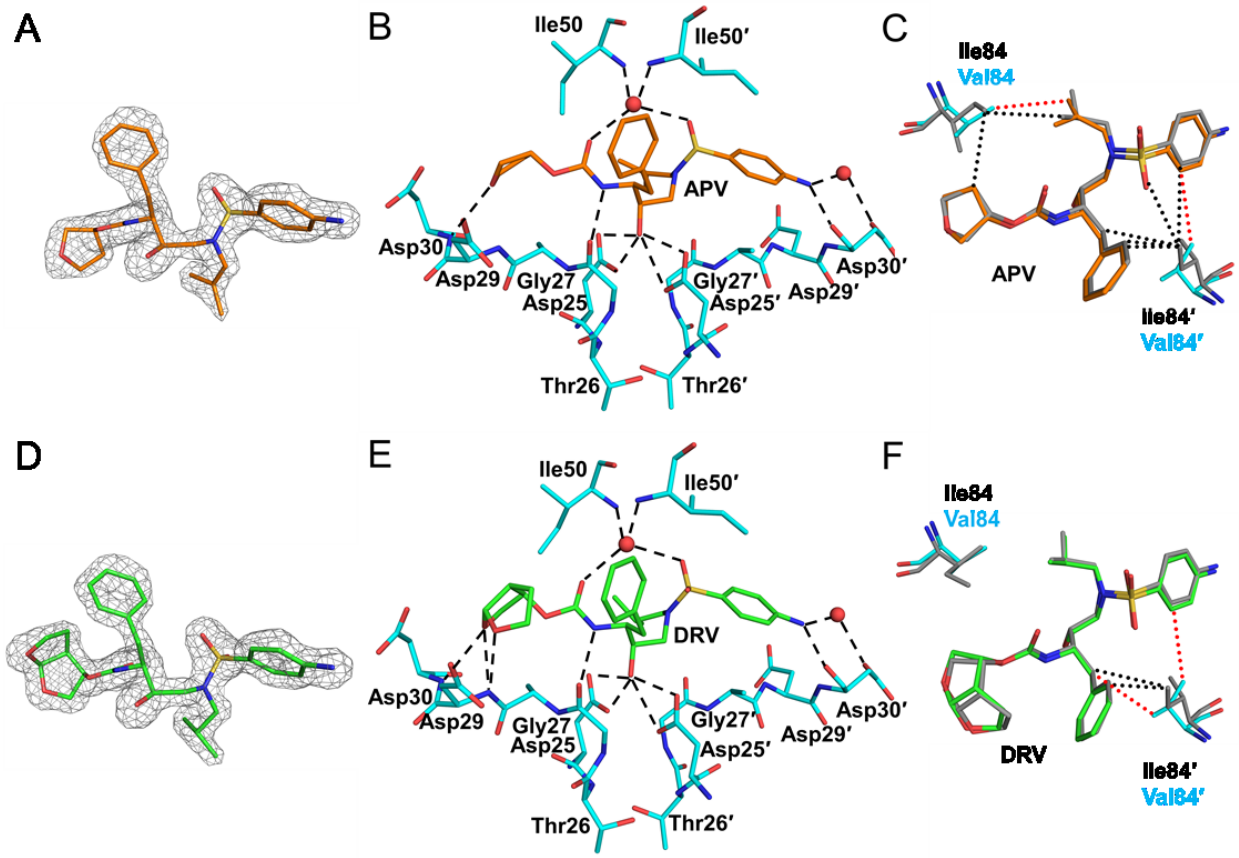


Figure 4.2: PRS5B interactions with clinical inhibitors APV and DRV
A. Omit map (grey mesh) for APV (orange sticks) B. APV hydrogen bonds with PRS5B. C. Comparison of van der Waals contacts of I84V with APV in PR and PRS5B. D. Omit map (grey mesh) for DRV (green sticks) E. DRV hydrogen bonds with PRS5B. F. Comparison of van der Waals contacts of I84V with DRV in PR and PRS5B. Omit maps are Fo-Fc maps contoured at 3σ (grey mesh) for APV and DRV. PRS5B residues are shown as cyan sticks. PR/APV and PR/DRV are shown as grey sticks. Water molecules are red spheres. Van der Waals interactions are shown as dotted lines in red and black for PRS5B and PR, respectively.

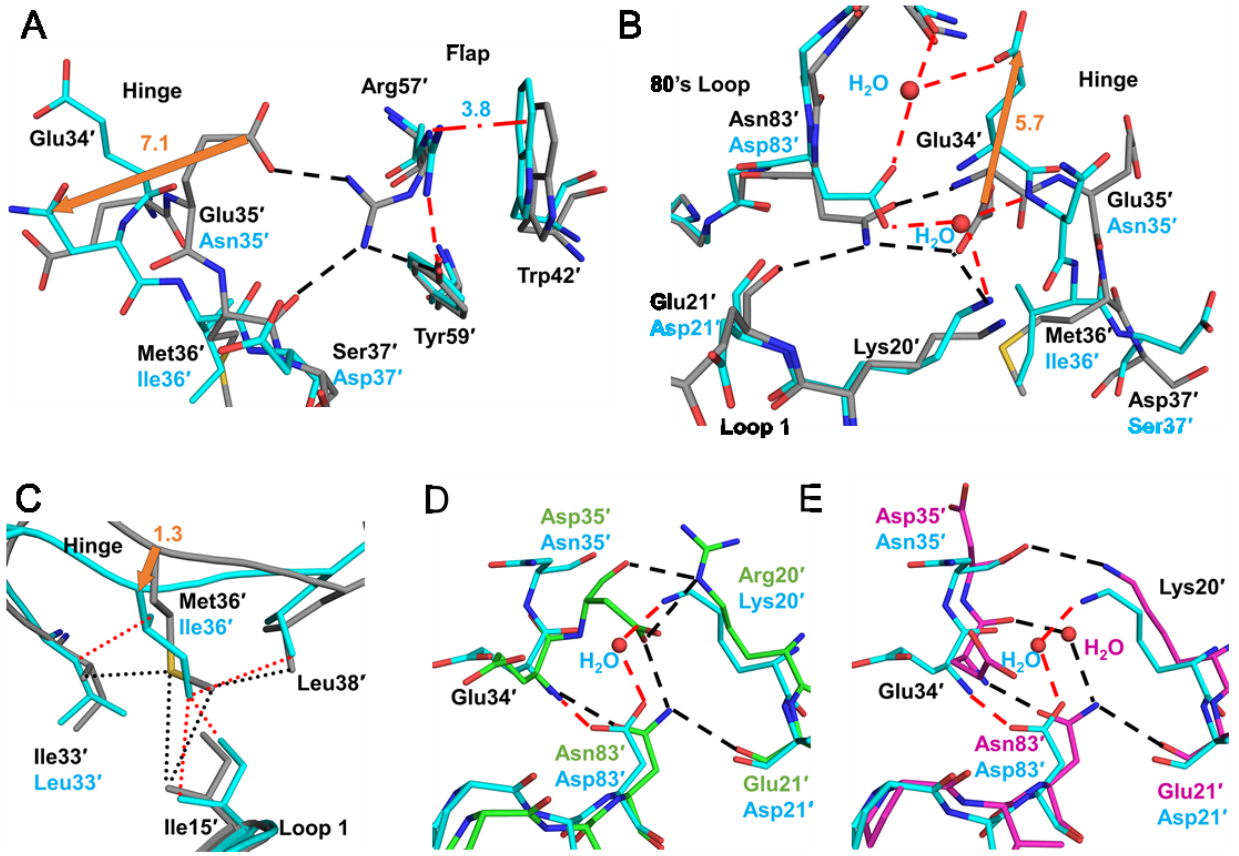


Figure 4.3: Conformational changes as a result of hinge cluster mutations E35N, M36I, and N83D

A. Hinge interactions for APV structures of PR (grey) and PRS5B (cyan) and conformational change due to E35′N. E35′N mutation eliminates ion pair with Arg57′. In PRS5B, altered conformation of Arg57′ forms coplanar π-interactions with Trp42′ (red dash-dots) and a hydrogen bond with Tyr59′. B. Rearrangements in the PRS5B hinge. Shifts due to mutations N83′D and E21′D alter the distribution of hydrogen bonds and introduce partial occupancy water molecules. C. M36′I mutation retains hydrophobic contacts (dots) with sidechains of I33′L, Ile′15, and Leu38′ due to shift in hinge main chain confirmation. D. Hinge of PRS5B/DRV (cyan) compared to PR^{S17}/DRV (green). E. Hinge of PRS5B/DRV (cyan) compared to PR20/DRV (magenta). PR/APV (grey) and PRS5B/APV (cyan) carbons are shown as sticks. Hydrogen bonds are shown as dashes colored red for PRS5B and black for PR, PR^{S17}, and PR20 structures. Waters are shown as red spheres. Orange arrows indicate major shifts with distances in Å. Alternate conformations were omitted for clarity.

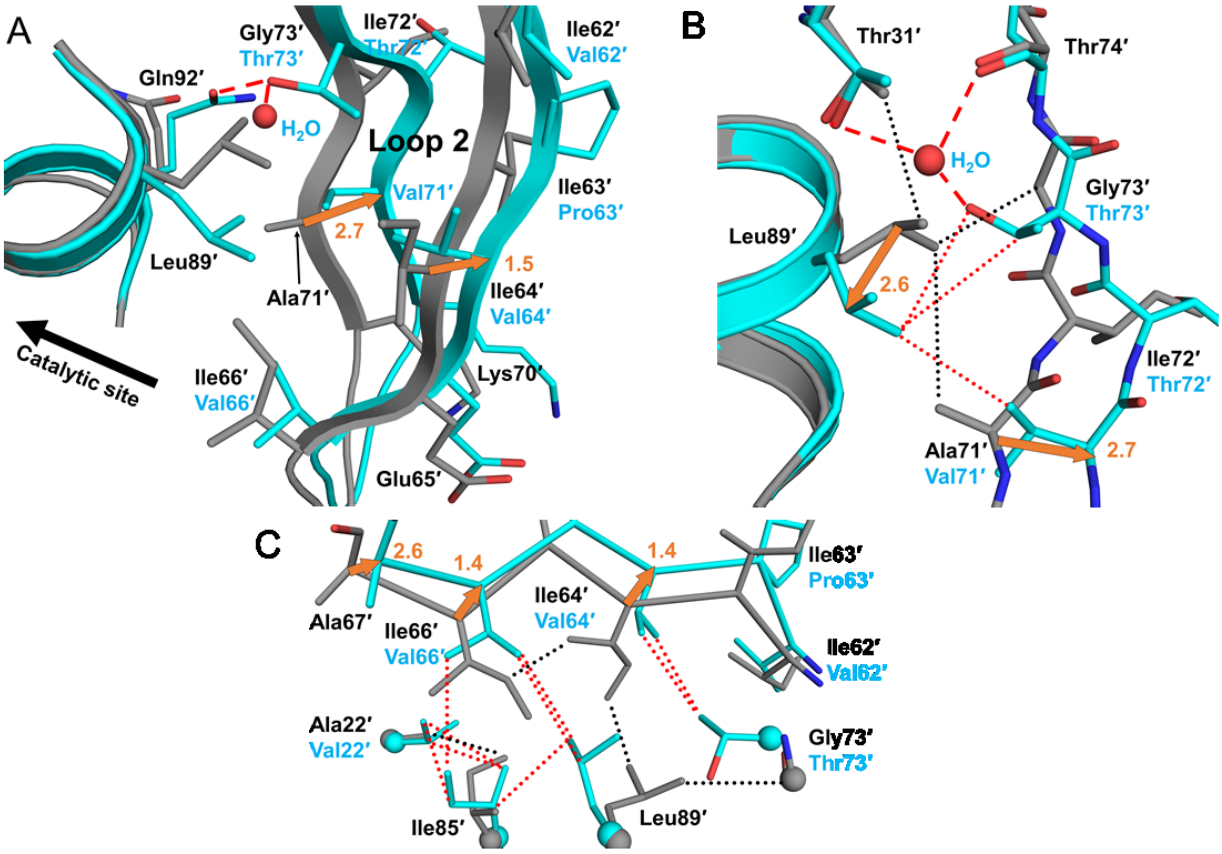


Figure 4.4: G73T and A71L drive distal movement of Loop 2 through flip in Leu89 sidechain A. B-subunit Loop 2 of PR/APV (grey cartoon) compared to PRS5B/APV (cyan cartoon) with G73′T stabilizing H-bonds shown as red dashed lines. G73′T induces flipped conformation of Leu89′ which combines with larger A71′V side chain to rearrange Loop 2. B. New water molecule coordinates with Thr31′, Thr74′ and mutated G73′T. Flipped conformation of Leu89′ faces the valine side chain of A71′V mutation. Hydrogen bonds shown as dashed lines and van der Waals contacts as dotted lines. Side chains of Asn92′ and Asn88′ were omitted for clarity. C. Smaller hydrophobic side chains for residues 62′, 64′, and 66′ dampen distal movement of Loop 2. Distances in Å are indicated as orange arrows.

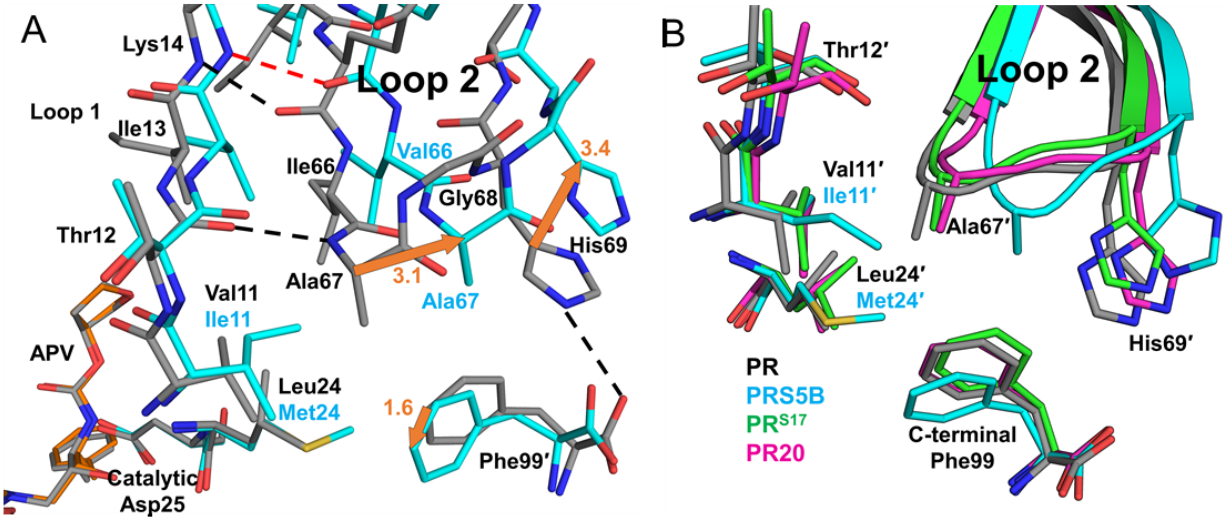


Figure 4.5: Interactions of Loop 2 tip and comparison of three highly resistant mutants PRS5B, PR^{S17}, and PR20 with wild-type PR

A. B-subunit beta-loop of Loop 2 (residues 67-69) interacting with Loop 1 (residues 24 and 11-14 shown) and the C-terminus of A subunit. PRS5B/APV is shown in cyan sticks. PR/APV is grey sticks. Hydrogen bonds and ion-pair are shown as dashed lines. Distances in Å are indicated as orange arrows. B. PRS5B Loop 2 tip comparison with drug resistant mutants PR^{S17} and PR20. Protease/DRV cartoon complexes colored as follows: Wild-type PR (grey), PRS5B (cyan), PR^{S17} (green), and PR20 (magenta). Side chains for tip residues and interacting residues, including PRS5B mutations L24M and V11I are shown as sticks.

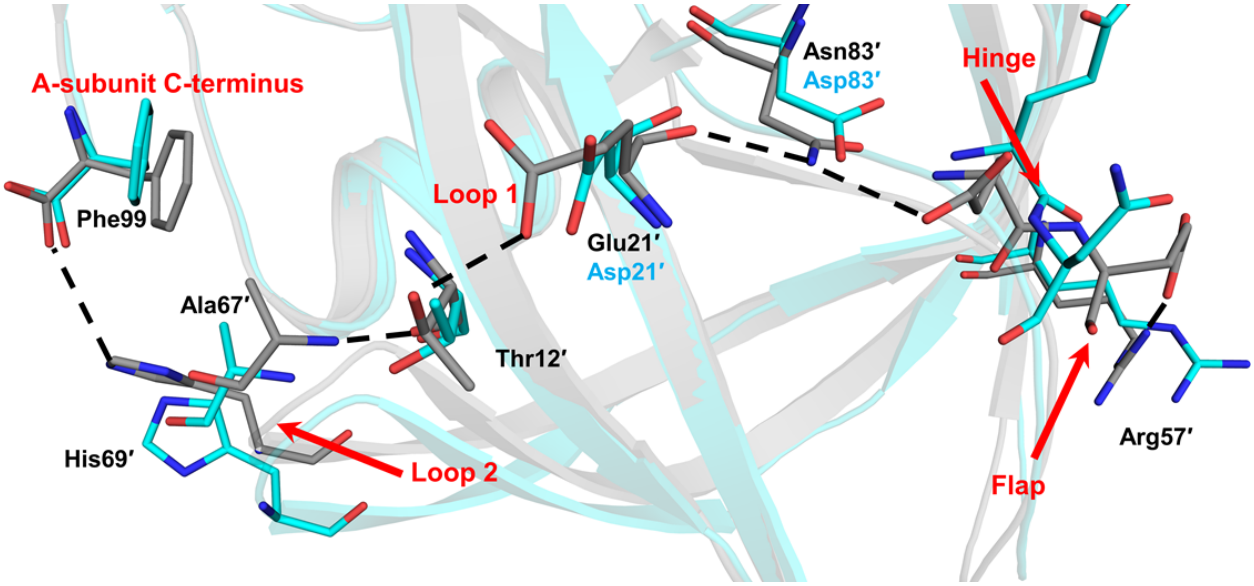


Figure 4.6: Mutations in PRS5B eliminate polar interactions spanning the length of the PRS5B monomer from the terminus to the flap at Arg57.
Hydrogen bonds present in PR/APV structure and absent in the PRS5B/APV structure are shown as black dashes. Notable regions of the protein are labeled in red. Alternate conformations omitted for clarity. PR/APV (3NU3) and PRS5B/APV are shown as grey and cyan sticks, respectively.

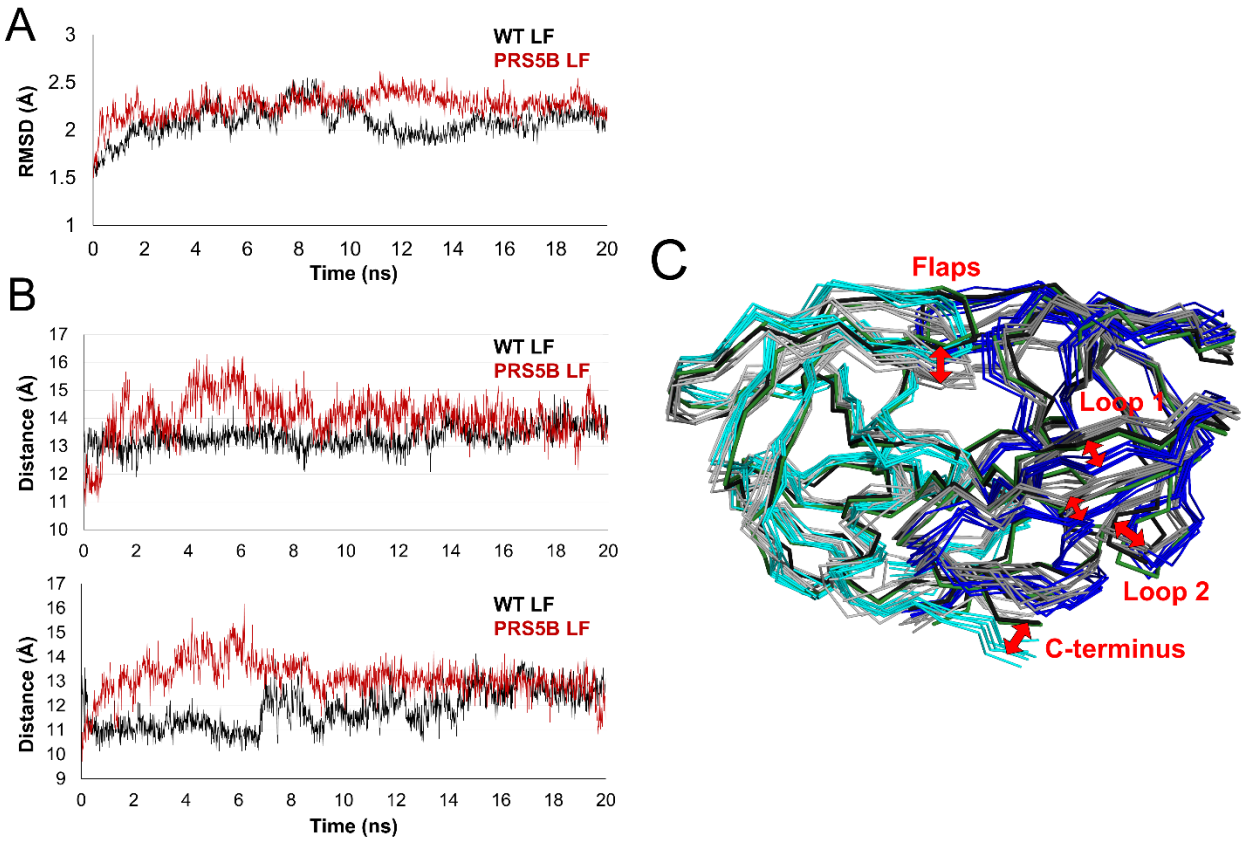


Figure 4.7: Molecular dynamics simulations of ligand-free PR and PRS5B show conformational variation in the flaps, loops 1 and 2, and termini

A. Trajectories of the simulations of ligand-free dimers. RMSD of Cα atoms after superposition on the starting crystal structure plotted for the 20 ns time course. PRS5B LF and PR LF simulations are colored in red and black respectively.

B. Distance between Cα atoms of the flap tip (Ile50) and the catalytic Asp25 for subunit A (top) and subunit B (bottom) for the time course of the simulation. Plots are colored as in A.

C. Comparison of PR and PRS5B ligand-free simulations. Snapshots of PR (grey shades) and PRS5B (cyan and blue) models for every 3 ns of the simulation shown as superimposed ribbons. The starting crystal structures of PR (black) and PRS5B (green) are shown. Labeled regions and red arrows indicate large conformational differences between wild-type and mutant protease.

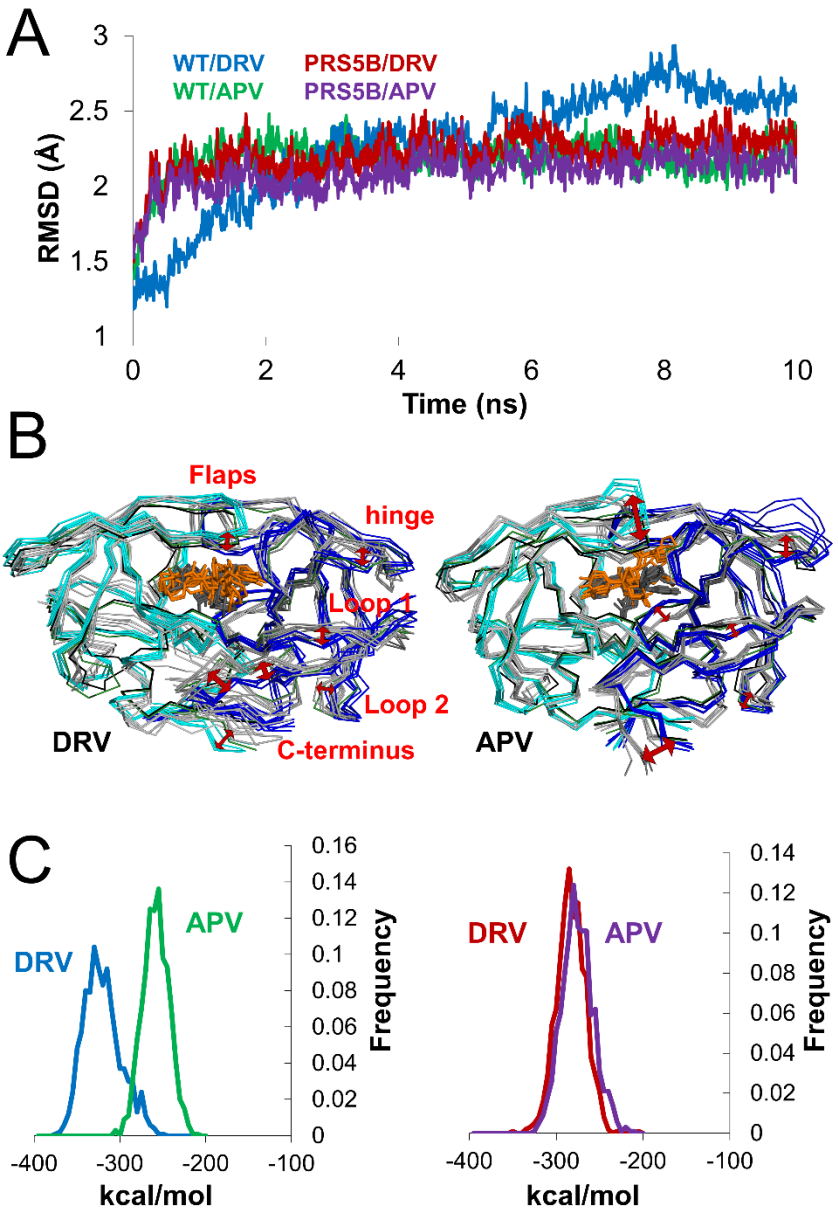


Figure 4.8: Molecular dynamics simulations of inhibitor-bound PR and PRS5B give conformational variation and calculated non-bonded interaction energies of inhibitors
A. Trajectories of the 10 ns simulations of APV and DRV complexes of PR and PRS5B. RMSD of Cα atoms after superposition with the starting crystal structure plotted for the 10 ns time course. Simulations are colored as follows: PR/DRV (blue), PR/APV (green), PRS5B/DRV (red), and PRS5B/APV (purple).
B. Frequency plot of total non-bonded interaction energies of inhibitors in PR (left) and PRS5B (right) simulations. Simulations are colored as in A.
C. Comparison of PR with PRS5B for DRV (left) and APV (right) simulations. Snapshots of PR (grey shades, grey inhibitor as sticks) and PRS5B (cyan and blue, orange inhibitor as sticks) models for every 2 ns of the simulation are shown as superimposed ribbons. The starting crystal structures PR (black) and PRS5B (green) are shown. Red arrows indicate large conformational differences between wild-type and mutant protease

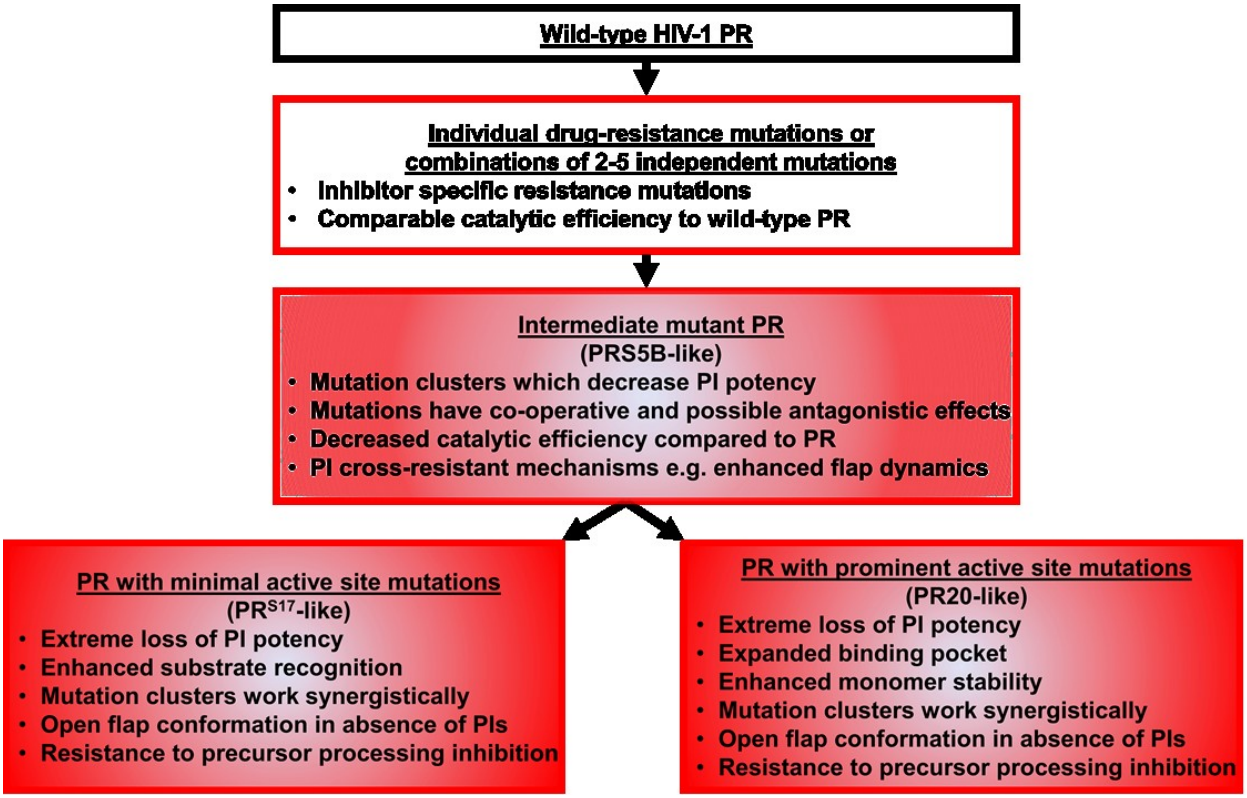


Figure 4.9: Scheme for evolution of highly-drug resistant protease variants.
Antiretroviral treatment with PR inhibitors selects for PR variants able to process substrate in the presence of inhibitors. PRS5B represents an example of an intermediate mutant protease between the initial accumulation of resistance-associated mutations and more highly-evolved protease mutants.

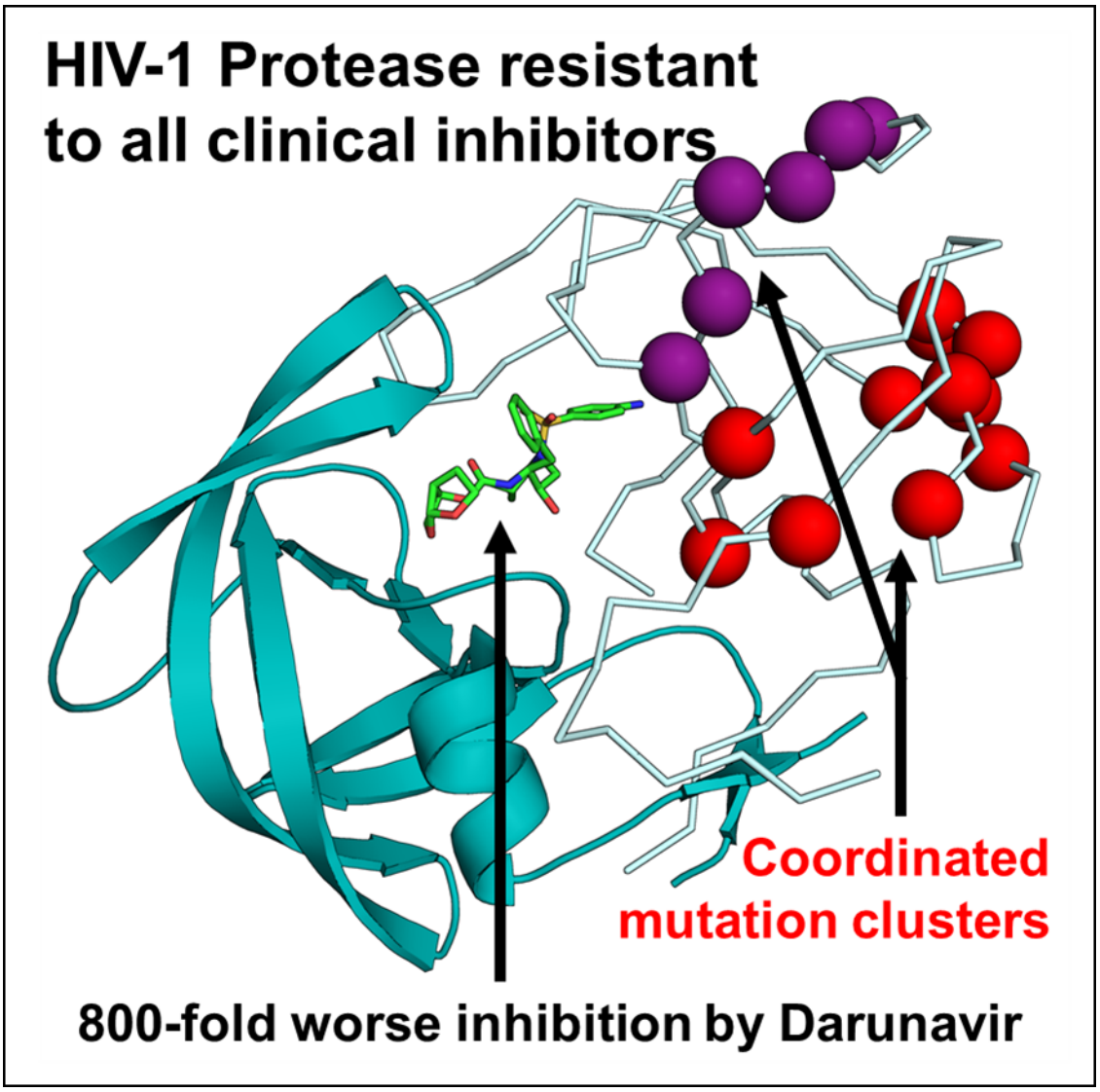


Figure 4.10: Graphical Abstract
A mutant of HIV protease selected for high-level drug resistance by machine-learning on genotype-phenotype data was analyzed for the molecular basis of resistance. Clinical inhibitors are 20 to 8000-fold less effective on this mutant relative to wild-type enzyme. Crystal structures reveal conserved interactions with inhibitor. Instead, coordinated rearrangements due to two clusters of mutations eliminate internal interactions likely contributing to resistance. We propose this mutant represents an intermediate in evolution of highly-resistant HIV-1 proteases.

5 CONCLUSIONS

HIV PR is a significant drug target in the global fight against the HIV/AIDS pandemic. Despite the success of antiretroviral therapy, drug resistance remains a threat in the wake of greater accessibility to ART. Structure-guided design of PIs has produced effective drugs such as DRV with picomolar affinity for PR. However, a greater understanding of the molecular mechanisms employed by DRV-resistant mutants, such as PR20, PRS17, and PRS5B, is important in order to devise new structure-guided design strategies to stem the increasing trend of drug-resistant PRs. This research expands our understanding of how mutations in these three highly-resistant PRs clinical isolates cooperate to lower susceptibility to PIs.

PR20 is an excellent prototype to test the inhibition of investigative compounds on due to its expanded ligand-binding site, highly mobile flaps, and greater dimer dissociation constant. A new compound, GRL-142, showed 16-fold better inhibition than DRV of PR20. Comparison of PR20/GRL142 and PR20/DRV crystal structures revealed that the bulkier P2 and P2′ groups on GRL-142 fill the expanded ligand-binding site of PR20 and the di-fluorinated P1-benzyl introduced a fluorine-bond bridge and restored an important inter-subunit ion pair. These insights show how strategies for drug-design incorporating larger and halogenated moieties onto the DRV scaffold can inhibit highly-mutated PR variants better than does DRV.

Viable Gag and Gag-Pol processing in the presence of inhibitors is required for positive selection of PR mutants, such as PRS17. Substrate analogs of Gag cleavage sites p2-NC and CA-p2 exhibited 3-fold better inhibition of PRS17 compared to wild-type enzyme. Crystallographic analysis of PRS17 complexed with p2-NC and CA-p2 revealed how mutations G48V and V82S contribute to improved interactions with substrate. This study gives insight into the dual roles of

G48V and V82S as mutations that reduce affinity for inhibitors whilst allowing for more favorable interactions with substrates.

The clinical isolate PRS5B provides an example of a mutant harboring distal mutations with minimal changes in the ligand-binding site. Enzyme kinetics measurements confirmed that PRS5B shows intermediate levels of inhibition by clinical PIs while retaining proteolytic efficiency and stability in the presence of urea comparable to PRS17 and PR20. Analysis of crystal structures of PRS5B bound to APV and DRV suggested that mutation clusters in the hinge-loop and Loop 2 regions of PRS5B eliminate interactions throughout the monomer. Interestingly, the combinations of mutations in these two clusters exhibit both synergistic and antagonistic effects. Molecular dynamics simulations of ligand-free PRS5B confirm the conformational variations in the flap, Loop 2, and termini as shown in the crystal structures. In the context of other studied variants, these results suggest PRS5B may be representative of an intermediate mutant in the selection process towards more synergistic or highly evolved drug-resistant PR mutants.

Crystallographic and enzyme kinetic studies of highly-resistant mutants of PR illustrate the molecular mechanisms involved in drug resistance. Structural studies by NMR, crystallography, and MD simulations demonstrate that drug-resistant PR mutants possess altered conformational dynamics and structural rearrangements. Future work is needed to better explain how these changes in protein dynamics in mutants alter drug binding and dissociation. More detailed MD simulations combined with neutron vibrational spectroscopy may provide new insights into the vibrational dynamics of HIV PR mutants bound with substrate or inhibitors in femtosecond to picosecond timescales[360–363]. Understanding the details by which mutations can

directly and indirectly affect inhibitor affinity for the PR ligand-binding site is crucial for efforts to design more advanced therapeutics.

To reach current UNAIDS 90-90-90 targets, major efforts to provide ART to the approximately 15 million HIV-infected people not receiving it will undoubtedly have impacts on the prevalence of drug resistance[364]. Structure-based drug design has been successful in producing the first and second generation of clinical protease inhibitors. The future of structure-based drug design for PIs should combine traditional X-ray crystallography with new ligand-design strategies[365] and computational methods[366] in order to combat the threat of HIV drug resistance.

REFERENCES

(1) Burton, D. R. Advancing an HIV Vaccine; Advancing Vaccinology. *Nat. Rev. Immunol.* **2019**, *19* (2), 77–78. https://doi.org/10.1038/s41577-018-0103-6.

(2) Hütter, G.; Nowak, D.; Mossner, M.; Ganepola, S.; Müßig, A.; Allers, K.; Schneider, T.; Hofmann, J.; Kücherer, C.; Blau, O.; et al. Long-Term Control of HIV by *CCR5* Delta32/Delta32 Stem-Cell Transplantation. *N. Engl. J. Med.* **2009**, *360* (7), 692–698. https://doi.org/10.1056/NEJMoa0802905.

(3) Allers, K.; Hütter, G.; Hofmann, J.; Loddenkemper, C.; Rieger, K.; Thiel, E.; Schneider, T. Evidence for the Cure of HIV Infection by CCR5Δ32/Δ32 Stem Cell Transplantation. *Blood* **2011**, *117* (10), 2791–2799. https://doi.org/10.1182/blood-2010-09-309591.

(4) Gupta, R. K.; Abdul-Jawad, S.; McCoy, L. E.; Mok, H. P.; Peppa, D.; Salgado, M.; Martinez-Picado, J.; Nijhuis, M.; Wensing, A. M. J.; Lee, H.; et al. HIV-1 Remission Following CCR5Δ32/Δ32 Haematopoietic Stem-Cell Transplantation. *Nature* **2019**, *568* (7751), 244–248. https://doi.org/10.1038/s41586-019-1027-4.

(5) WHO | Data and Statistics. *WHO* **2018**.

(6) Ayouba, A.; Akoua-Koffi, C.; Calvignac-Spencer, S.; Esteban, A.; Locatelli, S.; Li, H.; Li, Y.; Hahn, B. H.; Delaporte, E.; Leendertz, F. H.; et al. Evidence for Continuing Cross-Species Transmission of SIVsmm to Humans: Characterization of a New HIV-2 Lineage in Rural Côte d'Ivoire. *AIDS* **2013**, *27* (15), 2488–2491. https://doi.org/10.1097/01.aids.0000432443.22684.50.

(7) Worobey, M.; Gemmel, M.; Teuwen, D. E.; Haselkorn, T.; Kunstman, K.; Bunce, M.; Muyembe, J.-J.; Kabongo, J.-M. M.; Kalengayi, R. M.; Marck, E. Van; et al. Direct Evidence of Extensive Diversity of HIV-1 in Kinshasa by 1960. *Nature* **2008**, *455* (7213),

661. https://doi.org/10.1038/NATURE07390.

(8) Faria, N. R.; Rambaut, A.; Suchard, M. A.; Baele, G.; Bedford, T.; Ward, M. J.; Tatem, A. J.; Sousa, J. D.; Arinaminpathy, N.; Pépin, J.; et al. HIV Epidemiology. The Early Spread and Epidemic Ignition of HIV-1 in Human Populations. *Science* **2014**, *346* (6205), 56–61. https://doi.org/10.1126/science.1256739.

(9) Masur, H.; Michelis, M. A.; Greene, J. B.; Onorato, I.; Vande Stouwe, R. A.; Holzman, R. S.; Wormser, G.; Brettman, L.; Lange, M.; Murray, H. W.; et al. An Outbreak of Community-Acquired *Pneumocystis Carinii* Pneumonia. *N. Engl. J. Med.* **1981**, *305* (24), 1431–1438. https://doi.org/10.1056/NEJM198112103052402.

(10) Hymes, K. B.; Cheung, T.; Greene, J. B.; Prose, N. S.; Marcus, A.; Ballard, H.; William, D. C.; Laubenstein, L. J. Kaposi's Sarcoma in Homosexual Men-a Report of Eight Cases. *Lancet (London, England)* **1981**, *2* (8247), 598–600. https://doi.org/10.1016/s0140-6736(81)92740-9.

(11) Centers for Disease Control (CDC). Kaposi's Sarcoma and Pneumocystis Pneumonia among Homosexual Men--New York City and California. *MMWR. Morb. Mortal. Wkly. Rep.* **1981**, *30* (25), 305–308.

(12) Vilaseca, J.; Arnau, J. M.; Bacardi, R.; Mieras, C.; Serrano, A.; Navarro, C. Kaposi's Sarcoma and Toxoplasma Gondii Brain Abscess in a Spanish Homosexual. *Lancet (London, England)* **1982**, *1* (8271), 572. https://doi.org/10.1016/s0140-6736(82)92086-4.

(13) Francioli, P.; Vogt, M.; Schädelin, J.; Clément, F.; Russi, E.; Delacrétaz, F.; Perret, C.; Glauser, M. P. [Acquired Immunologic Deficiency Syndrome, Opportunistic Infections and Homosexuality. Presentation of 3 Cases Studied in Switzerland]. *Schweiz. Med. Wochenschr.* **1982**, *112* (47), 1682–1687.

(14) Rozenbaum, W.; Coulaud, J. P.; Saimot, A. G.; Klatzmann, D.; Mayaud, C.; Carette, M. F. Multiple Opportunistic Infection in a Male Homosexual in France. *Lancet (London, England)* **1982**, *1* (8271), 572–573. https://doi.org/10.1016/s0140-6736(82)92087-6.

(15) Current Trends Update on Acquired Immune Deficiency Syndrome (AIDS) --United States. *CDC MMWR* **1982**, *31* (37), 507-508,513-514.

(16) Barré-Sinoussi, F.; Chermann, J. C.; Rey, F.; Nugeyre, M. T.; Chamaret, S.; Gruest, J.; Dauguet, C.; Axler-Blin, C.; Vézinet-Brun, F.; Rouzioux, C.; et al. Isolation of a T-Lymphotropic Retrovirus from a Patient at Risk for Acquired Immune Deficiency Syndrome (AIDS). *Science* **1983**, *220* (4599), 868–871.

(17) Gallo, R. C.; Sarin, P. S.; Gelmann, E. P.; Robert-Guroff, M.; Richardson, E.; Kalyanaraman, V. S.; Mann, D.; Sidhu, G. D.; Stahl, R. E.; Zolla-Pazner, S.; et al. Isolation of Human T-Cell Leukemia Virus in Acquired Immune Deficiency Syndrome (AIDS). *Science* **1983**, *220* (4599), 865–867.

(18) Popovic, M.; Sarin, P. S.; Robert-Gurroff, M.; Kalyanaraman, V. S.; Mann, D.; Minowada, J.; Gallo, R. C. Isolation and Transmission of Human Retrovirus (Human t-Cell Leukemia Virus). *Science* **1983**, *219* (4586), 856–859. https://doi.org/10.1126/science.6600519.

(19) Sarngadharan, M.; Popovic, M.; Bruch, L.; Schupbach, J.; Gallo, R. Antibodies Reactive with Human T-Lymphotropic Retroviruses (HTLV-III) in the Serum of Patients with AIDS. *Science (80-.).* **1984**, *224* (4648), 506–508. https://doi.org/10.1126/SCIENCE.6324345.

(20) Popovic, M.; Sarngadharan, M. G.; Read, E.; Gallo, R. C. Detection, Isolation, and Continuous Production of Cytopathic Retroviruses (HTLV-III) from Patients with AIDS

and Pre-AIDS. *Science* **1984**, *224* (4648), 497–500. https://doi.org/10.1126/science.6200935.

(21) Reilly, B. *Disaster and Human History : Case Studies in Nature, Society and Catastrophe*; McFarland & Co, 2009.

(22) Veronka, G. *History of AIDS*; Lulu.com, 2013.

(23) Fischl, M. A.; Richman, D. D.; Grieco, M. H.; Gottlieb, M. S.; Volberding, P. A.; Laskin, O. L.; Leedom, J. M.; Groopman, J. E.; Mildvan, D.; Schooley, R. T.; et al. The Efficacy of Azidothymidine (AZT) in the Treatment of Patients with AIDS and AIDS-Related Complex. *N. Engl. J. Med.* **1987**, *317* (4), 185–191. https://doi.org/10.1056/NEJM198707233170401.

(24) Centers for Disease Control and Prevention (CDC). Update: Mortality Attributable to HIV Infection among Persons Aged 25-44 Years--United States, 1994. *MMWR. Morb. Mortal. Wkly. Rep.* **1996**, *45* (6), 121–125.

(25) Kitchen, V. S.; Skinner, C.; Ariyoshi, K.; Weber, J. N.; Pinching, A. J.; Lane, E. A.; Duncan, I. B.; Burckhardt, J.; Burger, H. U.; Bragman, K. Safety and Activity of Saquinavir in HIV Infection. *Lancet* **1995**, *345* (8955), 952–955. https://doi.org/10.1016/S0140-6736(95)90699-1.

(26) Pokorná, J.; Machala, L.; Rezáčová, P.; Konvalinka, J. Current and Novel Inhibitors of HIV Protease. *Viruses* **2009**, *1* (3), 1209–1239. https://doi.org/10.3390/v1031209.

(27) Scott, H.; Volberding, P. A. HIV Screening and Preexposure Prophylaxis Guidelines. *JAMA* **2019**, *321* (22), 2172. https://doi.org/10.1001/jama.2019.2590.

(28) Saag, M. S.; Benson, C. A.; Gandhi, R. T.; Hoy, J. F.; Landovitz, R. J.; Mugavero, M. J.; Sax, P. E.; Smith, D. M.; Thompson, M. A.; Buchbinder, S. P.; et al. Antiretroviral Drugs

for Treatment and Prevention of HIV Infection in Adults. *JAMA* **2018**, *320* (4), 379. https://doi.org/10.1001/jama.2018.8431.

(29) Eisinger, R. W.; Dieffenbach, C. W.; Fauci, A. S. HIV Viral Load and Transmissibility of HIV Infection. *JAMA* **2019**, *321* (5), 451. https://doi.org/10.1001/jama.2018.21167.

(30) McGrath, J. W.; Winchester, M. S.; Kaawa-Mafigiri, D.; Walakira, E.; Namutiibwa, F.; Birungi, J.; Ssendegye, G.; Nalwoga, A.; Kyarikunda, E.; Kisakye, S.; et al. Challenging the Paradigm: Anthropological Perspectives on HIV as a Chronic Disease. *Med. Anthropol.* **2014**, *33* (4), 303–317. https://doi.org/10.1080/01459740.2014.892483.

(31) 90–90–90 - An ambitious treatment target to help end the AIDS epidemic | UNAIDS https://www.unaids.org/en/resources/documents/2017/90-90-90 (accessed Jun 28, 2019).

(32) UNAIDS. Global HIV & AIDS statistics — 2019 fact sheet | UNAIDS https://www.unaids.org/en/resources/fact-sheet (accessed Jul 31, 2019).

(33) WHO | HIV Drug Resistance Report 2017. *WHO* **2017**.

(34) Luo, X.; Mo, L.; Su, G.; Huang, J.; Wu, J.; Su, H.; Huang, W.; Luo, S.; Ni, Z. Incidence and Types of HIV-1 Drug Resistance Mutation among Patients Failing First-Line Antiretroviral Therapy. *J. Pharmacol. Sci.* **2019**, *139* (4), 275–279. https://doi.org/10.1016/J.JPHS.2018.11.016.

(35) Patel, P.; Borkowf, C. B.; Brooks, J. T.; Lasry, A.; Lansky, A.; Mermin, J. Estimating Per-Act HIV Transmission Risk: A Systematic Review. *AIDS* **2014**, *28* (10), 1509–1519. https://doi.org/10.1097/QAD.0000000000000298.

(36) Patel, P.; Borkowf, C. B.; Brooks, J. T.; Lasry, A.; Lansky, A.; Mermin, J. Estimating Per-Act HIV Transmission Risk: A Systematic Review. *AIDS* **2014**, *28* (10), 1509–1519. https://doi.org/10.1097/QAD.0000000000000298.

(37) Eisinger, R. W.; Dieffenbach, C. W.; Fauci, A. S. HIV Viral Load and Transmissibility of HIV Infection. *JAMA* **2019**, *321* (5), 451. https://doi.org/10.1001/jama.2018.21167.

(38) Cohen, M. S.; Chen, Y. Q.; McCauley, M.; Gamble, T.; Hosseinipour, M. C.; Kumarasamy, N.; Hakim, J. G.; Kumwenda, J.; Grinsztejn, B.; Pilotto, J. H. S.; et al. Antiretroviral Therapy for the Prevention of HIV-1 Transmission. *N. Engl. J. Med.* **2016**, *375* (9), 830–839. https://doi.org/10.1056/NEJMoa1600693.

(39) Rodger, A. J.; Cambiano, V.; Bruun, T.; Vernazza, P.; Collins, S.; van Lunzen, J.; Corbelli, G. M.; Estrada, V.; Geretti, A. M.; Beloukas, A.; et al. Sexual Activity Without Condoms and Risk of HIV Transmission in Serodifferent Couples When the HIV-Positive Partner Is Using Suppressive Antiretroviral Therapy. *JAMA* **2016**, *316* (2), 171. https://doi.org/10.1001/jama.2016.5148.

(40) Bavinton, B. R.; Pinto, A. N.; Phanuphak, N.; Grinsztejn, B.; Prestage, G. P.; Zablotska-Manos, I. B.; Jin, F.; Fairley, C. K.; Moore, R.; Roth, N.; et al. Viral Suppression and HIV Transmission in Serodiscordant Male Couples: An International, Prospective, Observational, Cohort Study. *Lancet HIV* **2018**, *5* (8), e438–e447. https://doi.org/10.1016/S2352-3018(18)30132-2.

(41) Rodger, A. J.; Cambiano, V.; Bruun, T.; Vernazza, P.; Collins, S.; Degen, O.; Corbelli, G. M.; Estrada, V.; Geretti, A. M.; Beloukas, A.; et al. Risk of HIV Transmission through Condomless Sex in Serodifferent Gay Couples with the HIV-Positive Partner Taking Suppressive Antiretroviral Therapy (PARTNER): Final Results of a Multicentre, Prospective, Observational Study. *Lancet* **2019**, *0* (0). https://doi.org/10.1016/S0140-6736(19)30418-0.

(42) Attia, S.; Egger, M.; Müller, M.; Zwahlen, M.; Low, N. Sexual Transmission of HIV

According to Viral Load and Antiretroviral Therapy: Systematic Review and Meta-Analysis. *AIDS* **2009**, *23* (11), 1397–1404. https://doi.org/10.1097/QAD.0b013e32832b7dca.

(43) Quinn, T. C.; Wawer, M. J.; Sewankambo, N.; Serwadda, D.; Li, C.; Wabwire-Mangen, F.; Meehan, M. O.; Lutalo, T.; Gray, R. H. Viral Load and Heterosexual Transmission of Human Immunodeficiency Virus Type 1. *N. Engl. J. Med.* **2000**, *342* (13), 921–929. https://doi.org/10.1056/NEJM200003303421303.

(44) Deeks, S. G.; Overbaugh, J.; Phillips, A.; Buchbinder, S. HIV Infection. *Nat. Rev. Dis. Prim.* **2015**, *1* (1), 15035. https://doi.org/10.1038/nrdp.2015.35.

(45) Deeks, S. G.; Overbaugh, J.; Phillips, A.; Buchbinder, S. HIV Infection. *Nat. Rev. Dis. Prim.* **2015**, *1* (1), 15035. https://doi.org/10.1038/nrdp.2015.35.

(46) Laskey, S. B.; Siliciano, R. F. A Mechanistic Theory to Explain the Efficacy of Antiretroviral Therapy. *Nat. Rev. Microbiol.* **2014**, *12* (11), 772–780. https://doi.org/10.1038/nrmicro3351.

(47) Freed, E. O. HIV-1 Replication. *Somat. Cell Mol. Genet.* **2001**, *26* (1/6), 13–33. https://doi.org/10.1023/A:1021070512287.

(48) Freed, E. O. HIV-1 Assembly, Release and Maturation. *Nat. Rev. Microbiol.* **2015**, *13* (8), 484–496. https://doi.org/10.1038/nrmicro3490.

(49) Perelson, A. S.; Neumann, A. U.; Markowitz, M.; Leonard, J. M.; Ho, D. D. HIV-1 Dynamics in Vivo: Virion Clearance Rate, Infected Cell Life-Span, and Viral Generation Time. *Science* **1996**, *271* (5255), 1582–1586. https://doi.org/10.1126/science.271.5255.1582.

(50) Perelson, A. S.; Neumann, A. U.; Markowitz, M.; Leonard, J. M.; Ho, D. D. HIV-1

Dynamics in Vivo: Virion Clearance Rate, Infected Cell Life-Span, and Viral Generation Time. *Science* **1996**, *271* (5255), 1582–1586. https://doi.org/10.1126/science.271.5255.1582.

(51) Beyrer, C.; Pozniak, A. HIV Drug Resistance — An Emerging Threat to Epidemic Control. *N. Engl. J. Med.* **2017**, *377* (17), 1605–1607. https://doi.org/10.1056/NEJMp1710608.

(52) Martinez-Picado, J.; Deeks, S. G. Persistent HIV-1 Replication during Antiretroviral Therapy. *Curr. Opin. HIV AIDS* **2016**, *11* (4), 417. https://doi.org/10.1097/COH.0000000000000287.

(53) Panel on Antiretroviral Guidelines for Adults and Adolescents. *Guidelines for the Use of Antiretroviral Agents in Adults and Adolescents with HIV*; 2019.

(54) Estill, J.; Ford, N.; Salazar-Vizcaya, L.; Haas, A. D.; Blaser, N.; Habiyambere, V.; Keiser, O. Estimating the Need of Second-Line Antiretroviral Therapy in Adults in Sub-Saharan Africa up to 2030: A Mathematical Model. *lancet. HIV* **2016**, *3* (3), e132. https://doi.org/10.1016/S2352-3018(16)00016-3.

(55) Davey, R. T.; Bhat, N.; Yoder, C.; Chun, T.-W.; Metcalf, J. A.; Dewar, R.; Natarajan, V.; Lempicki, R. A.; Adelsberger, J. W.; Miller, K. D.; et al. HIV-1 and T Cell Dynamics after Interruption of Highly Active Antiretroviral Therapy (HAART) in Patients with a History of Sustained Viral Suppression. *Proc. Natl. Acad. Sci.* **1999**, *96* (26), 15109–15114. https://doi.org/10.1073/pnas.96.26.15109.

(56) UNAIDS. Global HIV & AIDS statistics — 2019 fact sheet | UNAIDS.

(57) Martin, A. R.; Siliciano, R. F. Progress Toward HIV Eradication: Case Reports, Current Efforts, and the Challenges Associated with Cure. *Annu. Rev. Med.* **2015**.

https://doi.org/10.1146/annurev-med-011514-023043.

(58) Brew, B. J.; Robertson, K.; Wright, E. J.; Churchill, M.; Crowe, S. M.; Cysique, L. A.; Deeks, S.; Garcia, J. V.; Gelman, B.; Gray, L. R.; et al. HIV Eradication Symposium: Will the Brain Be Left Behind? *J. Neurovirol.* **2015**. https://doi.org/10.1007/s13365-015-0322-6.

(59) Melhuish, A.; Lewthwaite, P. Natural History of HIV and AIDS. *Medicine (United Kingdom)*. 2018. https://doi.org/10.1016/j.mpmed.2018.03.010.

(60) Nakagawa, F.; Lodwick, R. K.; Smith, C. J.; Smith, R.; Cambiano, V.; Lundgren, J. D.; Delpech, V.; Phillips, A. N. Projected Life Expectancy of People with HIV According to Timing of Diagnosis. *AIDS* **2012**, *26* (3), 335–343. https://doi.org/10.1097/QAD.0b013e32834dcec9.

(61) Samji, H.; Cescon, A.; Hogg, R. S.; Modur, S. P.; Althoff, K. N.; Buchacz, K.; Burchell, A. N.; Cohen, M.; Gebo, K. A.; Gill, M. J.; et al. Closing the Gap: Increases in Life Expectancy among Treated HIV-Positive Individuals in the United States and Canada. *PLoS One* **2013**, *8* (12), e81355. https://doi.org/10.1371/journal.pone.0081355.

(62) Abdul-Quader, A. S.; Baughman, A. L.; Hladik, W. Estimating the Size of Key Populations. *Curr. Opin. HIV AIDS* **2014**, *9* (2), 107–114. https://doi.org/10.1097/COH.0000000000000041.

(63) Brown, T.; Peerapatanapokin, W. Evolving HIV Epidemics. *Curr. Opin. HIV AIDS* **2019**, *14* (5), 337–353. https://doi.org/10.1097/COH.0000000000000571.

(64) UNAIDS. UNAIDS Terminology Guidelines - 2015 | UNAIDS https://www.unaids.org/en/resources/documents/2015/2015_terminology_guidelines (accessed Aug 2, 2019).

(65) Levi, J.; Raymond, A.; Pozniak, A.; Vernazza, P.; Kohler, P.; Hill, A. Can the UNAIDS 90-90-90 Target Be Achieved? A Systematic Analysis of National HIV Treatment Cascades. *BMJ Glob. Heal.* **2016**, *1* (2), e000010. https://doi.org/10.1136/bmjgh-2015-000010.

(66) Dave, S.; Peter, T.; Fogarty, C.; Karatzas, N.; Belinsky, N.; Pant Pai, N. Which Community-Based HIV Initiatives Are Effective in Achieving UNAIDS 90-90-90 Targets? A Systematic Review and Meta-Analysis of Evidence (2007-2018). *PLoS One* **2019**, *14* (7), e0219826. https://doi.org/10.1371/journal.pone.0219826.

(67) Karatzas, N.; Peter, T.; Dave, S.; Fogarty, C.; Belinsky, N.; Pant Pai, N. Are Policy Initiatives Aligned to Meet UNAIDS 90-90-90 Targets Impacting HIV Testing and Linkages to Care? Evidence from a Systematic Review. *PLoS One* **2019**, *14* (6), e0216936. https://doi.org/10.1371/journal.pone.0216936.

(68) Figueroa, C.; Johnson, C.; Verster, A.; Baggaley, R. Attitudes and Acceptability on HIV Self-Testing Among Key Populations: A Literature Review. *AIDS Behav.* **2015**, *19* (11), 1949–1965. https://doi.org/10.1007/s10461-015-1097-8.

(69) Ratner, L.; Haseltine, W.; Patarca, R.; Livak, K. J.; Starcich, B.; Josephs, S. F.; Doran, E. R.; Rafalski, J. A.; Whitehorn, E. A.; Baumeister, K.; et al. Complete Nucleotide Sequence of the AIDS Virus, HTLV-III. *Nature* **1985**, *313* (6000), 277–284. https://doi.org/10.1038/313277a0.

(70) Meek, T. D. Inhibitors of HIV-1 Protease. *J. Enzyme Inhib.* **1992**, *6* (1), 65–98.

(71) Egger, M.; May, M.; Chêne, G.; Phillips, A. N.; Ledergerber, B.; Dabis, F.; Costagliola, D.; Monforte, A. D.; de Wolf, F.; Reiss, P.; et al. Prognosis of HIV-1-Infected Patients Starting Highly Active Antiretroviral Therapy: A Collaborative Analysis of Prospective

Studies. *Lancet* **2002**, *360* (9327), 119–129. https://doi.org/10.1016/S0140-6736(02)09411-4.

(72) Keefe, L. J.; Stoll, V. S. Accelerating Pharmaceutical Structure-Guided Drug Design: A Successful Model. *Drug Discov. Today* **2019**, *24* (2), 377–381. https://doi.org/10.1016/J.DRUDIS.2018.11.008.

(73) Johnson, R. J. Teaching Foundational Topics and Scientific Skills in Biochemistry within the Conceptual Framework of HIV Protease. *Biochem. Mol. Biol. Educ.* **2014**, *42* (4), 299–304. https://doi.org/10.1002/bmb.20793.

(74) Wlodawer, A.; Vondrasek, J. INHIBITORS OF HIV-1 PROTEASE: A Major Success of Structure-Assisted Drug Design. *Annu. Rev. Biophys. Biomol. Struct.* **1998**, *27* (1), 249–284. https://doi.org/10.1146/annurev.biophys.27.1.249.

(75) Ghosh, A. K.; Osswald, H. L.; Prato, G. Recent Progress in the Development of HIV-1 Protease Inhibitors for the Treatment of HIV/AIDS. *J. Med. Chem.* **2016**, *59* (11), 5172–5208. https://doi.org/10.1021/acs.jmedchem.5b01697.

(76) Kohl, N. E.; Emini, E. A.; Schleif, W. A.; Davis, L. J.; Heimbach, J. C.; Dixon, R. A.; Scolnick, E. M.; Sigal, I. S. Active Human Immunodeficiency Virus Protease Is Required for Viral Infectivity. *Proc. Natl. Acad. Sci. U. S. A.* **1988**, *85* (13), 4686–4690. https://doi.org/10.1073/pnas.85.13.4686.

(77) Seelmeier, S.; Schmidt, H.; Turk, V.; von der Helm, K. Human Immunodeficiency Virus Has an Aspartic-Type Protease That Can Be Inhibited by Pepstatin A. *Proc. Natl. Acad. Sci. U. S. A.* **1988**, *85* (18), 6612–6616. https://doi.org/10.1073/pnas.85.18.6612.

(78) McQuade, T.; Tomasselli, A.; Liu, L.; Karacostas, V.; Moss, B.; Sawyer, T.; Heinrikson, R.; Tarpley, W. A Synthetic HIV-1 Protease Inhibitor with Antiviral Activity Arrests

HIV-like Particle Maturation. *Science (80-.).* **1990**, *247* (4941), 454–456. https://doi.org/10.1126/science.2405486.

(79) Ganser, B. K.; Li, S.; Klishko, V. Y.; Finch, J. T.; Sundquist, W. I. Assembly and Analysis of Conical Models for the HIV-1 Core. *Science* **1999**, *283* (5398), 80–83. https://doi.org/10.1126/science.283.5398.80.

(80) Li, S.; Hill, C. P.; Sundquist, W. I.; Finch, J. T. Image Reconstructions of Helical Assemblies of the HIV-1 CA Protein. *Nature* **2000**, *407* (6802), 409–413. https://doi.org/10.1038/35030177.

(81) Solbak, S. M. Ø.; Reksten, T. R.; Hahn, F.; Wray, V.; Henklein, P.; Henklein, P.; Halskau, Ø.; Schubert, U.; Fossen, T. HIV-1 P6 — a Structured to Flexible Multifunctional Membrane-Interacting Protein. *Biochim. Biophys. Acta - Biomembr.* **2013**, *1828* (2), 816–823. https://doi.org/10.1016/j.bbamem.2012.11.010.

(82) Huang, L.; Chen, C. Autoprocessing of Human Immunodeficiency Virus Type 1 Protease Miniprecursor Fusions in Mammalian Cells. *AIDS Res. Ther.* **2010**. https://doi.org/10.1186/1742-6405-7-27.

(83) Pettit, S. C.; Moody, M. D.; Wehbie, R. S.; Kaplan, A. H.; Nantermet, P. V; Klein, C. A.; Swanstrom, R. The P2 Domain of Human Immunodeficiency Virus Type 1 Gag Regulates Sequential Proteolytic Processing and Is Required to Produce Fully Infectious Virions. *J. Virol.* **1994**, *68* (12), 8017–8027.

(84) Ivanchenko, S.; Godinez, W. J.; Lampe, M.; Kräusslich, H.-G.; Eils, R.; Rohr, K.; Bräuchle, C.; Müller, B.; Lamb, D. C. Dynamics of HIV-1 Assembly and Release. *PLoS Pathog.* **2009**, *5* (11), e1000652. https://doi.org/10.1371/journal.ppat.1000652.

(85) Pettit, S. C.; Lindquist, J. N.; Kaplan, A. H.; Swanstrom, R. Processing Sites in the Human

Immunodeficiency Virus Type 1 (HIV-1) Gag-Pro-Pol Precursor Are Cleaved by the Viral Protease at Different Rates. *Retrovirology* **2005**. https://doi.org/10.1186/1742-4690-2-66.

(86) Konvalinka, J.; Kräusslich, H.-G.; Müller, B. Retroviral Proteases and Their Roles in Virion Maturation. *Virology* **2015**, *479*, 403–417. https://doi.org/10.1016/j.virol.2015.03.021.

(87) Blundell, T. L.; Lapatto, R.; Wilderspin, A. F.; Hemmings, A. M.; Hobart, P. M.; Danley, D. E.; Whittle, P. J. The 3-D Structure of HIV-1 Proteinase and the Design of Antiviral Agents for the Treatment of AIDS. *Trends in Biochemical Sciences*. 1990. https://doi.org/10.1016/0968-0004(90)90280-O.

(88) Miller, M.; Jaskólski, M.; Rao, J. K. M.; Leis, J.; Wlodawer, A. Crystal Structure of a Retroviral Protease Proves Relationship to Aspartic Protease Family. *Nature* **1989**. https://doi.org/10.1038/337576a0.

(89) Navia, M. A.; Fitzgerald, P. M. D.; McKeever, B. M.; Leu, C.-T.; Heimbach, J. C.; Herber, W. K.; Sigal, I. S.; Darke, P. L.; Springer, J. P. Three-Dimensional Structure of Aspartyl Protease from Human Immunodeficiency Virus HIV-1. *Nature* **1989**, *337* (6208), 615–620. https://doi.org/10.1038/337615a0.

(90) Pearl, L. H.; Taylor, W. R. A Structural Model for the Retroviral Proteases. *Nature* **1987**. https://doi.org/10.1038/329351a0.

(91) Weber, I. T.; Miller, M.; Jaskólski, M.; Leis, J.; Skalka, A. M.; Wlodawer, A. Molecular Modeling of the HIV-1 Protease and Its Substrate Binding Site. *Science (80-.).* **1989**. https://doi.org/10.1126/science.2537531.

(92) Wlodawer, A.; Miller, M.; Jaskolski, M.; Sathyanarayana, B.; Baldwin, E.; Weber, I.; Selk, L.; Clawson, L.; Schneider, J.; Kent, S. Conserved Folding in Retroviral Proteases:

Crystal Structure of a Synthetic HIV-1 Protease. *Science (80-.).* **1989**, *245* (4918), 616–621. https://doi.org/10.1126/science.2548279.

(93) Lapatto, R.; Blundell, T.; Hemmings, a; Overington, J.; Wilderspin, a; Wood, S.; Merson, J. R.; Whittle, P. J.; Danley, D. E.; Geoghegan, K. F. X-Ray Analysis of HIV-1 Proteinase at 2.7 A Resolution Confirms Structural Homology among Retroviral Enzymes. *Nature* **1989**, *342* (6247), 299–302. https://doi.org/10.1038/342299a0.

(94) Miller, M. The Early Years of Retroviral Protease Crystal Structures. *Biopolymers* **2010**, *94* (4), 521–529. https://doi.org/10.1002/bip.21387.

(95) Miller, M.; Schneider, J.; Sathyanarayana, B. K.; Toth, M. V.; Marshall, G. R.; Clawson, L.; Selk, L.; Kent, S. B. H.; Wlodawer, A. Structure of Complex of Synthetic HIV-1 Protease with a Substrate-Based Inhibitor at 2.3 Å Resolution. *Science (80-.).* **1989**. https://doi.org/10.1126/science.2686029.

(96) Swain, A. L.; Miller, M. M.; Green, J.; Rich, D. H.; Schneider, J.; Kent, S. B.; Wlodawer, A. X-Ray Crystallographic Structure of a Complex between a Synthetic Protease of Human Immunodeficiency Virus 1 and a Substrate-Based Hydroxyethylamine Inhibitor. *Proc. Natl. Acad. Sci.* **1990**. https://doi.org/10.1073/pnas.87.22.8805.

(97) Jaskólski, M.; Wlodawer, A.; Tomasselli, A. G.; Sawyer, T. K.; Staples, D. G.; Heinrikson, R. L.; Schneider, J.; Kent, S. B. H.; Kent, S. B. H. Structure at 2.5-Å Resolution of Chemically Synthesized Human Immunodeficiency Virus Type 1 Protease Complexed with a Hydroxyethylene-Based Inhibitor. *Biochemistry* **1991**. https://doi.org/10.1021/bi00220a023.

(98) Sansom, C. E.; Wu, J.; Weber, I. T. Molecular Mechanics Analysis of Inhibitor Binding to HIV-1 Protease. *Protein Eng.* **1992**, *5* (7), 659–667.

(99) Gustchina, A.; Sansom, C.; Prevost, M.; Richelle, J.; Wodak, S. Y.; Wlodawer, A.; Weber, I. T. Energy Calculations and Analysis of HIV-1 Protease-Inhibitor Crystal Structures. *Protein Eng.* **1994**, *7* (3), 309–317.

(100) 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.; Kröhn, A. Rational Design of Peptide-Based HIV Proteinase Inhibitors. *Science* **1990**, *248* (4953), 358–361.

(101) Dunn, B. M.; Rao, M. Human Immunodeficiency Virus 1 Retropepsin. *Handb. Proteolytic Enzym.* **2004**, 144–153. https://doi.org/10.1016/B978-0-12-079611-3.50049-5.

(102) Krohn, A.; Redshaw, S.; Ritchie, J. C.; Graves, B. J.; Hatada, M. H. Novel Binding Mode of Highly Potent HIV-Proteinase Inhibitors Incorporating the (R)-Hydroxyethylamine Isostere. *J. Med. Chem.* **1991**, *34* (11), 3340–3342. https://doi.org/10.1021/jm00115a028.

(103) Rutenber, E. E.; McPhee, F.; Kaplan, A. P.; Gallion, S. L.; Hogan, J. C.; Craik, C. S.; Stroud, R. M. A New Class of HIV-1 Protease Inhibitor: The Crystallographic Structure, Inhibition and Chemical Synthesis of an Aminimide Peptide Isostere. *Bioorganic Med. Chem.* **1996**. https://doi.org/10.1016/0968-0896(96)00147-2.

(104) de Oliveira, T.; Engelbrecht, S.; Janse van Rensburg, E.; Gordon, M.; Bishop, K.; zur Megede, J.; Barnett, S. W.; Cassol, S. Variability at Human Immunodeficiency Virus Type 1 Subtype C Protease Cleavage Sites: An Indication of Viral Fitness? *J. Virol.* **2003**. https://doi.org/10.1128/jvi.77.17.9422-9430.2003.

(105) Torrecilla, E.; Llaćer Delicado, T.; Holguín, Á. New Findings in Cleavage Sites Variability across Groups, Subtypes and Recombinants of Human Immunodeficiency Virus Type 1. *PLoS One* **2014**. https://doi.org/10.1371/journal.pone.0088099.

(106) Louis, J. M.; Ishima, R.; Torchia, D. A.; Weber, I. T. HIV-1 Protease: Structure,

Dynamics, and Inhibition. *Adv. Pharmacol.* **2007**, *55*, 261–298. https://doi.org/10.1016/S1054-3589(07)55008-8.

(107) Weber, I. T. Comparison of the Crystal Structures and Intersubunit Interactions of Human Immunodeficiency and Rous Sarcoma Virus Proteases. *J. Biol. Chem.* **1990**, *265* (18), 10492–10496.

(108) Agniswamy, J.; Sayer, J. M.; Weber, I. T.; Louis, J. M. Terminal Interface Conformations Modulate Dimer Stability Prior to Amino Terminal Autoprocessing of HIV-1 Protease. *Biochemistry* **2012**, *51* (5), 1041–1050. https://doi.org/10.1021/bi201809s.

(109) Zutshi, R.; Franciskovich, J.; Shultz, M.; Schweitzer, B.; Bishop, P.; Wilson, M.; Chmielewski, J. Targeting the Dimerization Interface of HIV-1 Protease: Inhibition with Cross-Linked Interfacial Peptides. *J. Am. Chem. Soc.* **1997**. https://doi.org/10.1021/ja962496j.

(110) Boggetto, N.; Reboud-Ravaux, M. Dimerization Inhibitors of HIV-1 Protease. *Biol. Chem.* **2002**. https://doi.org/10.1515/BC.2002.150.

(111) Frutos, S.; Rodriguez-Mias, R. A.; Madurga, S.; Collinet, B.; Reboud-Ravaux, M.; Ludevid, D.; Giralt, E. Disruption of the HIV-1 Protease Dimer with Interface Peptides: Structural Studies Using NMR Spectroscopy Combined with [2-13C]-Trp Selective Labeling. In *Biopolymers - Peptide Science Section*; 2007. https://doi.org/10.1002/bip.20685.

(112) Ingr, M.; Uhlíková, T.; Střížovský, K.; Majerová, E.; Konvalinka, J. Kinetics of the Dimerization of Retroviral Proteases: The “Fireman’s Grip” and Dimerization. *Protein Sci.* **2003**, *12* (10), 2173–2182. https://doi.org/10.1110/ps.03171903.

(113) Ishima, R.; Ghirlando, R.; Tözsér, J.; Gronenborn, A. M.; Torchia, D. A.; Louis, J. M.

Folded Monomer of HIV-1 Protease. *J. Biol. Chem.* **2001**. https://doi.org/10.1074/jbc.M108136200.

(114) Louis, J. M.; Ishima, R.; Nesheiwat, I.; Pannell, L. K.; Lynch, S. M.; Torchia, D. A.; Gronenborn, A. M. Revisiting Monomeric HIV-1 Protease: Characterization and Redesign for Improved Properties. *J. Biol. Chem.* **2003**. https://doi.org/10.1074/jbc.M209726200.

(115) Todd, M. J.; Semo, N.; Freire, E. The Structural Stability of the HIV-1 Protease. *J. Mol. Biol.* **1998**. https://doi.org/10.1006/jmbi.1998.2090.

(116) Freedberg, D. I.; Ishima, R.; Jacob, J.; Wang, Y.-X.; Kustanovich, I.; Louis, J. M.; Torchia, D. A. Rapid Structural Fluctuations of the Free HIV Protease Flaps in Solution: Relationship to Crystal Structures and Comparison with Predictions of Dynamics Calculations. *Protein Sci.* **2009**. https://doi.org/10.1110/ps.33202.

(117) Ishima, R.; Torchia, D. A.; Louis, J. M. Mutational and Structural Studies Aimed at Characterizing the Monomer of HIV-1 Protease and Its Precursor. *J. Biol. Chem.* **2007**. https://doi.org/10.1074/jbc.M701304200.

(118) Agniswamy, J.; Louis, J. M.; Roche, J.; Harrison, R. W.; Weber, I. T. Structural Studies of a Rationally Selected Multi-Drug Resistant HIV-1 Protease Reveal Synergistic Effect of Distal Mutations on Flap Dynamics. *PLoS One* **2016**, *11* (12), e0168616. https://doi.org/10.1371/journal.pone.0168616.

(119) Wensing, A. M.; Calvez, V.; Günthard, H. F.; Ceccherini-Silberstein, F.; Charpentier, C.; Günthard, H. F.; Paredes, R.; Shafer, R. W.; Richman, D. D. 2019 Update of the Drug Resistance Mutations in HIV-1. *Top. Antivir. Med.* **2019**, *27* (3).

(120) Pearson, A. R.; Owen, R. L. Combining X-Ray Crystallography and Single-Crystal Spectroscopy to Probe Enzyme Mechanisms. *Biochem. Soc. Trans.* **2009**.

https://doi.org/10.1042/BST0370378.

(121) Veerapandian, B.; Cooper, J. B.; Šali, A.; Blundell, T. L.; Rosati, R. L.; Dominy, B. W.; Damon, D. B.; Hoover, D. J. Direct Observation by X-ray Analysis of the Tetrahedral "Intermediate" of Aspartic Proteinases. *Protein Sci.* **1992**. https://doi.org/10.1002/pro.5560010303.

(122) Coates, L.; Tuan, H. F.; Tomanicek, S.; Kovalevsky, A.; Mustyakimov, M.; Erskine, P.; Cooper, J. The Catalytic Mechanism of an Aspartic Proteinase Explored with Neutron and X-Ray Diffraction. *J. Am. Chem. Soc.* **2008**. https://doi.org/10.1021/ja801269x.

(123) Weber, I. T.; Waltman, M. J.; Mustyakimov, M.; Blakeley, M. P.; Keen, D. A.; Ghosh, A. K.; Langan, P.; Kovalevsky, A. Y. Joint X-Ray/Neutron Crystallographic Study of HIV-1 Protease with Clinical Inhibitor Amprenavir: Insights for Drug Design. *J. Med. Chem.* **2013**, *56* (13), 5631–5635. https://doi.org/10.1021/jm400684f.

(124) Gerlits, O.; Keen, D. A.; Blakeley, M. P.; Louis, J. M.; Weber, I. T.; Kovalevsky, A. Room Temperature Neutron Crystallography of Drug Resistant HIV-1 Protease Uncovers Limitations of X-Ray Structural Analysis at 100 K. *J. Med. Chem.* **2017**, *60* (5), 2018–2025. https://doi.org/10.1021/acs.jmedchem.6b01767.

(125) Brik, A.; Wong, C.-H. HIV-1 Protease: Mechanism and Drug Discovery. *Org. Biomol. Chem.* **2003**, *1* (1), 5–14. https://doi.org/10.1039/b208248a.

(126) Smith, R.; Brereton, I. M.; Chai, R. Y.; Kent, S. B. H. Ionization States of the Catalytic Residues in HIV-1 Protease. *Nat. Struct. Biol.* **1996**. https://doi.org/10.1038/nsb1196-946.

(127) Liu, H.; Müller-Plathe, F.; van Gunsteren, W. F. A Combined Quantum/Classical Molecular Dynamics Study of the Catalytic Mechanism of HIV Protease. *J. Mol. Biol.* **1996**, *261* (3), 454–469. https://doi.org/10.1006/JMBI.1996.0476.

(128) Hyland, L. J.; Tomaszek, T. A.; Meek, T. D.; Roberts, G. D.; Carr, S. A.; DesJarlais, R. L.; Magaard, V. W.; Bryan, H. L.; Fakhoury, S. A.; Moore, M. L.; et al. Human Immunodeficiency Virus-1 Protease. 1. Initial Velocity Studies and Kinetic Characterization of Reaction Intermediates by 18 O Isotope Exchange. *Biochemistry* **1991**. https://doi.org/10.1021/bi00098a023.

(129) Kumar, M.; Prashar, V.; Mahale, S.; Hosur, M. V. Observation of a Tetrahedral Reaction Intermediate in the HIV-1 Protease-Substrate Complex. *Biochem. J.* **2005**. https://doi.org/10.1042/BJ20041804.

(130) Das, A.; Mahale, S.; Prashar, V.; Bihani, S.; Ferrer, J. L.; Hosur, M. V. X-Ray Snapshot of HIV-1 Protease in Action: Observation of Tetrahedral Intermediate and Short Ionic Hydrogen Bond SIHB with Catalytic Aspartate. *J. Am. Chem. Soc.* **2010**. https://doi.org/10.1021/ja100002b.

(131) Kovalevsky, A. Y.; Chumanevich, A. A.; Liu, F.; Louis, J. M.; Weber, I. T. Caught in the Act: The 1.5 A Resolution Crystal Structures of the HIV-1 Protease and the I54V Mutant Reveal a Tetrahedral Reaction Intermediate. *Biochemistry* **2007**, *46* (51), 14854–14864. https://doi.org/10.1021/bi700822g.

(132) Shen, C.; Tie, Y.; Yu, X.; Wang, Y.; Andrey, Y.; Harrison, R. W.; Weber, I. T. Capturing the Reaction Pathway in Near-Atomic Resolution Crystal Structures of HIV-1 Protease. *Biochemistry* **2013**, *51* (39), 7726–7732. https://doi.org/10.1021/bi3008092.Capturing.

(133) Gerlits, O.; Wymore, T.; Das, A.; Shen, C.-H.; Parks, J. M.; Smith, J. C.; Weiss, K. L.; Keen, D. A.; Blakeley, M. P.; Louis, J. M.; et al. Long-Range Electrostatics-Induced Two-Proton Transfer Captured by Neutron Crystallography in an Enzyme Catalytic Site. *Angew. Chemie Int. Ed.* **2016**, *55* (16), 4924–4927.

https://doi.org/10.1002/anie.201509989.

(134) De Clercq, E. The History of Antiretrovirals: Key Discoveries over the Past 25 Years. *Rev. Med. Virol.* **2009**, *19* (5), 287–299. https://doi.org/10.1002/rmv.624.

(135) Wlodawer, A. Rational Approach to AIDS Drug Design Through Structural Biology. *Annu. Rev. Med.* **2002**, *53* (1), 595–614. https://doi.org/10.1146/annurev.med.53.052901.131947.

(136) Tarpley;, B. L. H. B. E. D. W. F. Tipranavir Inhibits Broadly Protease Inhibitor-Resistant HIV-1 Clinical Samples. *Aids* **2000**, *14* (13), 1943–1948.

(137) Louis, J. M.; Aniana, A.; Weber, I. T.; Sayer, J. M. Inhibition of Autoprocessing of Natural Variants and Multidrug Resistant Mutant Precursors of HIV-1 Protease by Clinical Inhibitors. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108* (22), 9072–9077. https://doi.org/10.1073/pnas.1102278108.

(138) Dierynck, I.; De Wit, M.; Gustin, E.; Keuleers, I.; Vandersmissen, J.; Hallenberger, S.; Hertogs, K. Binding Kinetics of Darunavir to Human Immunodeficiency Virus Type 1 Protease Explain the Potent Antiviral Activity and High Genetic Barrier. *J. Virol.* **2007**, *81* (24), 13845–13851. https://doi.org/10.1128/JVI.01184-07.

(139) Wang, Y.; Lv, Z.; Chu, Y. HIV Protease Inhibitors: A Review of Molecular Selectivity and Toxicity. *HIV/AIDS - Res. Palliat. Care* **2015**, *7*, 95. https://doi.org/10.2147/HIV.S79956.

(140) Agniswamy, J.; Louis, J. M.; Shen, C.-H.; Yashchuk, S.; Ghosh, A. K.; Weber, I. T. Substituted Bis-THF Protease Inhibitors with Improved Potency against Highly Resistant Mature HIV-1 Protease PR20. *J. Med. Chem.* **2015**, *58* (12), 5088–5095. https://doi.org/10.1021/acs.jmedchem.5b00474.

(141) Salcedo Gómez, P. M.; Amano, M.; Yashchuk, S.; Mizuno, A.; Das, D.; Ghosh, A. K.; Mitsuya, H. GRL-04810 and GRL-05010, Difluoride-Containing Nonpeptidic HIV-1 Protease Inhibitors (PIs) That Inhibit the Replication of Multi-PI-Resistant HIV-1 in Vitro and Possess Favorable Lipophilicity That May Allow Blood-Brain Barrier Penetration. *Antimicrob. Agents Chemother.* **2013**, *57* (12), 6110–6121. https://doi.org/10.1128/AAC.01420-13.

(142) Hedskog, C.; Mild, M.; Jernberg, J.; Sherwood, E.; Bratt, G.; Leitner, T.; Lundeberg, J.; Andersson, B.; Albert, J. Dynamics of HIV-1 Quasispecies during Antiviral Treatment Dissected Using Ultra-Deep Pyrosequencing. *PLoS One* **2010**, *5* (7), e11345. https://doi.org/10.1371/journal.pone.0011345.

(143) Lloyd, S. B.; Kent, S. J.; Winnall, W. R. The High Cost of Fidelity. *AIDS Res. Hum. Retroviruses* **2014**, *30* (1), 8–16. https://doi.org/10.1089/AID.2013.0153.

(144) Garforth, S. J.; Kim, T. W.; Parniak, M. A.; Kool, E. T.; Prasad, V. R. Site-Directed Mutagenesis in the Fingers Subdomain of HIV-1 Reverse Transcriptase Reveals a Specific Role for the B3–B4 Hairpin Loop in DNTP Selection. *J. Mol. Biol.* **2007**, *365* (1), 38–49. https://doi.org/10.1016/J.JMB.2006.09.057.

(145) Temin, H. M. Retrovirus Variation and Reverse Transcription: Abnormal Strand Transfers Result in Retrovirus Genetic Variation. *Proc. Natl. Acad. Sci. U. S. A.* **1993**, *90* (15), 6900–6903. https://doi.org/10.1073/pnas.90.15.6900.

(146) Pathak, V. K.; Hu, W.-S. "Might as Well Jump!" Template Switching by Retroviral Reverse Transcriptase, Defective Genome Formation, and Recombination. *Semin. Virol.* **1997**, *8* (2), 141–150. https://doi.org/10.1006/SMVY.1997.0114.

(147) Mansky, L. M.; Temin, H. M. Lower in Vivo Mutation Rate of Human Immunodeficiency

Virus Type 1 than That Predicted from the Fidelity of Purified Reverse Transcriptase. *J. Virol.* **1995**, *69* (8), 5087–5094.

(148) Nanni, R. G.; Ding, J.; Jacobo-Molina, A.; Hughes, S. H.; Arnold, E. Review of HIV-1 Reverse Transcriptase Three-Dimensional Structure: Implications for Drug Design. *Perspect. Drug Discov. Des.* **1993**, *1* (1), 129–150. https://doi.org/10.1007/BF02171659.

(149) Rose, R. B.; Craik, C. S.; Stroud, R. M. Domain Flexibility in Retroviral Proteases: Structural Implications for Drug Resistant Mutations † , ‡. *Biochemistry* **1998**, *37* (8), 2607–2621. https://doi.org/10.1021/bi9716074.

(150) Tramuto, F.; Bonura, F.; Mancuso, S.; Romano, N.; Vitale, F. Detection of a New 3-Base Pair Insertion Mutation in the Protease Gene of Human Immunodeficiency Virus Type 1 during Highly Active Antiretroviral Therapy (HAART). *AIDS Res. Hum. Retroviruses* **2005**, *21* (5), 420–423. https://doi.org/10.1089/aid.2005.21.420.

(151) Pereira-Vaz, J.; Duque, V.; Trindade, L.; Saraiva-da-Cunha, J.; Meliço-Silvestre, A. Detection of the Protease Codon 35 Amino Acid Insertion in Sequences from Treatment-Naïve HIV-1 Subtype C Infected Individuals in the Central Region of Portugal. *J. Clin. Virol.* **2009**, *46* (2), 169–172. https://doi.org/10.1016/j.jcv.2009.06.019.

(152) Kozísek, M.; Sasková, K. G.; Rezácová, P.; Brynda, J.; van Maarseveen, N. M.; De Jong, D.; Boucher, C. A.; Kagan, R. M.; Nijhuis, M.; Konvalinka, J. Ninety-Nine Is Not Enough: Molecular Characterization of Inhibitor-Resistant Human Immunodeficiency Virus Type 1 Protease Mutants with Insertions in the Flap Region. *J. Virol.* **2008**, *82* (12), 5869–5878. https://doi.org/10.1128/JVI.02325-07.

(153) Kozísek, M.; Henke, S.; Sasková, K. G.; Jacobs, G. B.; Schuch, A.; Buchholz, B.; Müller, V.; Kräusslich, H.-G. G.; Rezácová, P.; Konvalinka, J.; et al. Mutations in HIV-1 Gag and

Pol Compensate for the Loss of Viral Fitness Caused by a Highly Mutated Protease. *Antimicrob. Agents Chemother.* **2012**, *56* (8), 4320–4330. https://doi.org/10.1128/AAC.00465-12.

(154) Dam, E.; Quercia, R.; Glass, B.; Descamps, D.; Launay, O.; Duval, X.; Kräusslich, H.-G.; Hance, A. J.; Clavel, F.; Group, A. 109 S. Gag Mutations Strongly Contribute to HIV-1 Resistance to Protease Inhibitors in Highly Drug-Experienced Patients besides Compensating for Fitness Loss. *PLoS Pathog.* **2009**, *5* (3), e1000345. https://doi.org/10.1371/journal.ppat.1000345.

(155) Clavel, F.; Mammano, F. Role of Gag in HIV Resistance to Protease Inhibitors. *Viruses* **2010**, *2* (7), 1411–1426. https://doi.org/10.3390/v2071411.

(156) Bangsberg, D. R.; Acosta, E. P.; Gupta, R.; Guzman, D.; Riley, E. D.; Harrigan, P. R.; Parkin, N.; Deeks, S. G. Adherence–Resistance Relationships for Protease and Non-Nucleoside Reverse Transcriptase Inhibitors Explained by Virological Fitness. *AIDS* **2006**, *20* (2), 223–231. https://doi.org/10.1097/01.aids.0000199825.34241.49.

(157) Paterson, D. L.; Swindells, S.; Mohr, J.; Brester, M.; Vergis, E. N.; Squier, C.; Wagener, M. M.; Singh, N. Adherence to Protease Inhibitor Therapy and Outcomes in Patients with HIV Infection. *Ann. Intern. Med.* **2000**, *133* (1), 21–30.

(158) Nolan, D.; Reiss, P.; Mallal, S. Adverse Effects of Antiretroviral Therapy for HIV Infection: A Review of Selected Topics. *Expert Opin. Drug Saf.* **2005**, *4* (2), 201–218.

(159) Yasuda, J. M.; Miller, C.; Currier, J. S.; Forthal, D. N.; Kemper, C. A.; Beall, G. N.; Tilles, J. G.; Capparelli, E. V; McCutchan, J. A.; Haubrich, R. H. The Correlation between Plasma Concentrations of Protease Inhibitors, Medication Adherence and Virological Outcome in HIV-Infected Patients. *Antivir. Ther.* **2004**, *9* (5), 753–761.

(160) Harrigan, P. R.; Hogg, R. S.; Dong, W. W. Y.; Yip, B.; Wynhoven, B.; Woodward, J.; Brumme, C. J.; Brumme, Z. L.; Mo, T.; Alexander, C. S.; et al. Predictors of HIV Drug-Resistance Mutations in a Large Antiretroviral-Naive Cohort Initiating Triple Antiretroviral Therapy. *J. Infect. Dis.* **2005**, *191* (3), 339–347. https://doi.org/10.1086/427192.

(161) Pham, Q. D.; Wilson, D. P.; Law, M. G.; Kelleher, A. D.; Zhang, L. Global Burden of Transmitted HIV Drug Resistance and HIV-Exposure Categories. *AIDS* **2014**, *28* (18), 2751–2762. https://doi.org/10.1097/QAD.0000000000000494.

(162) Bonura, F.; Tramuto, F.; Vitale, F.; Perna, A. M.; Viviano, E.; Romano, N. Transmission of Drug-Resistant HIV Type 1 Strains in HAART-Naive Patients: A 5-Year Retrospective Study in Sicily, Italy. *AIDS Res. Hum. Retroviruses* **2010**, *26* (9), 961–965. https://doi.org/10.1089/aid.2009.0250.

(163) Weber, I. T.; Agniswamy, J. HIV-1 Protease: Structural Perspectives on Drug Resistance. *Viruses* **2009**, *1* (3), 1110–1136. https://doi.org/10.3390/v1031110.

(164) Wensing, A. M.; Calvez, V.; Günthard, H. F.; Johnson, V. A.; Paredes, R.; Pillay, D.; Shafer, R. W.; Richman, D. D. 2017 Update of the Drug Resistance Mutations in HIV-1. *Top. Antivir. Med.* **2017**, *24* (4), 132–133.

(165) Weber, I. T.; Kneller, D. W.; Wong-Sam, A. Highly Resistant HIV-1 Proteases and Strategies for Their Inhibition. *Future Med. Chem.* **2015**, *7* (8), 1023–1038. https://doi.org/10.4155/fmc.15.44.

(166) Shen, C.-H.; Wang, Y.-F.; Kovalevsky, A. Y.; Harrison, R. W.; Weber, I. T. Amprenavir Complexes with HIV-1 Protease and Its Drug-Resistant Mutants Altering Hydrophobic Clusters. *FEBS J.* **2010**, *277* (18), 3699–3714. https://doi.org/10.1111/j.1742-

4658.2010.07771.x.

(167) Wensing, A. M.; Calvez, V.; Günthard, H. F.; Ceccherini-Silberstein, F.; Charpentier, C.; Günthard, H. F.; Paredes, R.; Shafer, R. W.; Richman, D. D. 2019 Update of the Drug Resistance Mutations in HIV-1. *Top. Antivir. Med.* **2019**, *27* (3).

(168) Doyon, L.; Tremblay, S.; Bourgon, L.; Wardrop, E.; Cordingley, M. G. Selection and Characterization of HIV-1 Showing Reduced Susceptibility to the Non-Peptidic Protease Inhibitor Tipranavir. *Antiviral Res.* **2005**, *68* (1), 27–35. https://doi.org/10.1016/J.ANTIVIRAL.2005.07.003.

(169) Wang, Y.-F.; Tie, Y.; Boross, P. I.; Tozser, J.; Ghosh, A. K.; Harrison, R. W.; Weber, I. T. Potent New Antiviral Compound Shows Similar Inhibition and Structural Interactions with Drug Resistant Mutants and Wild Type HIV-1 Protease. *J. Med. Chem.* **2007**, *50* (18), 4509–4515. https://doi.org/10.1021/jm070482q.

(170) Tie, Y.; Boross, P. I.; Wang, Y. F.; Gaddis, L.; Liu, F.; Chen, X.; Tozser, J.; Harrison, R. W.; Weber, I. T. Molecular Basis for Substrate Recognition and Drug Resistance from 1.1 to 1.6 Å Resolution Crystal Structures of HIV-1 Protease Mutants with Substrate Analogs. *FEBS J.* **2005**, *272* (20), 5265–5277. https://doi.org/10.1111/j.1742-4658.2005.04923.x.

(171) Sasková, K. G.; Kozísek, M.; Lepsík, M.; Brynda, J.; Rezácová, P.; Václavíková, J.; Kagan, R. M.; Machala, L.; Konvalinka, J. Enzymatic and Structural Analysis of the I47A Mutation Contributing to the Reduced Susceptibility to HIV Protease Inhibitor Lopinavir. *Protein Sci.* **2008**, *17* (9), 1555–1564. https://doi.org/10.1110/ps.036079.108.

(172) Goldfarb, N. E.; Ohanessian, M.; Biswas, S.; McGee Jr, T. D.; Mahon, B. P.; Ostrov, D. A.; Garcia, J. P.; Tang, Y.; McKenna, R.; Roitberg, A. E.; et al. Defective Hydrophobic Sliding Mechanism and Active Site Expansion in HIV-1 Protease Drug Resistant Variant

Gly48Thr/Leu89Met: Mechanisms for the Loss of Saquinavir Binding Potency. *Biochemistry* **2014**. https://doi.org/10.1021/bi501088e.

(173) Hayashi, H.; Takamune, N.; Nirasawa, T.; Aoki, M.; Morishita, Y.; Das, D.; Koh, Y.; Ghosh, A. K.; Misumi, S.; Mitsuya, H. Dimerization of HIV-1 Protease Occurs through Two Steps Relating to the Mechanism of Protease Dimerization Inhibition by Darunavir. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111* (33), 12234–12239. https://doi.org/10.1073/pnas.1400027111.

(174) Aoki, M.; Danish, M. L.; Aoki-Ogata, H.; Amano, M.; Ide, K.; Das, D.; Koh, Y.; Mitsuya, H. Loss of the Protease Dimerization Inhibition Activity of Tipranavir (TPV) and Its Association with the Acquisition of Resistance to TPV by HIV-1. *J. Virol.* **2012**, *86* (24), 13384–13396. https://doi.org/10.1128/JVI.07234-11.

(175) Koh, Y.; Aoki, M.; Danish, M. L.; Aoki-Ogata, H.; Amano, M.; Das, D.; Shafer, R. W.; Ghosh, A. K.; Mitsuya, H. Loss of Protease Dimerization Inhibition Activity of Darunavir Is Associated with the Acquisition of Resistance to Darunavir by HIV-1. *J. Virol.* **2011**, *85* (19), 10079–10089. https://doi.org/10.1128/JVI.05121-11.

(176) Chang, Y.-C. E.; Yu, X.; Zhang, Y.; Tie, Y.; Wang, Y.-F.; Yashchuk, S.; Ghosh, A. K.; Harrison, R. W.; Weber, I. T. Potent Antiviral HIV-1 Protease Inhibitor GRL-02031 Adapts to the Structures of Drug Resistant Mutants with Its P1'-Pyrrolidinone Ring. *J. Med. Chem.* **2012**, *55* (7), 3387–3397. https://doi.org/10.1021/jm300072d.

(177) Louis, J. M.; Zhang, Y.; Sayer, J. M.; Wang, Y.; Robert, W.; Weber, I. T. Drug Resistance Mutation L76V Decreases the Dimer Stability and Rate of Autoprocessing of HIV-1 Protease by Reducing Internal Hydrophobic Contacts. **2012**, *50* (21), 4786–4795. https://doi.org/10.1021/bi200033z.Drug.

(178) Wong-Sam, A.; Wang, Y.-F.; Zhang, Y.; Ghosh, A. K.; Harrison, R. W.; Weber, I. T. Drug Resistance Mutation L76V Alters Nonpolar Interactions at the Flap–Core Interface of HIV-1 Protease. *ACS Omega* **2018**, *3* (9), 12132–12140. https://doi.org/10.1021/acsomega.8b01683.

(179) Ragland, D. A.; Nalivaika, E. A.; Nalam, M. N. L.; Prachanronarong, K. L.; Cao, H.; Bandaranayake, R. M.; Cai, Y.; Kurt-Yilmaz, N.; Schiffer, C. A. Drug Resistance Conferred by Mutations Outside the Active Site through Alterations in the Dynamic and Structural Ensemble of HIV-1 Protease. *J. Am. Chem. Soc.* **2014**, *136* (34), 11956–11963. https://doi.org/10.1021/ja504096m.

(180) Mahalingam, B.; Wang, Y.-F.; Boross, P. I.; Tozser, J.; Louis, J. M.; Harrison, R. W.; Weber, I. T. Crystal Structures of HIV Protease V82A and L90M Mutants Reveal Changes in the Indinavir-Binding Site. *Eur. J. Biochem.* **2004**, *271* (8), 1516–1524. https://doi.org/10.1111/j.1432-1033.2004.04060.x.

(181) Hirotaka Ode, *; Saburo Neya; Masayuki Hata; Wataru Sugiura, † and; Hoshino‡, T. Computational Simulations of HIV-1 ProteasesMulti-Drug Resistance Due to Nonactive Site Mutation L90M. **2006**. https://doi.org/10.1021/JA060682B.

(182) Henes, M.; Kosovrasti, K.; Lockbaum, G. J.; Leidner, F.; Nachum, G. S.; Nalivaika, E. A.; Bolon, D. N. A.; Kurt Yilmaz, N.; Schiffer, C. A.; Whitfield, T. W. Molecular Determinants of Epistasis in HIV-1 Protease: Elucidating the Interdependence of L89V and L90M Mutations in Resistance. *Biochemistry* **2019**, *58* (35), 3711–3726. https://doi.org/10.1021/acs.biochem.9b00446.

(183) Liu, F.; Boross, P. I.; Wang, Y. F.; Tozser, J.; Louis, J. M.; Harrison, R. W.; Weber, I. T. Kinetic, Stability, and Structural Changes in High-Resolution Crystal Structures of HIV-1

Protease with Drug-Resistant Mutations L24I, I50V, and G73S. *J. Mol. Biol.* **2005**, *354* (4), 789–800. https://doi.org/10.1016/j.jmb.2005.09.095.

(184) Wang, Y.; Liu, Z.; Brunzelle, J. S.; Kovari, I. A.; Dewdney, T. G.; Reiter, S. J.; Kovari, L. C. The Higher Barrier of Darunavir and Tipranavir Resistance for HIV-1 Protease. *Biochem. Biophys. Res. Commun.* **2011**, *412* (4), 737–742. https://doi.org/10.1016/j.bbrc.2011.08.045.

(185) Yedidi, R. S.; Proteasa, G.; Martin, P. D.; Liu, Z.; Vickrey, J. F.; Kovari, I. A.; Kovari, L. C. A Multi-Drug Resistant HIV-1 Protease Is Resistant to the Dimerization Inhibitory Activity of TLF-PafF. *J. Mol. Graph. Model.* **2014**, *53*, 105–111. https://doi.org/10.1016/j.jmgm.2014.06.010.

(186) Martin, P.; Vickrey, J. F.; Proteasa, G.; Jimenez, Y. L.; Wawrzak, Z.; Winters, M. A.; Merigan, T. C.; Kovari, L. C. "Wide-Open" 1.3 A Structure of a Multidrug-Resistant HIV-1 Protease as a Drug Target. *Structure* **2005**, *13* (12), 1887–1895. https://doi.org/10.1016/j.str.2005.11.005.

(187) Yedidi, R. S.; Proteasa, G.; Martinez, J. L.; Vickrey, J. F.; Martin, P. D.; Wawrzak, Z.; Liu, Z.; Kovari, I. A.; Kovari, L. C. Contribution of the 80s Loop of HIV-1 Protease to the Multidrug-Resistance Mechanism: Crystallographic Study of MDR769 HIV-1 Protease Variants. *Acta Crystallogr. D. Biol. Crystallogr.* **2011**, *67* (Pt 6), 524–532. https://doi.org/10.1107/S0907444911011541.

(188) Logsdon, B. C.; Vickrey, J. F.; Martin, P.; Proteasa, G.; Koepke, J. I.; Terlecky, S. R.; Wawrzak, Z.; Winters, M. A.; Merigan, T. C.; Kovari, L. C. Crystal Structures of a Multidrug-Resistant Human Immunodeficiency Virus Type 1 Protease Reveal an Expanded Active-Site Cavity. *J. Virol.* **2004**, *78* (6), 3123–3132.

(189) Huang, X.; de Vera, I. M. S.; Veloro, A. M.; Rocca, J. R.; Simmerling, C.; Dunn, B. M.; Fanucci, G. E. Backbone 1H, 13C, and 15N Chemical Shift Assignment for HIV-1 Protease Subtypes and Multi-Drug Resistant Variant MDR 769. *Biomol. NMR Assign.* **2013**, *7* (2), 199–202. https://doi.org/10.1007/s12104-012-9409-7.

(190) Fong, C. W. The Effect of Desolvation on the Binding of Inhibitors to HIV-1 Protease and Cyclin-Dependent Kinases: Causes of Resistance. *Bioorg. Med. Chem. Lett.* **2016**, *26* (15), 3705–3713. https://doi.org/10.1016/J.BMCL.2016.05.080.

(191) Agniswamy, J.; Shen, C.-H.; Wang, Y.-F.; Ghosh, A. K.; Rao, K. V.; Xu, C.-X.; Sayer, J. M.; Louis, J. M.; Weber, I. T. Extreme Multidrug Resistant HIV-1 Protease with 20 Mutations Is Resistant to Novel Protease Inhibitors with P1'-Pyrrolidinone or P2-Tris-Tetrahydrofuran. *J. Med. Chem.* **2013**, *56* (10), 4017–4027. https://doi.org/10.1021/jm400231v.

(192) Agniswamy, J.; Shen, C. H.; Aniana, A.; Sayer, J. M.; Louis, J. M.; Weber, I. T. HIV-1 Protease with 20 Mutations Exhibits Extreme Resistance to Clinical Inhibitors through Coordinated Structural Rearrangements. *Biochemistry* **2012**, *51* (13), 2819–2828. https://doi.org/10.1021/bi2018317.

(193) Louis, J. M.; Tözsér, J.; Roche, J.; Matúz, K.; Aniana, A.; Sayer, J. M. Enhanced Stability of Monomer Fold Correlates with Extreme Drug Resistance of HIV-1 Protease. *Biochemistry* **2013**, *52* (43), 7678–7688. https://doi.org/10.1021/bi400962r.

(194) Shen, C.-H.; Chang, Y.-C.; Agniswamy, J.; Harrison, R. W.; Weber, I. T. Conformational Variation of an Extreme Drug Resistant Mutant of HIV Protease. *J. Mol. Graph. Model.* **2015**, *62*, 87–96. https://doi.org/10.1016/j.jmgm.2015.09.006.

(195) Chetty, S.; Bhakat, S.; Martin, A. J. M.; Soliman, M. E. S. Multi-Drug Resistance Profile

of PR20 HIV-1 Protease Is Attributed to Distorted Conformational and Drug Binding Landscape: Molecular Dynamics Insights. *J. Biomol. Struct. Dyn.* **2016**, *34* (1). https://doi.org/10.1080/07391102.2015.1018326.

(196) Kneller, D. W.; Agniswamy, J.; Ghosh, A. K.; Weber, I. T. Potent Antiviral HIV-1 Protease Inhibitor Combats Highly Drug Resistant Mutant PR20. *Biochem. Biophys. Res. Commun.* **2019**, *519* (1), 61–66. https://doi.org/10.1016/J.BBRC.2019.08.126.

(197) Sasková, K. G.; Kozísek, M.; Rezácová, P.; Brynda, J.; Yashina, T.; Kagan, R. M.; Konvalinka, J. Molecular Characterization of Clinical Isolates of Human Immunodeficiency Virus Resistant to the Protease Inhibitor Darunavir. *J. Virol.* **2009**, *83* (17), 8810–8818. https://doi.org/10.1128/JVI.00451-09.

(198) Kožíšek, M.; Lepšík, M.; Grantz Šašková, K.; Brynda, J.; Konvalinka, J.; Rezáčová, P. Thermodynamic and Structural Analysis of HIV Protease Resistance to Darunavir - Analysis of Heavily Mutated Patient-Derived HIV-1 Proteases. *FEBS J.* **2014**, *281* (7), 1834–1847. https://doi.org/10.1111/febs.12743.

(199) Kagan, R. M.; Cheung, P. K.; Huard, T. K.; Lewinski, M. A. Increasing Prevalence of HIV-1 Protease Inhibitor-Associated Mutations Correlates with Long-Term Non-Suppressive Protease Inhibitor Treatment. *Antiviral Res.* **2006**, *71* (1), 42–52. https://doi.org/10.1016/j.antiviral.2006.02.008.

(200) Koh, Y.; Amano, M.; Towata, T.; Danish, M.; Leshchenko-Yashchuk, S.; Das, D.; Nakayama, M.; Tojo, Y.; Ghosh, A. K.; Mitsuya, H. In Vitro Selection of Highly Darunavir-Resistant and Replication-Competent HIV-1 Variants by Using a Mixture of Clinical HIV-1 Isolates Resistant to Multiple Conventional Protease Inhibitors. *J. Virol.* **2010**, *72* (6), 4798–4810. https://doi.org/10.1128/jvi.00967-10.

(201) Zhang, Y.; Chang, Y. C. E.; Louis, J. M.; Wang, Y. F.; Harrison, R. W.; Weber, I. T. Structures of Darunavir-Resistant HIV-1 Protease Mutant Reveal Atypical Binding of Darunavir to Wide Open Flaps. *ACS Chem. Biol.* **2014**, *9* (6), 1351–1358. https://doi.org/10.1021/cb4008875.

(202) Yu, X.; Weber, I. T.; Harrison, R. W. Sparse Representation for Prediction of HIV-1 Protease Drug Resistance. *Proc. SIAM Int. Conf. Data Min.* **2013**, *2013*, 342–349. https://doi.org/10.1137/1.9781611972832.38.

(203) Yu, X.; Weber, I. T.; Harrison, R. W. Prediction of HIV Drug Resistance from Genotype with Encoded Three-Dimensional Protein Structure. *BMC Genomics* **2014**, *15* (Suppl 5), S1. https://doi.org/10.1186/1471-2164-15-S5-S1.

(204) Yu, X.; Weber, I. T.; Harrison, R. W. Identifying Representative Drug Resistant Mutants of HIV. *BMC Bioinformatics* **2015**, *16* (Suppl 17), S1. https://doi.org/10.1186/1471-2105-16-S17-S1.

(205) Rhee, S.-Y.; Taylor, J.; Fessel, W. J.; Kaufman, D.; Towner, W.; Troia, P.; Ruane, P.; Hellinger, J.; Shirvani, V.; Zolopa, A.; et al. HIV-1 Protease Mutations and Protease Inhibitor Cross-Resistance. *Antimicrob. Agents Chemother.* **2010**, *54* (10), 4253–4261. https://doi.org/10.1128/AAC.00574-10.

(206) Park, J. H.; Sayer, J. M.; Aniana, A.; Yu, X.; Weber, I. T.; Harrison, R. W.; Louis, J. M. Binding of Clinical Inhibitors to a Model Precursor of a Rationally Selected Multidrug Resistant HIV-1 Protease Is Significantly Weaker Than That to the Released Mature Enzyme. *Biochemistry* **2016**, *55* (16), 2390–2400. https://doi.org/10.1021/acs.biochem.6b00012.

(207) Agniswamy, J.; Kneller, D. W.; Brothers, R.; Wang, Y.-F.; Harrison, R. W.; Weber, I. T.

Highly Drug-Resistant HIV-1 Protease Mutant PRS17 Shows Enhanced Binding to Substrate Analogues. *ACS Omega* **2019**, *4* (5), 8707–8719. https://doi.org/10.1021/acsomega.9b00683.

(208) Kepler, J. A New Year's Gift of Hexagonal Snow, or On the Six-Cornered Snowflake. **1611**.

(209) Röntgen, W. On a New Kind of Rays. *Science (80-.).* **1896**, *3* (59), 227–231.

(210) Bragg, W. H. The Reflection of X-Rays by Crystals. (II.). *Proc. R. Soc. A Math. Phys. Eng. Sci.* **1913**, *89* (610), 246–248. https://doi.org/10.1098/rspa.1913.0082.

(211) Bragg, W. L. The Structure of Some Crystals as Indicated by Their Diffraction of X-Rays. *Proc. R. Soc. A Math. Phys. Eng. Sci.* **1913**, *89* (610), 248–277. https://doi.org/10.1098/rspa.1913.0083.

(212) Bragg, W. H.; Bragg, W. L. The Structure of the Diamond. *Nature* **1913**, *91* (2283), 557–557. https://doi.org/10.1038/091557a0.

(213) Dickinson, R. G.; Raymond, A. L. The Crystal Structure of Hexamethlene-Tetramine. *J. Am. Chem. Soc.* **1923**, *45* (1), 22–29. https://doi.org/10.1021/ja01654a003.

(214) Kendrew, J. C.; Bodo, G.; Dintzis, H. M.; Parrish, R. G.; Wyckoff, H.; Phillips, D. C. A Three-Dimensional Model of the Myoglobin Molecule Obtained by X-Ray Analysis. *Nature* **1958**, *181* (4610), 662–666. https://doi.org/10.1038/181662a0.

(215) Brink, C.; Hodgkin, D. C.; Lindsey, J.; Pickworth, J.; Robertson, J. H.; White, J. G. Structure of Vitamin B12: X-Ray Crystallographic Evidence on the Structure of Vitamin B12. *Nature* **1954**, *174* (4443), 1169–1171. https://doi.org/10.1038/1741169a0.

(216) ADAMS, M. J.; BLUNDELL, T. L.; DODSON, E. J.; DODSON, G. G.; VIJAYAN, M.; BAKER, E. N.; HARDING, M. M.; HODGKIN, D. C.; RIMMER, B.; SHEAT, S.

Structure of Rhombohedral 2 Zinc Insulin Crystals. *Nature* **1969**, *224* (5218), 491–495. https://doi.org/10.1038/224491a0.

(217) Watson, J. D.; Crick, F. H. C. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature* **1953**, *171* (4356), 737–738. https://doi.org/10.1038/171737a0.

(218) Brown, D. G.; Shotton, E. J.; Anderson, A.; Verlinde, C.; Hol, W.; Brown, D.; Flocco, M.; Tari, L.; Maynes, J.; Zheng, J.; et al. Diamond: Shedding Light on Structure-Based Drug Discovery. *Philos. Trans. A. Math. Phys. Eng. Sci.* **2015**, *373* (2036), 787–797. https://doi.org/10.1098/rsta.2014.0468.

(219) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N. The Protein Data Bank (Www.Rcsb.Org). *Nucleic Acids Res.* **2000**. https://doi.org/10.1093/nar/28.1.235.

(220) Blakeley, M. P.; Hasnain, S. S.; Antonyuk, S. V. Sub-Atomic Resolution X-Ray Crystallography and Neutron Crystallography: Promise, Challenges and Potential. *IUCrJ* **2015**, *2* (4), 464–474. https://doi.org/10.1107/S2052252515011239.

(221) Betzel, C.; Martirosyan, A.; Ruyters, G. Protein Crystallization on the International Space Station ISS; Springer, Cham, 2017; pp 27–39. https://doi.org/10.1007/978-3-319-64054-9_3.

(222) Benvenuti, M.; Mangani, S. Crystallization of Soluble Proteins in Vapor Diffusion for X-Ray Crystallography. *Nat. Protoc.* **2007**, *2* (7), 1633–1651. https://doi.org/10.1038/nprot.2007.198.

(223) Chayen, N. E.; Saridakis, E. Protein Crystallization: From Purified Protein to Diffraction-Quality Crystal. *Nat. Methods* **2008**, *5* (2), 147–153. https://doi.org/10.1038/nmeth.f.203.

(224) Russo Krauss, I.; Merlino, A.; Vergara, A.; Sica, F. An Overview of Biological Macromolecule Crystallization. *Int. J. Mol. Sci.* **2013**, *14* (6), 11643–11691. https://doi.org/10.3390/ijms140611643.

(225) Cowtan, K. Phase Problem in X-Ray Crystallography, and Its Solution. In *Encyclopedia of Life Sciences*; John Wiley & Sons, Ltd: Chichester, 2003. https://doi.org/10.1038/npg.els.0002722.

(226) Scapin, G. Molecular Replacement Then and Now. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2013**, *69* (11), 2266–2275. https://doi.org/10.1107/S0907444913011426.

(227) The Structure of Haemoglobin - V. Imidazole-Methaemoglobin: A Further Check of the Signs. *Proc. R. Soc. London. Ser. A. Math. Phys. Sci.* **1954**, *225* (1162), 308–314. https://doi.org/10.1098/rspa.1954.0204.

(228) Hendrickson, W. A.; Ogata, C. M. [28] Phase Determination from Multiwavelength Anomalous Diffraction Measurements. *Methods Enzymol.* **1997**, *276*, 494–523. https://doi.org/10.1016/S0076-6879(97)76074-9.

(229) Jones, T. A.; Zou, J. Y.; Cowan, S. W.; Kjeldgaard, M.; IUCr. Improved Methods for Building Protein Models in Electron Density Maps and the Location of Errors in These Models. *Acta Crystallogr. Sect. A Found. Crystallogr.* **1991**, *47* (2), 110–119. https://doi.org/10.1107/S0108767390010224.

(230) Mooij, W. T. M.; Mitsiki, E.; Perrakis, A. ProteinCCD: Enabling the Design of Protein Truncation Constructs for Expression and Crystallization Experiments. *Nucleic Acids Res.* **2009**, *37* (Web Server), W402–W405. https://doi.org/10.1093/nar/gkp256.

(231) Höfer, N.; Aragão, D.; Caffrey, M. Crystallizing Transmembrane Peptides in Lipidic Mesophases. *Biophys. J.* **2010**, *99* (3), L23–L25.

https://doi.org/10.1016/J.BPJ.2010.05.011.

(232) Qutub, Y.; Reviakine, I.; Maxwell, C.; Navarro, J.; Landau, E. M.; Vekilov, P. G. Crystallization of Transmembrane Proteins in Cubo: Mechanisms of Crystal Growth and Defect Formation. *J. Mol. Biol.* **2004**, *343* (5), 1243–1254. https://doi.org/10.1016/J.JMB.2004.09.022.

(233) Niedzialkowska, E.; Gasiorowska, O.; Handing, K. B.; Majorek, K. A.; Porebski, P. J.; Shabalin, I. G.; Zasadzinska, E.; Cymborowski, M.; Minor, W. Protein Purification and Crystallization Artifacts: The Tale Usually Not Told. *Protein Sci.* **2016**, *25* (3), 720–733. https://doi.org/10.1002/pro.2861.

(234) Wlodawer, A.; Sjoelin, L. Structure of Ribonuclease A: Results of Joint Neutron and x-Ray Refinement at 2.0-.ANG. Resolution. *Biochemistry* **1983**, *22* (11), 2720–2728. https://doi.org/10.1021/bi00280a021.

(235) Afonine, P. V.; Mustyakimov, M.; Grosse-Kunstleve, R. W.; Moriarty, N. W.; Langan, P.; Adams, P. D.; IUCr. Joint X-Ray and Neutron Refinement with *Phenix.Refine*. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2010**, *66* (11), 1153–1163. https://doi.org/10.1107/S0907444910026582.

(236) Blum, M. M.; Tomanicek, S. J.; John, H.; Hanson, B. L.; Rüterjans, H.; Schoenborn, B. P.; Langan, P.; Chen, J. C. H. X-Ray Structure of Perdeuterated Diisopropyl Fluorophosphatase (DFPase): Perdeuteration of Proteins for Neutron Diffraction. *Acta Crystallogr. Sect. F. Struct. Biol. Cryst. Commun.* **2010**, *66* (Pt 4), 379–385. https://doi.org/10.1107/S1744309110004318.

(237) Coates, L.; Stoica, A. D.; Hoffmann, C.; Richards, J.; Cooper, R.; IUCr. The Macromolecular Neutron Diffractometer (MaNDi) at the Spallation Neutron Source, Oak

Ridge: Enhanced Optics Design, High-Resolution Neutron Detectors and Simulated Diffraction. *J. Appl. Crystallogr.* **2010**, *43* (3), 570–577. https://doi.org/10.1107/S0021889810008587.

(238) Kikhney, A. G.; Svergun, D. I. A Practical Guide to Small Angle X-Ray Scattering (SAXS) of Flexible and Intrinsically Disordered Proteins. *FEBS Lett.* **2015**, *589* (19), 2570–2577. https://doi.org/10.1016/j.febslet.2015.08.027.

(239) Vestergaard, B. Analysis of Biostructural Changes, Dynamics, and Interactions – Small-Angle X-Ray Scattering to the Rescue. *Arch. Biochem. Biophys.* **2016**, *602*, 69–79. https://doi.org/10.1016/j.abb.2016.02.029.

(240) Mertens, H. D. T.; Svergun, D. I. Structural Characterization of Proteins and Complexes Using Small-Angle X-Ray Solution Scattering. *J. Struct. Biol.* **2010**, *172* (1), 128–141. https://doi.org/10.1016/j.jsb.2010.06.012.

(241) Hammel, M. Validation of Macromolecular Flexibility in Solution by Small-Angle X-Ray Scattering (SAXS). *Eur. Biophys. J.* **2012**, *41* (10), 789–799. https://doi.org/10.1007/s00249-012-0820-x.

(242) Grishaev, A. Sample Preparation, Data Collection, and Preliminary Data Analysis in Biomolecular Solution X-Ray Scattering. *Curr. Protoc. protein Sci.* **2012**, *Chapter 17*, Unit17.14. https://doi.org/10.1002/0471140864.ps1714s70.

(243) Wüthrich, K. Protein Structure Determination in Solution by NMR Spectroscopy. *Journal of Biological Chemistry*. 1990.

(244) Howard, M. J. Protein NMR Spectroscopy. *Curr. Biol.* **1998**, *8* (10), R331-3. https://doi.org/10.1016/s0960-9822(98)70214-3.

(245) Schur, F. K. M.; Hagen, W. J. H.; Rumlová, M.; Ruml, T.; Müller, B.; Kräusslich, H.-G.;

Briggs, J. A. G. Structure of the Immature HIV-1 Capsid in Intact Virus Particles at 8.8 Å Resolution. *Nature* **2015**, *517* (7535), 505–508. https://doi.org/10.1038/nature13838.

(246) Sirohi, D.; Chen, Z.; Sun, L.; Klose, T.; Pierson, T. C.; Rossmann, M. G.; Kuhn, R. J. The 3.8 Å Resolution Cryo-EM Structure of Zika Virus. *Science* **2016**, *352* (6284), 467–470. https://doi.org/10.1126/science.aaf5316.

(247) Zhang, X.; Ge, P.; Yu, X.; Brannan, J. M.; Bi, G.; Zhang, Q.; Schein, S.; Zhou, Z. H. Cryo-EM Structure of the Mature Dengue Virus at 3.5-Å Resolution. *Nat. Struct. Mol. Biol.* **2013**, *20* (1), 105–110. https://doi.org/10.1038/nsmb.2463.

(248) Bai, X.; McMullan, G.; Scheres, S. H. . How Cryo-EM Is Revolutionizing Structural Biology. *Trends Biochem. Sci.* **2015**, *40* (1), 49–57. https://doi.org/10.1016/J.TIBS.2014.10.005.

(249) Karplus, M.; McCammon, J. A. Molecular Dynamics Simulations of Biomolecules. *Nat. Struct. Biol.* **2002**, *9* (9), 646–652. https://doi.org/10.1038/nsb0902-646.

(250) Durrant, J. D.; McCammon, J. A. Molecular Dynamics Simulations and Drug Discovery. *BMC Biol.* **2011**, *9* (1), 71. https://doi.org/10.1186/1741-7007-9-71.

(251) Klepeis, J. L.; Lindorff-Larsen, K.; Dror, R. O.; Shaw, D. E. Long-Timescale Molecular Dynamics Simulations of Protein Structure and Function. *Curr. Opin. Struct. Biol.* **2009**, *19* (2), 120–127. https://doi.org/10.1016/J.SBI.2009.03.004.

(252) Congreve, M.; Murray, C. W.; Blundell, T. L. Keynote Review: Structural Biology and Drug Discovery. *Drug Discov. Today* **2005**, *10* (13), 895–907. https://doi.org/10.1016/S1359-6446(05)03484-7.

(253) Chen, Z.; Wilmanns, M.; Zeng, A.-P. Structural Synthetic Biotechnology: From Molecular Structure to Predictable Design for Industrial Strain Development. *Trends*

Biotechnol. **2010**, *28* (10), 534–542. https://doi.org/10.1016/J.TIBTECH.2010.07.004.

(254) Tanokura, M.; Miyakawa, T.; Guan, L.; Hou, F. Structural Analysis of Enzymes Used for Bioindustry and Bioremediation. *Biosci. Biotechnol. Biochem.* **2015**, *79* (9), 1391–1401. https://doi.org/10.1080/09168451.2015.1052770.

(255) Joerger, A. C.; Fersht, A. R. Structural Biology of the Tumor Suppressor P53. *Annu. Rev. Biochem.* **2008**, *77* (1), 557–582. https://doi.org/10.1146/annurev.biochem.77.060806.091238.

(256) Phillips, A. N.; Stover, J.; Cambiano, V.; Nakagawa, F.; Jordan, M. R.; Pillay, D.; Doherty, M.; Revill, P.; Bertagnolio, S. Impact of HIV Drug Resistance on HIV/AIDS-Associated Mortality, New Infections, and Antiretroviral Therapy Program Costs in Sub–Saharan Africa. *J. Infect. Dis.* **2017**, *215* (9), 1362–1365. https://doi.org/10.1093/infdis/jix089.

(257) Koh, Y.; Nakata, H.; Maeda, K.; Ogata, H.; Bilcer, G.; Devasamudram, T.; Kincaid, J. F.; Boross, P.; Wang, Y.-F.; Tie, Y.; et al. Novel Bis-Tetrahydrofuranylurethane-Containing Nonpeptidic Protease Inhibitor (PI) UIC-94017 (TMC114) with Potent Activity against Multi-PI-Resistant Human Immunodeficiency Virus In Vitro. *Antimicrob. Agents Chemother.* **2003**, *47* (10), 3123–3129. https://doi.org/10.1128/AAC.47.10.3123-3129.2003.

(258) Panel on Antiretroviral Guidelines for Adults and Adolescents. Guidelines for the Use of Antiretroviral Agents in Adults and Adolescents Living with HIV. **2016**.

(259) Lathouwers, E.; Wong, E. Y.; Luo, D.; Seyedkazemi, S.; De Meyer, S.; Brown, K. HIV-1 Resistance Rarely Observed in Subjects Using Darunavir Once-Daily Regimens across Clinical Studies. *HIV Clin. Trials* **2017**, *18* (5–6), 196–204.

https://doi.org/10.1080/15284336.2017.1387690.

(260) Orkin, C.; DeJesus, E.; Khanlou, H.; Stoehr, A.; Supparatpinyo, K.; Lathouwers, E.; Lefebvre, E.; Opsomer, M.; Van de Casteele, T.; Tomaka, F. Final 192-Week Efficacy and Safety of Once-Daily Darunavir/Ritonavir Compared with Lopinavir/Ritonavir in HIV-1-Infected Treatment-Naïve Patients in the ARTEMIS Trial. *HIV Med.* **2013**, *14* (1), 49–59. https://doi.org/10.1111/j.1468-1293.2012.01060.x.

(261) Huang, L.; Chen, C. Autoprocessing of Human Immunodeficiency Virus Type 1 Protease Miniprecursor Fusions in Mammalian Cells. *AIDS Res. Ther.* **2010**, *7* (1), 27. https://doi.org/10.1186/1742-6405-7-27.

(262) Huang, D.; Caflisch, A. How Does Darunavir Prevent HIV-1 Protease Dimerization? *J. Chem. Theory Comput.* **2012**, *8* (5), 1786–1794. https://doi.org/10.1021/ct300032r.

(263) Godfrey, C.; Thigpen, M. C.; Crawford, K. W.; Jean-Phillippe, P.; Pillay, D.; Persaud, D.; Kuritzkes, D. R.; Wainberg, M.; Raizes, E.; Fitzgibbon, J. Global HIV Antiretroviral Drug Resistance. *J. Infect. Dis.* **2017**, *216* (suppl_9), S798–S800. https://doi.org/10.1093/infdis/jix137.

(264) Ghosh, A. K.; Martyr, C. D.; Kassekert, L. A.; Nyalapatla, P. R.; Steffey, M.; Agniswamy, J.; Wang, Y.-F. Y.-F.; Weber, I. T.; Amano, M.; Mitsuya, H. Design, Synthesis, Biological Evaluation and X-Ray Structural Studies of HIV-1 Protease Inhibitors Containing Substituted Fused-Tetrahydropyranyl Tetrahydrofuran as P2-Ligands. *Org. Biomol. Chem.* **2015**, *13* (48), 11607–11621. https://doi.org/10.1039/c5ob01930c.

(265) Calcagno, A.; Di Perri, G.; Bonora, S. Treating HIV Infection in the Central Nervous System. *Drugs* **2017**, *77* (2), 145–157. https://doi.org/10.1007/s40265-016-0678-9.

(266) Ghosh, A. K.; Rao, K. V.; Nyalapatla, P. R.; Kovela, S.; Brindisi, M.; Osswald, H. L.;

Sekhara Reddy, B.; Agniswamy, J.; Wang, Y.-F.; Aoki, M.; et al. Design of Highly Potent, Dual-Acting and Central-Nervous-System-Penetrating HIV-1 Protease Inhibitors with Excellent Potency against Multidrug-Resistant HIV-1 Variants. *ChemMedChem* **2018**, *13* (8), 803–815. https://doi.org/10.1002/cmdc.201700824.

(267) Aoki, M.; Hayashi, H.; Rao, K. V.; Das, D.; Higashi-Kuwata, N.; Bulut, H.; Aoki-Ogata, H.; Takamatsu, Y.; Yedidi, R. S.; Davis, D. A.; et al. A Novel Central Nervous System-Penetrating Protease Inhibitor Overcomes Human Immunodeficiency Virus 1 Resistance with Unprecedented AM to PM Potency. *Elife* **2017**, *6*, 1–25. https://doi.org/10.7554/eLife.28020.

(268) Ghosh, A. K.; Rao, K. V.; Nyalapatla, P. R.; Osswald, H. L.; Martyr, C. D.; Aoki, M.; Hayashi, H.; Agniswamy, J.; Wang, Y.-F.; Bulut, H.; et al. Design and Development of Highly Potent HIV-1 Protease Inhibitors with a Crown-Like Oxotricyclic Core as the P2-Ligand To Combat Multidrug-Resistant HIV Variants. *J. Med. Chem.* **2017**, acs.jmedchem.7b00172. https://doi.org/10.1021/acs.jmedchem.7b00172.

(269) Hattori, S.-I.; Hayashi, H.; Bulut, H.; Rao, K. V.; Nyalapatla, P. R.; Hasegawa, K.; Aoki, M.; Ghosh, A. K.; Mitsuya, H. Halogen Bond Interactions of Novel HIV-1 Protease Inhibitors (PI)(GRL-001-15 and GRL-003-15) with the Flap of Protease Are Critical for Their Potent Activity against Wild-Type and Multi-PI-Resistant HIV-1 Variants. *Antimicrob. Agents Chemother.* **2019**, AAC.02635-18. https://doi.org/10.1128/AAC.02635-18.

(270) Ghosh, A. K.; Yashchuk, S.; Mizuno, A.; Chakraborty, N.; Agniswamy, J.; Wang, Y.-F.; Aoki, M.; Gomez, P. M. S.; Amano, M.; Weber, I. T.; et al. Design of *Gem* -Difluoro- *Bis* -Tetrahydrofuran as P2 Ligand for HIV-1 Protease Inhibitors to Improve Brain

Penetration: Synthesis, X-Ray Studies, and Biological Evaluation. *ChemMedChem* **2015**, *10* (1), 107–115. https://doi.org/10.1002/cmdc.201402358.

(271) Otwinowski, Z.; Minor, W. *Processing of X-Ray Diffraction Data Collected in Oscillation Mode*; Methods in Enzymology; Elsevier, 1997; Vol. 276. https://doi.org/10.1016/S0076-6879(97)76066-X.

(272) McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J. Phaser Crystallographic Software. *J. Appl. Crystallogr.* **2007**, *40* (Pt 4), 658–674. https://doi.org/10.1107/S0021889807021206.

(273) Winn, M. D.; Ballard, C. C.; Cowtan, K. D.; Dodson, E. J.; Emsley, P.; Evans, P. R.; Keegan, R. M.; Krissinel, E. B.; Leslie, A. G. W.; McCoy, A.; et al. Overview of the CCP4 Suite and Current Developments. *Acta Crystallogr. D. Biol. Crystallogr.* **2011**, *67* (Pt 4), 235–242. https://doi.org/10.1107/S0907444910045749.

(274) Emsley, P.; Cowtan, K. Coot: Model-Building Tools for Molecular Graphics. *Acta Crystallogr. D. Biol. Crystallogr.* **2004**, *60* (Pt 12 Pt 1), 2126–2132. https://doi.org/10.1107/S0907444904019158.

(275) Murshudov, G. N.; Skubák, P.; Lebedev, A. A.; Pannu, N. S.; Steiner, R. A.; Nicholls, R. A.; Winn, M. D.; Long, F.; Vagin, A. A. REFMAC5 for the Refinement of Macromolecular Crystal Structures. *Acta Crystallogr. D. Biol. Crystallogr.* **2011**, *67* (4), 355–367. https://doi.org/10.1107/S0907444911001314.

(276) DeLano, W. L. Pymol: An Open-Source Molecular Graphics Tool. *CCP4 Newsl. Protein Crystallogr.* **2002**, *40*, 82–92.

(277) King, N. M.; Prabu-Jeyabalan, M.; Nalivaika, E. A.; Wigerinck, P.; de Béthune, M.-P.; Schiffer, C. A. Structural and Thermodynamic Basis for the Binding of TMC114, a next-

Generation Human Immunodeficiency Virus Type 1 Protease Inhibitor. *J. Virol.* **2004**, *78* (21), 12012–12021. https://doi.org/10.1128/JVI.78.21.12012-12021.2004.

(278) Brower, E. T.; Bacha, U. M.; Kawasaki, Y.; Freire, E. Inhibition of HIV-2 Protease by HIV-1 Protease Inhibitors in Clinical Use. *Chem. Biol. Drug Des.* **2008**, *71* (4), 298–305. https://doi.org/10.1111/j.1747-0285.2008.00647.x.

(279) Flor-Parra, F.; Pérez-Pulido, A. J.; Pachón, J.; Pérez-Romero, P. The HIV Type 1 Protease L10I Minor Mutation Decreases Replication Capacity and Confers Resistance to Protease Inhibitors. *AIDS Res. Hum. Retroviruses* **2011**, *27* (1), 65–70. https://doi.org/10.1089/aid.2010.0072.

(280) Deeks, S. G.; Lewin, S. R.; Ross, A. L.; Ananworanich, J.; Benkirane, M.; Cannon, P.; Chomont, N.; Douek, D.; Lifson, J. D.; Lo, Y.-R.; et al. International AIDS Society Global Scientific Strategy: Towards an HIV Cure 2016. *Nat. Med.* **2016**, *22* (8), 839–850. https://doi.org/10.1038/nm.4108.

(281) Ghosh, A. K.; Osswald, H. L.; Prato, G. Recent Progress in the Development of HIV-1 Protease Inhibitors for the Treatment of HIV/AIDS. *J. Med. Chem.* **2016**, *59* (11), 5172–5208. https://doi.org/10.1021/acs.jmedchem.5b01697.

(282) Piot, P.; Abdool Karim, S. S.; Hecht, R.; Legido-Quigley, H.; Buse, K.; Stover, J.; Resch, S.; Ryckman, T.; Møgedal, S.; Dybul, M.; et al. Defeating AIDS—Advancing Global Health. *Lancet* **2015**, *386* (9989), 171–218. https://doi.org/10.1016/S0140-6736(15)60658-4.

(283) Shen, C.-H.; Tie, Y.; Yu, X.; Wang, Y.-F.; Kovalevsky, A. Y.; Harrison, R. W.; Weber, I. T. Capturing the Reaction Pathway in Near-Atomic-Resolution Crystal Structures of HIV-1 Protease. *Biochemistry* **2012**, *51* (39), 7726–7732. https://doi.org/10.1021/bi3008092.

(284) Hué, S.; Gifford, R. J.; Dunn, D.; Fernhill, E.; Pillay, D.; UK Collaborative Group on HIV Drug Resistance. Demonstration of Sustained Drug-Resistant Human Immunodeficiency Virus Type 1 Lineages Circulating among Treatment-Naïve Individuals. *J. Virol.* **2009**, *83* (6), 2645–2654. https://doi.org/10.1128/JVI.01556-08.

(285) Babrzadeh, F.; Varghese, V.; Pacold, M.; Liu, T. F.; Nyrén, P.; Schiffer, C.; Fessel, W. J.; Shafer, R. W. Collinearity of Protease Mutations in HIV-1 Samples with High-Level Protease Inhibitor Class Resistance. *J. Antimicrob. Chemother.* **2013**, *68* (2), 414–418. https://doi.org/10.1093/jac/dks409.

(286) Perryman, A. L.; Lin, J.-H.; McCammon, J. A. HIV-1 Protease Molecular Dynamics of a Wild-Type and of the V82F/I84V Mutant: Possible Contributions to Drug Resistance and a Potential New Target Site for Drugs. *Protein Sci.* **2004**, *13* (4), 1108–1123. https://doi.org/10.1110/ps.03468904.

(287) Sayer, J. M.; Agniswamy, J.; Weber, I. T.; Louis, J. M. Autocatalytic Maturation, Physical/Chemical Properties, and Crystal Structure of Group N HIV-1 Protease: Relevance to Drug Resistance. *Protein Sci.* **2010**, *19* (11), 2055–2072. https://doi.org/10.1002/pro.486.

(288) Louis, J. M.; Ishima, R.; Aniana, A.; Sayer, J. M. Revealing the Dimer Dissociation and Existence of a Folded Monomer of the Mature HIV-2 Protease. *Protein Sci.* **2009**, *18* (12), 2442–2453. https://doi.org/10.1002/pro.261.

(289) Storoni, L. C.; McCoy, A. J.; Read, R. J. Likelihood-Enhanced Fast Rotation Functions. *Acta Crystallogr. D. Biol. Crystallogr.* **2004**, *60* (Pt 3), 432–438. https://doi.org/10.1107/S0907444903028956.

(290) McCoy, A. J.; Grosse-Kunstleve, R. W.; Storoni, L. C.; Read, R. J. Likelihood-Enhanced

Fast Translation Functions. *Acta Crystallogr. D. Biol. Crystallogr.* **2005**, *61* (Pt 4), 458–464. https://doi.org/10.1107/S0907444905001617.

(291) Clemente, J. C.; Moose, R. E.; Hemrajani, R.; Whitford, L. R. S.; Govindasamy, L.; Reutzel, R.; McKenna, R.; Agbandje-McKenna, M.; Goodenow, M. M.; Dunn, B. M. Comparing the Accumulation of Active- and Nonactive-Site Mutations in the HIV-1 Protease. *Biochemistry* **2004**, *43* (38), 12141–12151. https://doi.org/10.1021/bi049459m.

(292) Sheldrick, G. M.; Schneider, T. R. *SHELXL: High-Resolution Refinement*; Methods in Enzymology; Elsevier, 1997; Vol. 277. https://doi.org/10.1016/S0076-6879(97)77018-6.

(293) Murshudov, G. N.; Vagin, A. A.; Dodson, E. J. Refinement of Macromolecular Structures by the Maximum-Likelihood Method. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **1997**, *53* (3), 240–255. https://doi.org/10.1107/S0907444996012255.

(294) Sheldrick, G. M.; IUCr. Crystal Structure Refinement with *SHELXL*. *Acta Crystallogr. Sect. C Struct. Chem.* **2015**, *71* (1), 3–8. https://doi.org/10.1107/S2053229614024218.

(295) Jones, T. A.; Zou, J. Y.; Cowan, S. W.; Kjeldgaard, M.; IUCr. Improved Methods for Building Protein Models in Electron Density Maps and the Location of Errors in These Models. *Acta Crystallogr. Sect. A Found. Crystallogr.* **1991**, *47* (2), 110–119. https://doi.org/10.1107/S0108767390010224.

(296) Tomasselli, A. G.; Olsen, M. K.; Hui, J. O.; Staples, D. J.; Sawyer, T. K.; Heinrikson, R. L.; Tomich, C. S. C. Substrate Analog Inhibition and Active Site Titration of Purified Recombinant HIV-1 Protease. *Biochemistry* **1990**, *29* (1), 264–269. https://doi.org/10.1021/bi00453a036.

(297) Tözsér, J.; Bláha, I.; Copeland, T. D.; Wondrak, E. M.; Oroszlan, S. Comparison of the HIV-1 and HIV-2 Proteinases Using Oligopeptide Substrates Representing Cleavage Sites

in Gag and Gag-Pol Polyproteins. *FEBS Lett.* **1991**, *281* (1–2), 77–80. https://doi.org/10.1016/0014-5793(91)80362-7.

(298) Liu, F.; Kovalevsky, A. Y.; Tie, Y.; Ghosh, A. K.; Harrison, R. W.; Weber, I. T. Effect of Flap Mutations on Structure of HIV-1 Protease and Inhibition by Saquinavir and Darunavir. *J. Mol. Biol.* **2008**, *381* (1), 102–115. https://doi.org/10.1016/j.jmb.2008.05.062.

(299) Kovalevsky, A. Y.; Liu, F.; Leshchenko, S.; Ghosh, A. K.; Louis, J. M.; Harrison, R. W.; Weber, I. T. Ultra-High Resolution Crystal Structure of HIV-1 Protease Mutant Reveals Two Binding Sites for Clinical Inhibitor TMC114. *J. Mol. Biol.* **2006**, *363* (1), 161–173. https://doi.org/10.1016/j.jmb.2006.08.007.

(300) Shafer, R. W. Rationale and Uses of a Public HIV Drug-Resistance Database. *J. Infect. Dis.* **2006**, *194* (s1), S51–S58. https://doi.org/10.1086/505356.

(301) Palmer, S.; Shafer, R. W.; Merigan, T. C. Highly Drug-Resistant HIV-1 Clinical Isolates Are Cross-Resistant to Many Antiretroviral Compounds in Current Clinical Development. *AIDS* **1999**, *13* (6), 661–667.

(302) Ronald M. Klabe, *,‡; Lee T. Bacheler, ‡; Paul J. Ala, §; Susan Erickson-Viitanen, ‡ and; Meek‡, J. L. Resistance to HIV Protease Inhibitors: A Comparison of Enzyme Inhibition and Antiviral Potency. **1998**. https://doi.org/10.1021/BI972555L.

(303) Gulnik, S. V.; Suvorov, L. I.; Liu, B.; Yu, B.; Anderson, B.; Mitsuya, H.; Erickson, J. W. Kinetic Characterization and Cross-Resistance Patterns Of HIV-1 Protease Mutants Selected under Drug Pressure. *Biochemistry* **1995**, *34* (29), 9282–9287. https://doi.org/10.1021/bi00029a002.

(304) Tie, Y.; Boross, P. I.; Wang, Y.-F.; Gaddis, L.; Hussain, A. K.; Leshchenko, S.; Ghosh, A.

K.; Louis, J. M.; Harrison, R. W.; Weber, I. T. High Resolution Crystal Structures of HIV-1 Protease with a Potent Non-Peptide Inhibitor (UIC-94017) Active against Multi-Drug-Resistant Clinical Strains. *J. Mol. Biol.* **2004**, *338* (2), 341–352. https://doi.org/10.1016/j.jmb.2004.02.052.

(305) 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. ABT-538 Is a Potent Inhibitor of Human Immunodeficiency Virus Protease and Has High Oral Bioavailability in Humans. *Proc. Natl. Acad. Sci.* **1995**, *92* (7), 2484–2488. https://doi.org/10.1073/PNAS.92.7.2484.

(306) Prabu-Jeyabalan, M.; Nalivaika, E. A.; King, N. M.; Schiffer, C. A. Viability of a Drug-Resistant Human Immunodeficiency Virus Type 1 Protease Variant: Structural Insights for Better Antiviral Therapy. *J. Virol.* **2003**, *77* (2), 1306–1315. https://doi.org/10.1128/jvi.77.2.1306-1315.2003.

(307) Bally, F.; Martinez, R.; Peters, S.; Sudre, P.; Telenti, A. Polymorphism of HIV Type 1 Gag P7/P1 and P1/P6 Cleavage Sites: Clinical Significance and Implications for Resistance to Protease Inhibitors. *AIDS Res. Hum. Retroviruses* **2000**, *16* (13), 1209–1213. https://doi.org/10.1089/08892220050116970.

(308) Mammano, F.; Petit, C.; Clavel, F. Resistance-Associated Loss of Viral Fitness in Human Immunodeficiency Virus Type 1: Phenotypic Analysis of Protease and Gag Coevolution in Protease Inhibitor-Treated Patients. *J. Virol.* **1998**, *72* (9), 7632–7637.

(309) Pettit, S. C.; Sheng, N.; Tritch, R.; Erickson-Viitanen, S.; Swanstrom, R. The Regulation of Sequential Processing of HIV-1 Gag by the Viral Protease; Springer, Boston, MA, 1998; pp 15–25. https://doi.org/10.1007/978-1-4615-5373-1_2.

(310) La Seta Catamancio, S.; De Pasquale, M. P.; Citterio, P.; Kurtagic, S.; Galli, M.; Rusconi,

S. In Vitro Evolution of the Human Immunodeficiency Virus Type 1 Gag-Protease Region and Maintenance of Reverse Transcriptase Resistance Following Prolonged Drug Exposure. *J. Clin. Microbiol.* **2001**, *39* (3), 1124–1129. https://doi.org/10.1128/JCM.39.3.1124-1129.2001.

(311) Kantor, R.; Fessel, W. J.; Zolopa, A. R.; Israelski, D.; Shulman, N.; Montoya, J. G.; Harbour, M.; Schapiro, J. M.; Shafer, R. W. Evolution of Primary Protease Inhibitor Resistance Mutations during Protease Inhibitor Salvage Therapy. *Antimicrob. Agents Chemother.* **2002**, *46* (4), 1086–1092. https://doi.org/10.1128/aac.46.4.1086-1092.2002.

(312) José C. Clemente, ‡; Roxana M. Coman, §; Michele M. Thiaville, §; Linda K. Janka, §; Jennifer A. Jeung, §; Sarawut Nukoolkarn, ∥; Lakshmanan Govindasamy, §; Mavis Agbandje-McKenna, §; Robert McKenna, §; Wichet Leelamanit, ∥; et al. Analysis of HIV-1 CRF_01 A/E Protease Inhibitor Resistance: Structural Determinants for Maintaining Sensitivity and Developing Resistance to Atazanavir†. **2006**. https://doi.org/10.1021/BI051886S.

(313) Shahriar, R.; Rhee, S.-Y.; Liu, T. F.; Fessel, W. J.; Scarsella, A.; Towner, W.; Holmes, S. P.; Zolopa, A. R.; Shafer, R. W. Nonpolymorphic Human Immunodeficiency Virus Type 1 Protease and Reverse Transcriptase Treatment-Selected Mutations. *Antimicrob. Agents Chemother.* **2009**, *53* (11), 4869–4878. https://doi.org/10.1128/AAC.00592-09.

(314) Svicher, V.; Ceccherini-Silberstein, F.; Erba, F.; Santoro, M.; Gori, C.; Bellocchi, M. C.; Giannella, S.; Trotta, M. P.; Monforte, A. d'Arminio; Antinori, A.; et al. Novel Human Immunodeficiency Virus Type 1 Protease Mutations Potentially Involved in Resistance to Protease Inhibitors. *Antimicrob. Agents Chemother.* **2005**, *49* (5), 2015–2025. https://doi.org/10.1128/AAC.49.5.2015-2025.2005.

(315) Prashar, V.; Bihani, S. C.; Das, A.; Rao, D. R.; Hosur, M. V. Insights into the Mechanism of Drug Resistance: X-Ray Structure Analysis of G48V/C95F Tethered HIV-1 Protease Dimer/Saquinavir Complex. *Biochem. Biophys. Res. Commun.* **2010**, *396* (4), 1018–1023. https://doi.org/10.1016/J.BBRC.2010.05.049.

(316) Boden, D.; Markowitz, M. Resistance to Human Immunodeficiency Virus Type 1 Protease Inhibitors. *Antimicrob. Agents Chemother.* **1998**, *42* (11), 2775–2783. https://doi.org/10.1128/AAC.42.11.2775.

(317) Liu, F.; Kovalevsky, A. Y.; Louis, J. M.; Boross, P. I.; Wang, Y.-F.; Harrison, R. W.; Weber, I. T. Mechanism of Drug Resistance Revealed by the Crystal Structure of the Unliganded HIV-1 Protease with F53L Mutation. *J. Mol. Biol.* **2006**, *358* (5), 1191–1199. https://doi.org/10.1016/j.jmb.2006.02.076.

(318) Nakashima, M.; Ode, H.; Suzuki, K.; Fujino, M.; Maejima, M.; Kimura, Y.; Masaoka, T.; Hattori, J.; Matsuda, M.; Hachiya, A.; et al. Unique Flap Conformation in an HIV-1 Protease with High-Level Darunavir Resistance. *Front. Microbiol.* **2016**, *7* (FEB). https://doi.org/10.3389/fmicb.2016.00061.

(319) Clutter, D. S.; Jordan, M. R.; Bertagnolio, S.; Shafer, R. W. HIV-1 Drug Resistance and Resistance Testing. *Infect. Genet. Evol.* **2016**, *46*, 292–307. https://doi.org/10.1016/J.MEEGID.2016.08.031.

(320) Menéndez-Arias, L. Molecular Basis of Human Immunodeficiency Virus Type 1 Drug Resistance: Overview and Recent Developments. *Antiviral Res.* **2013**, *98* (1), 93–120. https://doi.org/10.1016/J.ANTIVIRAL.2013.01.007.

(321) Nijhuis, M.; Schuurman, R.; de Jong, D.; Erickson, J.; Gustchina, E.; Albert, J.; Schipper, P.; Gulnik, S.; Boucher, C. A. Increased Fitness of Drug Resistant HIV-1 Protease as a

Result of Acquisition of Compensatory Mutations during Suboptimal Therapy. *AIDS* **1999**, *13* (17), 2349–2359.

(322) van Maarseveen, N. M.; Wensing, A. M. J.; Jong, D. de; Taconis, M.; Borleffs, J. C. C.; Boucher, C. A. B.; Nijhuis, M. Persistence of HIV-1 Variants with Multiple Protease Inhibitor (PI)–Resistance Mutations in the Absence of PI Therapy Can Be Explained by Compensatory Fixation. *J. Infect. Dis.* **2007**, *195* (3), 399–409. https://doi.org/10.1086/510533.

(323) Roche, J.; Louis, J. M.; Bax, A. Conformation of Inhibitor-Free HIV-1 Protease Derived from NMR Spectroscopy in a Weakly Oriented Solution. *Chembiochem* **2014**, *16* (2), 214–218. https://doi.org/10.1002/cbic.201402585.

(324) Rhee, S.-Y.; Gonzales, M. J.; Kantor, R.; Betts, B. J.; Ravela, J.; Shafer, R. W. Human Immunodeficiency Virus Reverse Transcriptase and Protease Sequence Database. *Nucleic Acids Res.* **2003**, *31* (1), 298–303.

(325) Mildner, A. M.; Rothrock, D. J.; Leone, J. W.; Bannow, C. A.; Lull, J. M.; Reardon, I. M.; Sarcich, J. L.; Howe, W. J.; Tomich, C.-S. C.; Smith, C. W.; et al. The HIV-1 Protease as Enzyme and Substrate: Mutagenesis of Autolysis Sites and Generation of a Stable Mutant with Retained Kinetic Properties. *Biochemistry* **1994**, *33* (32), 9405–9413. https://doi.org/10.1021/bi00198a005.

(326) Szeltner, Z.; Polgár, L. Conformational Stability and Catalytic Activity of HIV-1 Protease Are Both Enhanced at High Salt Concentration. *J. Biol. Chem.* **1996**, *271* (10), 5458–5463. https://doi.org/10.1074/jbc.271.10.5458.

(327) Pawar, S.; Wang, Y.-F.; Wong-Sam, A.; Agniswamy, J.; Ghosh, A. K.; Harrison, R. W.; Weber, I. T. Structural Studies of Antiviral Inhibitor with HIV-1 Protease Bearing Drug

Resistant Substitutions of V32I, I47V and V82I. *Biochem. Biophys. Res. Commun.* **2019**. https://doi.org/10.1016/J.BBRC.2019.05.064.

(328) Krissinel, E.; Henrick, K.; IUCr. Secondary-Structure Matching (SSM), a New Tool for Fast Protein Structure Alignment in Three Dimensions. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2004**, *60* (12), 2256–2268. https://doi.org/10.1107/S0907444904026460.

(329) Harrison, R. W.; Weber, I. T. Molecular Dynamics Simulations of HIV-1 Protease with Peptide Substrate. *"Protein Eng. Des. Sel.* **1994**, *7* (11), 1353–1363. https://doi.org/10.1093/protein/7.11.1353.

(330) Harrison, R. W. Stiffness and Energy Conservation in Molecular Dynamics: An Improved Integrator. *J. Comput. Chem.* **1993**, *14* (9), 1112–1122. https://doi.org/10.1002/jcc.540140912.

(331) Fang, B.; Fu, G.; Agniswamy, J.; Harrison, R. W.; Weber, I. T. Caspase-3 Binds Diverse P4 Residues in Peptides as Revealed by Crystallography and Structural Modeling. *Apoptosis* **2009**, *14* (5), 741–752. https://doi.org/10.1007/s10495-009-0333-y.

(332) Bagossi, P.; Zahuczky, G.; Tözsér, J.; Weber, I. T.; Harrison, R. W. Improved Parameters for Generating Partial Charges: Correlation with Observed Dipole Moments. *J. Mol. Model.* **1999**, *5* (9), 143–152. https://doi.org/10.1007/s008940050114.

(333) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. Ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from Ff99SB. *J. Chem. Theory Comput.* **2015**, *11* (8), 3696–3713. https://doi.org/10.1021/acs.jctc.5b00255.

(334) Stigliani, J.-L.; Bernardes-Génisson, V.; Bernadou, J.; Pratviel, G. Cross-Docking Study on InhA Inhibitors: A Combination of Autodock Vina and PM6-DH2 Simulations to

Retrieve Bio-Active Conformations. *Org. Biomol. Chem.* **2012**, *10* (31), 6341. https://doi.org/10.1039/c2ob25602a.

(335) 188.ammp: SPEC CPU2000 Benchmark Description https://www.spec.org/osg/cpu2000/CFP2000/188.ammp/docs/188.ammp.html (accessed Nov 7, 2019).

(336) Chen, X.; Weber, I. T.; Harrison, R. W. Molecular Dynamics Simulations of 14 HIV Protease Mutants in Complexes with Indinavir. *J. Mol. Model.* **2004**, *10* (5–6), 373–381. https://doi.org/10.1007/s00894-004-0205-x.

(337) Muzammil, S.; Armstrong, A. A.; Kang, L. W.; Jakalian, A.; Bonneau, P. R.; Schmelmer, V.; Amzel, L. M.; Freire, E. Unique Thermodynamic Response of Tipranavir to Human Immunodeficiency Virus Type 1 Protease Drug Resistance Mutations. *J. Virol.* **2007**, *81* (10), 5144–5154. https://doi.org/10.1128/JVI.02706-06.

(338) Ohtaka, H.; Schön, A.; Freire, E. Multidrug Resistance to HIV-1 Protease Inhibition Requires Cooperative Coupling between Distal Mutations. *Biochemistry* **2003**, *42* (46), 13659–13666. https://doi.org/10.1021/bi0350405.

(339) Muzammil, S.; Ross, P.; Freire, E. A Major Role for a Set of Non-Active Site Mutations in the Development of HIV-1 Protease Drug Resistance. *Biochemistry* **2003**, *42* (3), 631–638. https://doi.org/10.1021/bi027019u.

(340) Meher, B. R.; Wang, Y. Exploring the Drug Resistance of V32I and M46L Mutant HIV-1 Protease to Inhibitor TMC114: Flap Dynamics and Binding Mechanism. *J. Mol. Graph. Model.* **2015**, *56*, 60–73. https://doi.org/10.1016/j.jmgm.2014.11.003.

(341) Liu, Z.; Huang, X.; Hu, L.; Pham, L.; Poole, K. M.; Tang, Y.; Mahon, B. P.; Tang, W.; Li, K.; Goldfarb, N. E.; et al. Effects of Hinge-Region Natural Polymorphisms on Human

Immunodeficiency Virus-Type 1 Protease Structure, Dynamics, and Drug Pressure Evolution. *J. Biol. Chem.* **2016**, *291* (43), 22741–22756. https://doi.org/10.1074/jbc.M116.747568.

(342) Henes, M.; Kosovrasti, K.; Lockbaum, G. J.; Leidner, F.; Nachum, G. S.; Nalivaika, E. A.; Bolon, D. N. A.; Kurt Yilmaz, N.; Schiffer, C. A.; Whitfield, T. W. Molecular Determinants of Epistasis in HIV-1 Protease: Elucidating the Interdependence of L89V and L90M Mutations in Resistance. *Biochemistry* **2019**, *58* (35), 3711–3726. https://doi.org/10.1021/acs.biochem.9b00446.

(343) Gardner, J. M.; Abrams, C. F. Energetics of Flap Opening in HIV-1 Protease: String Method Calculations. *J. Phys. Chem. B* **2019**. https://doi.org/10.1021/ACS.JPCB.9B08348.

(344) Henes, M.; Lockbaum, G. J.; Kosovrasti, K.; Leidner, F.; Nachum, G. S.; Nalivaika, E. A.; Lee, S.-K.; Spielvogel, E.; Zhou, S.; Swanstrom, R.; et al. Picomolar to Micromolar: Elucidating the Role of Distal Mutations in HIV-1 Protease in Conferring Drug Resistance. *ACS Chem. Biol.* **2019**, acschembio.9b00370. https://doi.org/10.1021/acschembio.9b00370.

(345) Yu, Y.; Wang, J.; Shao, Q.; Shi, J.; Zhu, W. Effects of Drug-Resistant Mutations on the Dynamic Properties of HIV-1 Protease and Inhibition by Amprenavir and Darunavir. *Sci. Rep.* **2015**, *5* (1), 10517. https://doi.org/10.1038/srep10517.

(346) Clemente, J. C.; Hemrajani, R.; Blum, L. E.; Goodenow, M. M.; Dunn, B. M. Secondary Mutations M36I and A71V in the Human Immunodeficiency Virus Type 1 Protease Can Provide an Advantage for the Emergence of the Primary Mutation D30N †. *Biochemistry* **2003**, *42* (51), 15029–15035. https://doi.org/10.1021/bi035701y.

(347) Gonzalez, L. M. F.; Santos, A. F.; Abecasis, A. B.; Van Laethem, K.; Soares, E. A.; Deforche, K.; Tanuri, A.; Camacho, R.; Vandamme, A.-M.; Soares, M. A. Impact of HIV-1 Protease Mutations A71V/T and T74S on M89I/V-Mediated Protease Inhibitor Resistance in Subtype G Isolates. *J. Antimicrob. Chemother.* **2008**, *61* (6), 1201–1204. https://doi.org/10.1093/jac/dkn099.

(348) Rajasekaran, N.; Suresh, S.; Gopi, S.; Raman, K.; Naganathan, A. N. A General Mechanism for the Propagation of Mutational Effects in Proteins. *Biochemistry* **2017**, *56* (1), 294–305. https://doi.org/10.1021/acs.biochem.6b00798.

(349) Naganathan, A. N. Modulation of Allosteric Coupling by Mutations: From Protein Dynamics and Packing to Altered Native Ensembles and Function. *Curr. Opin. Struct. Biol.* **2019**, *54*, 1–9. https://doi.org/10.1016/J.SBI.2018.09.004.

(350) Foulkes-Murzycki, J. E.; Scott, W. R. P.; Schiffer, C. A. Hydrophobic Sliding: A Possible Mechanism for Drug Resistance in Human Immunodeficiency Virus Type 1 Protease. *Structure* **2007**, *15* (2), 225–233. https://doi.org/10.1016/j.str.2007.01.006.

(351) Baxter, J. D.; Schapiro, J. M.; Boucher, C. A. B.; Kohlbrenner, V. M.; Hall, D. B.; Scherer, J. R.; Mayers, D. L. Genotypic Changes in Human Immunodeficiency Virus Type 1 Protease Associated with Reduced Susceptibility and Virologic Response to the Protease Inhibitor Tipranavir. *J. Virol.* **2006**, *80* (21), 10794–10801. https://doi.org/10.1128/JVI.00712-06.

(352) Llibre, J. M.; Schapiro, J. M.; Clotet, B. Clinical Implications of Genotypic Resistance to the Newer Antiretroviral Drugs in HIV-1–Infected Patients with Virological Failure. *Clin. Infect. Dis.* **2010**, *50* (6), 872–881. https://doi.org/10.1086/650732.

(353) Bennett, D. E.; Camacho, R. J.; Otelea, D.; Kuritzkes, D. R.; Fleury, H.; Kiuchi, M.;

Heneine, W.; Kantor, R.; Jordan, M. R.; Schapiro, J. M.; et al. Drug Resistance Mutations for Surveillance of Transmitted HIV-1 Drug-Resistance: 2009 Update. *PLoS One* **2009**, *4* (3), e4724. https://doi.org/10.1371/journal.pone.0004724.

(354) Varghese, V.; Mitsuya, Y.; Fessel, W. J.; Liu, T. F.; Melikian, G. L.; Katzenstein, D. A.; Schiffer, C. A.; Holmes, S. P.; Shafer, R. W. Prototypical Recombinant Multi-Protease-Inhibitor-Resistant Infectious Molecular Clones of Human Immunodeficiency Virus Type 1. *Antimicrob. Agents Chemother.* **2013**, *57* (9), 4290–4299. https://doi.org/10.1128/AAC.00614-13.

(355) Le, T.; Chiarella, J.; Simen, B. B.; Hanczaruk, B.; Egholm, M.; Landry, M. L.; Dieckhaus, K.; Rosen, M. I.; Kozal, M. J. Low-Abundance HIV Drug-Resistant Viral Variants in Treatment-Experienced Persons Correlate with Historical Antiretroviral Use. *PLoS One* **2009**, *4* (6), e6079. https://doi.org/10.1371/journal.pone.0006079.

(356) Rhee, S.-Y.; Taylor, J.; Wadhera, G.; Ben-Hur, A.; Brutlag, D. L.; Shafer, R. W. Genotypic Predictors of Human Immunodeficiency Virus Type 1 Drug Resistance. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103* (46), 17355–17360. https://doi.org/10.1073/pnas.0607274103.

(357) Chang, M. W.; Torbett, B. E. Accessory Mutations Maintain Stability in Drug-Resistant HIV-1 Protease. *J. Mol. Biol.* **2011**, *410* (4), 756–760. https://doi.org/10.1016/j.jmb.2011.03.038.

(358) de Vera, I. M. S.; Smith, A. N.; Dancel, M. C. A.; Huang, X.; Dunn, B. M.; Fanucci, G. E. Elucidating a Relationship between Conformational Sampling and Drug Resistance in HIV-1 Protease. *Biochemistry* **2013**, *52* (19), 3278–3288. https://doi.org/10.1021/bi400109d.

(359) Hornak, V.; Okur, A.; Rizzo, R. C.; Simmerling, C. HIV-1 Protease Flaps Spontaneously Open and Reclose in Molecular Dynamics Simulations. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103* (4), 915–920. https://doi.org/10.1073/pnas.0508452103.

(360) Cordone, L.; Ferrand, M.; Vitrano, E.; Zaccai, G. Harmonic Behavior of Trehalose-Coated Carbon-Monoxy-Myoglobin at High Temperature. *Biophys. J.* **1999**, *76* (2), 1043–1047. https://doi.org/10.1016/S0006-3495(99)77269-3.

(361) Anunciado, D. B.; Nyugen, V. P.; Hurst, G. B.; Doktycz, M. J.; Urban, V.; Langan, P.; Mamontov, E.; O'Neill, H. *In Vivo* Protein Dynamics on the Nanometer Length Scale and Nanosecond Time Scale. *J. Phys. Chem. Lett.* **2017**, *8* (8), 1899–1904. https://doi.org/10.1021/acs.jpclett.7b00399.

(362) P. A. Seeger; S. E. Rokop; P. D. Palmer; S. J. Henderson, †; D. E. Hobart, ‡ and; Trewhella*, J. Neutron Resonance Scattering Shows Specific Binding of Plutonium to the Calcium-Binding Sites of the Protein Calmodulin and Yields Precise Distance Information. **1997**. https://doi.org/10.1021/JA9633124.

(363) Balog, E.; Becker, T.; Oettl, M.; Lechner, R.; Daniel, R.; Finney, J.; Smith, J. C. Direct Determination of Vibrational Density of States Change on Ligand Binding to a Protein. *Phys. Rev. Lett.* **2004**, *93* (2), 028103. https://doi.org/10.1103/PhysRevLett.93.028103.

(364) Lu, D.-Y.; Wu, H.-Y.; Yarla, N. S.; Xu, B.; Ding, J.; Lu, T.-R. HAART in HIV/AIDS Treatments: Future Trends. *Infect. Disord. - Drug Targets* **2018**, *18* (1), 15–22. https://doi.org/10.2174/1871526517666170505122800.

(365) Baillie, T. A. Targeted Covalent Inhibitors for Drug Design. *Angew. Chemie Int. Ed.* **2016**, *55* (43), 13408–13421. https://doi.org/10.1002/anie.201601091.

(366) Zhao, H.; Caflisch, A. Molecular Dynamics in Drug Design. *Eur. J. Med. Chem.* **2015**, *91*,

4–14. https://doi.org/10.1016/J.EJMECH.2014.08.004.

APPENDICES

Appendix A: List of academic publications.

1. **Kneller DW**, Agniswamy J, Ghosh AK, Weber IT (2019) Potent antiviral HIV-1 protease inhibitor combats highly-drug resistant mutant PR20. *Biochemical and Biophysical Research Communications.* (519), 61-66.
2. Agniswamy J, **Kneller DW**, Brothers R, Wang Y-F, Harrison RW, Weber IT. (2019). Highly Drug-Resistant HIV-1 Protease Mutant PRS17 Shows Enhanced Binding to Substrate Analogues. *ACS Omega*. 4(5), 8707–8719.
3. **Kneller DW**, Gutzler SJ, Brewer MW. (2018). Using guided-inquiry experiments to characterize factors of osmosis and diffusion. *Principles of Biology Laboratory Manual 4th edition.* 23-34. Hayden-McNeil Publishing.
4. Weber IT, **Kneller DW**, Wong-Sam AE. (2015). Highly resistant HIV-1 proteases and strategies for their inhibition. *Future Medicinal Chemistry.* 7(8)1023-38.
5. Klimov, S, **Kneller DW,** Stone, RD, Mandric, I, Artsiomenka, A, Weber, IT, Harrison, R, Zelikovsky, A, Aneja, R, & Jiang, Y. (2015) Classifying Cancers from RNAseq Data through Machine Learning. Proceedings of ISBRA 2015, Norfolk, VA, VA, June 7-10, 2015.

Appendix B: Table of X-ray crystal structures deposited to the protein data bank

Protein	**Ligand (Ligand ID)**	**PDB accession code**	**Resolution (Å)**
PRS5B	DRV, **1** (017)	6P9A	1.66
PRS5B	APV (478)	6P9B	1.75
PR20	GRL-142, **2** (7OA)	6PRF	1.21

Appendix C: Table of published kinetic inhibition values for highly resistant PR mutants PR20, PRS17, and PRS5B

A dash (-) indicates data not published.

Inhibitor	K_i for PR20 (nM)	K_i for PRS17 (nM)	K_i for PRS5B (nM)
APV	178	11	3.6
DRV	41	50	4.0
ATV	-	70	8.4
TPV	-	-	8.6
NFV	-	5870	14.9
IDV	-	810	46
LPV	-	73	160
SQV	930	8390	1000
GRL 50-10A	1.7	-	-
GRL 44-10A	4.3	-	-
GRL 142-13A	2.5	-	-
GRL 0519A	39.5	-	-

Appendix D: Python Programs for AMMP Molecular Dynamics analysis

A 20 ns AMMP molecular dynamics simulation outputs a statistics and coordinate file every 10 ps for a total of 2000 individual files. Thus, analysis of MD simulations requires automated scripting as a tool to extract desired data. The Python programming language is an excellent choice for MD analysis due to a low-barrier learning curve and extensive library support. The following section details Python3 programs with selections of code used to analyze MD simulations from Chapter 4.

```
#measure_distance.py
#programs asks for file information and a range of frames to look for (filenames)
#returns the distance between two atoms over frame as a csv for plotting in excel

import math
import csv

def measure_distance(xyz1, xyz2):
      dist = math.sqrt(((xyz1[0] - xyz2[0])**2) + ((xyz1[1] - xyz2[1])**2) + ((xyz1[2] - xyz2[2])**2))
      return round(dist, 2)

#reads a .pdb file for the two given atoms, calls the measure_distance method
def readaframe(filename):
            with open(filename, 'r') as f:
                  line = csv.reader(f)
                  for i in line:
                        string = str(i)
                        parsed = string.split()
                        if parsed[2] == 'CA' and parsed[4] == str(aa1):
                              aa1_xyz1 = [parsed[5], parsed[6], parsed[7]]
                              xyz1 = [float(i) for i in aa1_xyz1]
                        if parsed[2] == 'CA' and parsed[4] == str(aa2):
                              aa2_xyz2 = [parsed[5], parsed[6], parsed[7]]
                              xyz2 = [float(i) for i in aa2_xyz2]
            return measure_distance(xyz1, xyz2)

#calls the readaframe method for every frame and puts each distance into a list
distances = [readaframe(i) for i in filenames]

#makes a iterator of tuples for the frame number and distance,  outputs csv file for plotting
data = zip(allframes, distances)
with open(prefix+"_"+str(aa1)+"_to_"+str(aa2)+".csv", 'w') as csvfile:
      writer = csv.writer(csvfile)
       for i in data:
            writer.writerow(i)
csvfile.close
```

Measure_distance.py measures the distance from given CA atoms for a given range of coordinate frames.

```
#getenergy.py
#takes filename information and returns the mean binding energy of inhibitors plus standard error

import statistics

#method opens stats.n file, looks for the desired energy number, and returns that number
def getenergy(filename):
    with open(filename, 'r') as f:
        for line in f:
            if "Vnonbon total external" in line:
                energyline = str(line)
                parsed = energyline.split()
        return float(parsed[3])

#makes a list containing the energy number for each file in the desired frame range
energies = [getenergy(i) for i in filenames]

mean = statistics.mean(energies)
sd = statistics.stdev(energies)
print ("The mean energy is " + str(round(mean, 2)) + " +- standard deviation of " + str(round(sd, 2)))
```

The script getenergy.py reads AMMP stats.n files and print the mean and standard deviation of the binding energy of inhibitors.

```
#gif.py
#creates a rendered ribbon image of the protease with inhibitor as sticks for each MD .pdb
#All .pdb files require a .pdb exention prior to running
#gif.py uses PyMol as the Python interpreter: pymol –c gif.py

#output names must be in 4 digit sequential numbers for proper conversion to .gif
outputnames = ["%04d" % i for i in range(1, 500)]

#loads file, orients view, sets colors and representaions, renders image, wipes for next file
def makeimage(pdb, outputname):
	pymol.cmd.load(pdb)
	pymol.cmd.hide("all")
	pymol.cmd.orient("resi 1-199")
	pymol.cmd.zoom("resi 1-199", -8)
	pymol.cmd.show("ribbon")
	pymol.cmd.color("marine", "resi 101-199")
	pymol.cmd.color("salmon", "resi 1-99")
	pymol.cmd.show("sticks", "resi 201") #inhibitor is residue 201
	pymol.cmd.hide("sticks", "hydrogen")
	pymol.cmd.util.cbag("resi 201")
	pymol.cmd.remove("resn HOH")
	pymol.cmd.png(outputname, 1280,720, 300, 1, 1)
	pymol.cmd.delete("all")

#makes an image for each .pdb file and names it as a sequential 4 digit number
for i,j in zip(pdbs, outputnames):
	makeimage(i,j)
pymol.cmd.quit()

#convert .pngs to give using imagemagick: convert -delay 10 -loop 0 *.png output.gif
```

Visualization of the MD simulations over the timescale are useful for understanding macro movements as well as being able to visual follow where specific elements, such as the flap water or catalytic hydrogen, are located over the timescale. When executed through the PyMol[276] python interpreter, gif.py creates an image for each .pdb file that can be converted into a .gif format for visualization.

Appendix E: Classifying Cancers from RNAseq Data through Machine Learning

Classifying Cancers from RNAseq Data through Machine Learning

An evaluation of machine learning methods for cancer origin using transcriptomic data

Sergey Klimov[1,3], Daniel Walter Kneller[1], Robert Douglas Stone Jr.[1], Igor Mandric[2], Aliaksandr Artsiomenka[2], Alex Zelikovsky[2], Ritu Aneja[1], Yi Jiang[3], Irene T. Weber[1], Robert W. Harrison[2]

[1]Georgia State University, Department of Biology, Atlanta, GA, United States; [2]Georgia State University, Department of Department of Computer Science, Atlanta, GA, United States; [3]Georgia State University, Department of Mathematics and Statistics, Atlanta, GA, United States.

***Abstract*— Cancer is the second leading cause of death in the United States, with a lifetime probability of cancer diagnosis 43% for men and 38% for women[1]. Early detection of cancer is paramount in effective treatment. Identification of tumor type can allow clinicians to provide personalized treatments, especially if clinical analysis of tumor type are inconclusive. A machine learning method to the analysis of transcriptomic data can identify primary tumor tissue type with 98% accuracy for five specific cancer types. Combined with an easy user interfaced website, clinician access to instant tumor type prediction is possible.**

Keywords*— *Cancer, Bioinformatics, Machine Learning, Tumor Typing, Gene Expression, SVM, Random Forests, Data Mining, Model Validation.

I. Introduction

Cancer genomics is a widely studied field within bioinformatics. Previous studies have used gene expression to identify risk with the help of different machine learning (ML) techniques including Artificial Neural Networks (ANNs), Bayesian Networks (BNs), Support Vector Machines (SVMs) and Decision Trees (DTs) [2]. While some studies look at genes unbiased[2 3 4 5 6] a few emphasize genes known to cause a changes in pathways which are well studied in tumor vs. normal cell analysis[3]. These studies are effective since they can assist in developing personalized therapies[4] and can be included with clinical factors to increase survival prediction[4], they are quite common.

One interesting application that can be used with gene expression information is the identification of tumor tissue type. This has a substantial benefit with metastatic tumors where the origin of the primary tumor is unknown[4 5]. It also can assist with the 3-5% of cases where even the primary tumor tissue of origin is ambiguous[3]. While this type of study has been attempted previously[3], we offer an alternative algorithm (SVM and random forests vs. a plain decision tree) and the clinical versatility of a website with instant prediction alongside powerful potential inferences by identifying peculiar genetic aberrations not seen in the average cancer.

II. Methods

A. Datasets

Sequence based gene expression data, with a GAF annotation, was mined from the International Cancer Genome Consortium (https://dcc.icgc.org/). Vectors of the 20,513, Lowess regression normalized, gene expressions were initialized for each unique submitted Sample ID. The introductory analysis was performed with data mined from 5 initial cancers (sample sizes): Breast (N = 1,177), Bladder (N = 318), Blood (N = 173), Brain (N = 619), and Cervix (N = 264).

B. Support Vector Machine (SVM)

The SVM was trained either based on a polynomial or linear kernel. Each of the constructed hyperplanes created were based on a one versus rest or a crammer and singer method (one versus all). A grid search allowed the combination of methods above, and the optimization of C (trades off misclassifications versus simplicity) and gamma (specifies the influence of individual training vectors) parameters.

C. Random Forest Classifier

In our model the random forest algorithm uses an ensemble of either 10, 20, or 50 decision trees. A grid search used the following hyper parameters for the data: Class weightage, which either gives classes an inverse proportional weight to the frequency of that class or none at all, and the ideal number of decision trees required.

D. Validation

Machine learning accuracy was tested by using the full samples of the 5 classes of cancers within a K-fold validation. The 2,551 samples were randomly divided into 5 equal sized (N = 510) groups with a single subsample used as validation of the model trained by the other four. This model building and validation was repeated 5 times, on a unique trained validation, and the accuracy averaged. The output of each testing vector can give the probabilities of belonging to a cancer, for validation purposes we simply used the cancer

with the largest probability. To highlight specific inconsistencies of cancer prediction a confusion matrix was created.

III. Results

A. Machine Learning Validation Results

The accuracy for each fold validation can be seen in Table I. The SVM's used and the random forest used 10 decision trees. The difference in the accuracy was found significant via a t-test ($p < 0.01$) and showed optimal performance with the random forest model.

TABLE I. Results of the 5-fold Cross Validation

Model\Fold	1	2	3	4	5	Mean(+/- std)
Polynomial SVM	0.45	0.45	0.44	0.47	0.46	0.45(0.01)
Linear SVM	0.93	0.92	0.90	0.91	0.89	0.91(0.03)
Random Forest	0.98	0.99	0.98	0.97	0.99	0.98(0.01)

The confusion matrices were generated by manually inputting novel cases into the finished models, either for the polynomial SVM (Table II) or the Random Forest (Table III). The polynomial SVM showed issues, especially stratifying blood cancer accurately, while the random forest was able to correctly discern all the test cases.

TABLE II. Confusion Matrix for Linear SVM

Cancer Type	Samples Tested	Tissue of Origin				
		BL	BLO	BR	BRE	CE
Bladder(BL)	5	3		1		1
Blood(BLO)	5			5		
Brain(BR)	5			5		
Breast(BRE)	6				5	1
Cervix(CE)	10	2				8

TABLE III. Confusion Matrix for Random Forest

Cancer Type	Samples Tested	Tissue of Origin				
		BL	BLO	BR	BRE	CE
Bladder(BL)	5	5				
Blood(BLO)	5		5			
Brain(BR)	5			5		
Breast(BRE)	5				5	
Cervix(CE)	5					5

IV. Discussion

We have evaluated multiple machine learning techniques for accurate identification of 5 cancer primary tissue types using gene expression information from clinical RNA sequencing data. Both the validation and the confusion matrix demonstrate the random forest as being the most accurate model, followed by the linear SVM. Interestingly the polynomial SVM either creates inadequate hyperplanes or provides evidence that some cancers may possess very similar expression (Brain and blood) profiles. The 98% accuracy of the random forest suggests it is a possible algorithm that can be used when traditional clinopathological methods may fail in identifying ambiguous cancer origins.

References.

[1] Cancer statistics, 2015. Siegel, R. L., Miller, K. D., & Jemal, A. CA: A Cancer Journal for Clinicians 2015.

[2] Machine learning applications in cancer prognosis and prediction. Konstantina Kourou et al. Computational and Structural Biotechnology Journal. 2015.

[3] Predicting breast cancer using an expression values weighted clinical classifier. Minta Thomas. BMC Bioinformatics 2014.

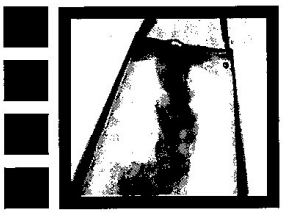
[4] MicroRNAs accurately identify cancer tissue origin. Nitzan Rosenfeld et al. Nature Biotechnology, April 2008.

[5] Case-Based Retrieval Framework for Gene Expression Data. Ali Anaissi et al. Cancer Inform 2015.

[6] Validation and Reproducibility of a Microarray-Based Gene Expression Test for Tumor Identification in Formalin-Fixed, Paraffin-Embedded Specimens. Raji Pillai et al. The Journal of Molecular Diagnostics, January 2011.

Appendix F: Lab Manual Chapter: Using Guided-Inquiry Experiments to Characterize Factors of Osmosis and Diffusion

MODULE 1



2B. Using Guided-Inquiry Experiments to Characterize Factors of Osmosis and Diffusion

Co-authored with Daniel W. Kneller

I. Objectives

In this lab, students will:

a. Utilize all steps of the scientific method to investigate different factors that influence diffusion of molecules across a semi-permeable membrane.

b. Work as a group to identify a simple experimental question and generate an experiment to test an original hypothesis.

c. Organize data and draw logical conclusions from the results.

d. Describe their results in a group report in which mechanics and formatting of general science writing will be assessed.

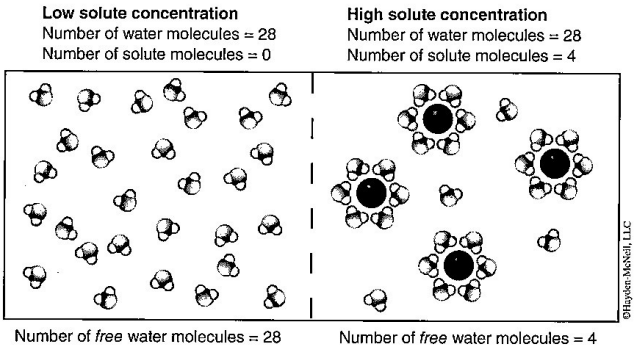
II. Introduction

In the previous experiment, you used different chemical indicators to test for the presence of macromolecules in unknown solutions. You followed predetermined instructions to perform an experiment that tested a prearranged hypothesis. In the present experiment, you will be tasked with identifying your own experimental question. In an effort to answer this question, you will derive a specific hypothesis and unique experiment protocol. The primary theme of your investigation is the variables that may influence diffusion and osmosis.

In order to ask an educated scientific question, you must start by collecting background information on the topic. Start with what you know, and then use inductive reasoning to develop a question about what you expect might be true. Begin by defining diffusion. **Diffusion** is the net movement of atoms or molecules from an area of high concentration to an area of low concentration. Molecules of a particular substance diffuse down the concentration gradient formed by that substance. Diffusion results from random motion of molecules driven by

thermal energy. It is important to understand that this is a *net* movement of solutes. Solute molecules are moving randomly in all directions but over time will eventually be distributed equally in a system. This movement does not require any additional energy input; therefore diffusion is a type of **passive transport**. In some cases, cells must force a substance against its concentration gradient, from an area of low concentration to an area of high concentration. Cells can utilize enzymes that require energy in the form of ATP to move molecules across a membrane. This process is known as **active transport**.

Osmosis is the diffusion of water across a semi-permeable membrane. Water will move from an area of low solute concentration, across a semi-permeable membrane, to an area of high solute concentration. When charged solutes, such as NaCl are dissolved, they dissociate and the individual sodium and chloride ions form hydrogen bonds with one or more water molecules. These water molecules are then associated with the ions and effectively lower the number of free water molecules in the solution. If fewer free water molecules occur on one side of a semi-permeable membrane, the free water concentration is effectively decreased on that side. The outcome of this is both a solute and free water concentration gradient. This will result in water moving from an area of higher free water concentration to an area of lower free water concentration. At the same time, Na^+ and Cl^- ions, if able, will pass through the membrane.

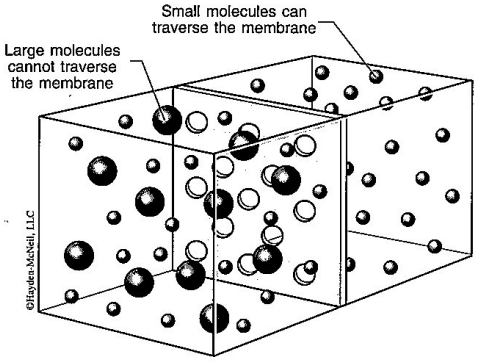


It is important for cells to allow substances in and out. For this to occur, molecules need to cross the semi-permeable cell membrane. The processes of diffusion and osmosis are critical to the physiology of the human body. For example, CO_2 diffuses out of capillaries that surround alveoli in the lungs. When in the alveolar air space, the CO_2 gas can then be expelled through exhalation. In this case, CO_2 moves through the semi-permeable membrane of the blood vessels, the area of its higher concentration, into the air sacs of the lungs where it is less concentrated. Many molecules are free to diffuse in or out of the cell. In order to do so, they must either be lipid-soluble, passing directly across the lipid bilayer, as in the case of steroid hormones; or, if they are polar, their passage must be facilitated by proteins. These channels, however, are not permeable to every polar molecule. They are selective based upon the size and charge of a specific solute, rendering the membrane semi-permeable, also referred to as selectively permeable.

Check Your Understanding

1. What does the term "semi-permeable" mean?
2. *Why* is the semi-permeable membrane critical to cell function?
3. *Why* would it be important for a cell to exchange substances with its surrounding environment?

Not all molecules can cross the cell membrane. The properties that are important for determining whether a substance can diffuse across a membrane include the size of the molecule, its polarity, and the presence of channels. Today you will use dialysis tubing to model a cell membrane. However, some properties differ between an actual plasma membrane and dialysis tubing. For example, polarity does not play a large role in the dialysis tube model. Dialysis tubing can act as a semi-permeable membrane by discriminating molecules based on size. The pore size of the dialysis tubing, in combination with the size of the molecules, determines whether a solute can diffuse freely though the pores. For example, a large molecule such as starch might be unable to traverse the pores of dialysis tubing while a small ion, such as sodium might freely move across.



III. EXPERIMENT—INQUIRY OF FACTORS INFLUENCING DIFFUSION OF MOLECULES ACROSS A SEMI-PERMEABLE MEMBRANE

Overview

As a lab group, you will now design an experiment to illustrate a factor that might affect the diffusion of water or solute molecules across a semi-permeable membrane. Each group will come up with an experimental question and original hypothesis regarding a variable thought to influence osmosis and/or diffusion. Your group must then design a feasible experiment to test this hypothesis using the resources available. Finally, as a group, you will convey the details of your experiment, along with your findings and conclusions, in a brief lab report.

Generating an Experimental Question, Hypothesis and Prediction—An Example
Consider different factors that might influence the movement of solutes or water across a semi-permeable membrane. You already know that simple diffusion requires kinetic energy and a concentration gradient. You can use this prior knowledge to propose an experimental question. Consider the following example experimental question:

Does the magnitude of the concentration gradient of sucrose affect the rate of osmosis?

In order to hypothesize an answer, the question should be clearly understood by dissecting into its components. "Magnitude of the concentration gradient of sucrose" refers to the size of the difference between two or more different concentrations of sucrose. "...rate of osmosis" refers to the movement of water through a semi-permeable membrane over a length of time. A good experimental question is specifically worded so that a single specific hypothesis would answer the question. You might use inductive reasoning to estimate that a large concentration gradient of sucrose would lead water to move through a water-permeable membrane faster than a small concentration gradient. This would be worded into a specific statement that addresses the experimental question like so:

A larger concentration gradient of sucrose will have a faster rate of osmosis relative to a small concentration gradient.

This hypothesis is valid because it is both specific to the experimental question and testable.

Brainstorm variables that might affect the way molecules and water move through a semi-permeable membrane. Generate a list of variables as well as ideas about how you might manipulate them. The following worksheets should serve as a guide to help you.

The following tables should be used as a guide to generate a hypothesis that you can test in the lab regarding osmosis and diffusion. This is strictly a brainstorming exercise. Once you have thought through several ideas, your group should come to a decision as to which experimental idea is most feasible to test in the lab with the materials available.

You should communicate with your TA(s) during this time to help clarify any parts of this process that you may have questions about. You may use any resources at your disposal to look up additional information—reference books, primary literature, university websites (.edu), etc. An example has been provided for you.

Name 2–3 topics/ideas/observed phenomena, relevant to factors affecting osmosis and diffusion, that you are interested in exploring in more detail:

1. The magnitude of a concentration gradient affects the rates of osmosis in a system.
2.
3.
4.

Are there current hypotheses out there that intrigue you about these topics? Or hypotheses that you have thought up on your own regarding what variables could be contributing to your phenomenon of interest?

HYPOTHESIS	HOW IS THIS PHENOMENON OBSERVED (QUALIFIED) OR MEASURED (QUANTIFIED)? (DEPENDENT VARIABLES)	POTENTIAL VARIABLES THAT CONTRIBUTE TO THE OCCURRENCE OF THE PHENOMENON (INDEPENDENT VARIABLES)
Osmosis occurs at a faster rate in the presence of a higher concentration gradient.	Weight of dialysis bag at different time points; Colorimetric verification of diffusion if applicable	Magnitude of concentration gradient Osmotic pressure Mass of solutes in solution System volume

What can you test in the lab and what manipulations will you use to test your hypothesis?

TOPIC/PHENOMENON	MODELS/CONDITIONS	POSSIBLE MANIPULATIONS OF THE SYSTEM (INDEPENDENT VARIABLES)
Water molecules in a solution diffuse more quickly in the presence of higher osmotic potentials.	Dialysis tubing bags with solute inside; Potential solute for testing: sucrose	Testing different concentrations of solute inside the dialysis bag; Volume of solutions both inside and outside of the dialysis bag

What will you measure? In other words, what are your possible dependent variables?

TOPIC/PHENOMENON	POSSIBLE MEASUREMENTS (DEPENDENT VARIABLES)	INSTRUMENTS/MATERIALS NEEDED TO DO SO
Water molecules surrounding a dialysis bag that contains a highly concentrated sucrose solution will move more quickly into the bag compared to those that surround a bag with a solution containing a lower concentration.	Weight of the dialysis bag at different time points; Potassium permanganate test for sucrose	Balance, timers, solute indicators ($KMnO_4$)

Are your hypotheses testable and falsifiable? State them here.

HYPOTHESIS	TESTABLE/FALSIFIABLE (YES OR NO?)
If a concentration differential affects the rate of osmosis across a membrane, then water will diffuse into a dialysis bag containing a highly concentrated solution more quickly compared to a solution of a lower concentration.	Yes

Is there one potential experiment from your thought process that seems more adaptable to testing in our lab? Can you utilize proper controls in the experiment to create a strong experimental design?

From your list, what factors can be easily tested in the lab with the materials available?

Materials Available

- Dialysis tubing of differing pore sizes
- Strong salts such as NaCl and $CaCl_2$
- Macromolecules for making solutions: oil, protein, starch, simple sugars
- Macromolecule indicators: Sudan IV, $CuSO_4$, Iodine, $KMnO_4$
- Thermometers
- Hot plates
- Stir plates and stir bars
- Ice
- Glassware
- Balances, weigh boats
- Graduated cylinders

What results do you expect? Do your expectations follow logically from your hypothesis? If not, it may be necessary to refine your hypothesis before you proceed with testing.

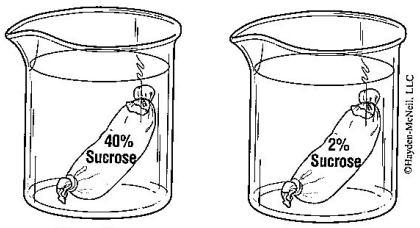
Once you have completed this worksheet, discuss with your TA what your group proposes to investigate. If he/she gives you the go-ahead, then **draft out an experimental procedure**. Make

sure to write down exactly what you plan to do and record any changes that are made when you are conducting the experiment yourself. Every step of the way, you should record your observations and information pertaining to your experiment in your laboratory notebook. This is your record of exactly what you planned to do, what you did do, and what you observed.

Draft out a table in which you plan to record your raw data. Do you have one independent variable with multiple levels? More than one independent variable? What is/are your dependent variable(s)? What units will be used for measurement?

Designing an Experiment

In order to test the statement from the previous example, the experimenter would need to understand that concentrations of sucrose must be set up on opposite sides of a semi-permeable membrane and that osmosis needs to be measured over a period of time. We know that we can use dialysis tubing as a model semi-permeable membrane. Sucrose solutions at different concentrations can be made to fill the dialysis bag and/or a beaker to compose the solution surrounding the dialysis bag. The experiment might be set up like the figure below.



"Large" concentration gradient

"Small" concentration gradient

Our hypothesis requires that osmosis be measured over a period of time. We know that osmosis is the movement of water through a semi-permeable membrane. We also know that water has weight. Weight can be easily measured on a balance. A measurable change in weight of the dialysis bag over a period of time might be a good method to measure rate of osmosis. Alternatively, volume changes over time could be measured using a graduated cylinder. Changes in solution volume inside or outside the dialysis bag would indicate a water movement has occurred. It is the job of the experimenter to design their experiment so that the dependent variable being measured is an accurate representation of the phenomena they are trying to characterize.

After establishing how the dependent variable will be measured, you need to attempt to establish positive and/or negative controls. In the case of our sucrose concentration gradient example, a proper negative control should be a beaker containing a dialysis tubing bag where both the beaker and bag contain a solution of same sucrose concentration. This system would provide no solute concentration gradient and thus no osmotic pressure.

How quickly will water move across the dialysis tubing? Will sucrose diffuse through the dialysis tubing? How can we test this? If so, how might that affect the movement of water? How would this affect the results? What controls need to be set up to demonstrate that molecules are or are not moving in a net direction? What substance concentrations are appropriate? A good scientist considers all the possibilities when designing an experiment.

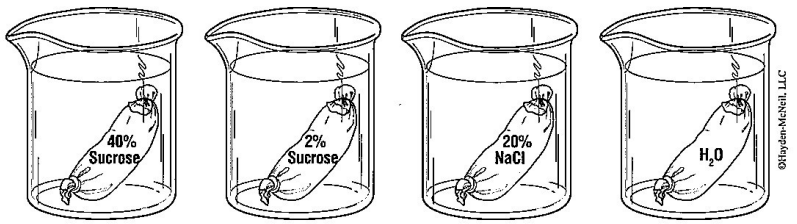
IV. AN EXAMPLE EXPERIMENTAL PROCEDURE

Every group will require a procedure that is specific to their experimental design.

As an example, let us set up an experiment to test our hypothesis regarding the effect of sucrose concentration on osmosis. To set up the experiment, equal volumes of 10%, and 2% sucrose solutions would be used to fill equal sized dialysis bags. Deionized water, or 0% sucrose, would be used as a negative control because there should be no net water movement because there is no solute concentration gradient. A high concentration of NaCl inside a dialysis bag could serve as a useful positive control because the large solute concentration gradient should allow a large movement of water from the beaker to the dialysis bag to be measured.

Setting Up a Dialysis Bag

1. Cut dialysis tubing to the desired size and soak in water for 5–10 minutes. A standard length is 6 cm.
2. Close one end of the tubing using a plastic clamp. A bag can be formed by rubbing the other end between your fingers until the edges separate. *Be sure to wear gloves so as not to introduce any unwanted contamination to your experiment.*
3. Use a transfer pipette to insert your desired solution into the dialysis bag. Seal the end with another dialysis clamp. A standard volume is 1–3 mL.
4. When placing your dialysis bag into a beaker, ensure the beaker solution volume chosen allows the bag to be completely submerged.



The weight of the dialysis bag could be measured over an incubation period to determine if water was moving from the beaker into bag or not. To measure a change in weight, a baseline

measurement must be recorded first. If we decide on a 30-minute incubation time with measurement intervals every 10 minutes, the data could be tabulated as shown below. Remember, this experiment is only an example—your experiment should be as unique as you see fit to measure your hypothesis.

Table 2b.1: Weight of dialysis bags filled with water, NaCl, or sucrose after incubation in water over time.

	WEIGHT OF DIALYSIS BAG (G)			
INCUBATION TIME (MINUTES)	H_2O (–)	20% NaCl (+)	10% SUCROSE	2% SUCROSE
0				
10				
20				
30				

This table provides an effective way to record raw data measurements during the experiment. But a more sophisticated analysis is expected as results for a written lab report assignment. This might be a data table that reports the change in weight of the bag over time or a line graph plotting the weight of the bags over time. You, as the experimenter, have the responsibility to illustrate the data objectively and as clearly as possible for the reader.

You might speculate that sucrose will diffuse out of the bag and into the beaker. What kind of effect might this have on the results? Is there a way to test for the presence of sucrose in the beaker solution?

Before you come to the lab next week you should have:

a. an experimental question

b. a testable hypothesis

c. a step-by-step protocol

d. operational definitions of all variables and controls

e. definitive concentrations of any applicable solutions and calculations that relate to how these solutions will be made

f. tables prepared in which to record data and observations

g. a preliminary list of references used to obtain information relevant to decisions made on the experimental design

Disposal Methods

Dialysis tubing can be discarded in the regular trash.

If you used $KMnO_4$ at any point, you should discard that in the appropriate container under the fume hood. All other test solutions can be rinsed down the drain.

Remove any tape or labeling off of the test tubes before rinsing and placing the tubes in the provided receptacle.

Rinse all beakers and other glassware you may have used.

All gloves should be discarded in the biohazard bin.

Don't forget to wipe down your bench space with 70% EtOH.

Wash your hands before exiting the lab!

QUESTIONS: ANSWER IN YOUR LAB NOTEBOOK

1. If resources were not a limitation, are there other ways of testing your hypothesis?
2. What is the difference, if any, between the terms osmosis and diffusion?
3. If you were to repeat your experiment, how would you alter your design to strengthen your conclusions?

Appendix G: Potent HIV-1 Protease inhibitors Containing Carboxylic and Boronic Acids: Effect on Enzyme Inhibition and Antiviral Activity and protein-Ligand X-ray structural studies

ChemPubSoc Europe
DOI: 10.1002/cmdc.201900508
CHEMMEDCHEM Full Papers
Check for updates

Potent HIV-1 Protease Inhibitors Containing Carboxylic and Boronic Acids: Effect on Enzyme Inhibition and Antiviral Activity and Protein-Ligand X-ray Structural Studies

Arun K. Ghosh,*[a] Zilei Xia,[a] Satish Kovela,[a] William L. Robinson,[a] Megan E. Johnson,[a] Daniel W. Kneller,[b] Yuan-Fang Wang,[b] Manabu Aoki,[d, e] Yuki Takamatsu,[e] Irene T. Weber,[b] and Hiroaki Mitsuya[c, d, e]

We report the synthesis and biological evaluation of phenylcarboxylic acid and phenylboronic acid containing HIV-1 protease inhibitors and their functional effect on enzyme inhibition and antiviral activity in MT-2 cell lines. Inhibitors bearing *bis*-THF ligand as P2 ligand and phenylcarboxylic acids and carboxamide as the P2′ ligands, showed very potent HIV-1 protease inhibitory activity. However, carboxylic acid containing inhibitors showed very poor antiviral activity relative to carboxamide-derived inhibitors which showed good antiviral IC_{50} value. Boronic acid derived inhibitor with *bis*-THF as the P2 ligand showed very potent enzyme inhibitory activity, but it showed lower antiviral activity than darunavir in the same assay. Boronic acid containing inhibitor with a P2-*Crn*-THF ligand also showed potent enzyme K_i but significantly decreased antiviral activity. We have evaluated antiviral activity against a panel of highly drug-resistant HIV-1 variants. One of the inhibitors maintained good antiviral activity against $HIV_{DRV^R_{P20}}$ and $HIV_{DRV^R_{P30}}$ viruses. We have determined high resolution X-ray structures of two synthetic inhibitors bound to HIV-1 protease and obtained molecular insight into the ligand-binding site interactions.

Introduction

Structure-based molecular design has become one of the most successful and widely used strategies for preclinical drug development. Over the years, it has led to numerous first-in-class approved drugs and preclinical candidates.[1,2] Structure-based design strategies have played a key role in the design and development of HIV-1 protease inhibitor drugs.[3,4] HIV-1 protease inhibitors are an important component of current antiretroviral therapy (ART) which is responsible for dramatic improvement of life expectancy and mortality rates of HIV/AIDS patients in developing nations. HIV-1 protease is an aspartic acid protease which plays a critical role in viral replication.[5,6] HIV-1 protease inhibitor (PI) drugs block HIV-1 protease and generate morphologically immature and noninfectious virions.[7,8] Darunavir (**1**, Figure 1), the latest FDA-approved PI, has been highly effica-

[a] Prof. Dr. A. K. Ghosh, Dr. Z. Xia, Dr. S. Kovela, W. L. Robinson, M. E. Johnson
Department of Chemistry and Department of Medicinal Chemistry, Purdue University, West Lafayette, IN 47907 (USA)
E-mail: akghosh@purdue.edu
[b] D. W. Kneller, Y.-F. Wang, Prof. Dr. I. T. Weber
Departments of Biology and Chemistry, Molecular Basis of Disease, Georgia State University, Atlanta, GA 30303 (USA)
[c] Dr. H. Mitsuya
Departments of Hematology and Infectious Diseases, Kumamoto University School of Medicine, Kumamoto 860-8556 (Japan)
[d] Dr. M. Aoki, Dr. H. Mitsuya
Experimental Retrovirology Section, HIV and AIDS Malignancy Branch, National Cancer Institute, Bethesda, MD 20892 (USA)
[e] Dr. M. Aoki, Y. Takamatsu, Dr. H. Mitsuya
Department of Refractory Viral Infections, National Center for Global Heath and Medicine Research Institute, Shinjuku, Tokyo 162-8655 (Japan)
Supporting information and the ORCID identification number(s) for the author(s) of this article can be found under: https://doi.org/10.1002/cmdc.201900508.

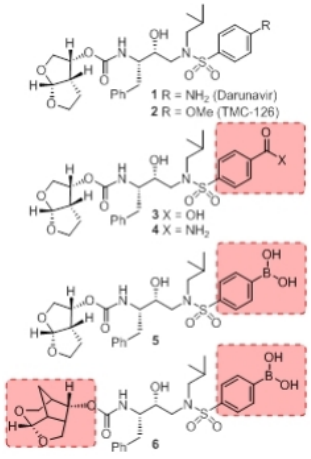


Figure 1. Structures of HIV-1 protease inhibitors **1**–**6**.

ChemPubSoc Europe

CHEMMEDCHEM Full Papers

cious in suppressing HIV-1 replication and showing significant clinical benefits to HIV/AIDS patients.[9,10] Darunavir has emerged as a widely used first-line therapy for rescue treatment.[11,12] Its exceptional resistance profile is due to a dual mechanism of action as it inhibits HIV-1 protease and inhibits dimerization of HIV-1 protease monomers.[13]

Darunavir was developed by structure-based design strategies with particular emphasis on maximizing active site interactions, especially promoting strong hydrogen bonding interactions with the HIV-1 backbone atoms in the S2 and S2′ subsites.[14,15] In this context, we created a stereochemically defined 3(*R*),6a(*R*)-*bis*-tetrahydrofuranyl urethane as the P2 ligand to form hydrogen bonds with the Asp29 and Asp30 backbone NHs in the S2-subsite. Also, the P2′-4-aminobenzene sulfonamide was incorporated to interact with the backbone atoms in the S2′ subsite.[16,17] These inhibitor design strategies targeting the protein backbone may be responsible for darunavir's exceptional antiviral activity against highly multidrug-resistant HIV-1 variants.[18,19] However, DRV-resistant HIV-1 variants have emerged, and development of novel antiviral drugs with a high genetic barrier against multidrug-resistant HIV-1 variants including DRV-resistant variants is urgently needed.[20,21] In this context, we investigated further optimization of structural elements of darunavir that would enhance backbone interactions in the active site of HIV-1 protease.

The X-ray structure of darunavir-bound HIV-1 protease (PDB ID: 4HLA) revealed that the P2′ 4-aminosulfonamide ligand is involved in water-mediated hydrogen bonding interactions with Gly48′ backbone amide NH through two crystallographic water molecules.[17,22] Based on this observation, we speculated that the modification of amine functionality with a carboxylic acid or with a carboxamide group would replace one of these two crystallographic water molecules closer to the amine functionality. Such modified ligands could not only maintain strong hydrogen bonds with backbone NH of Asp30, but also the carboxylic acid derivative can form a water-mediated hydrogen bond with the Gly48′ NH located in flap region of HIV-1 protease. Indeed, the resulting modified inhibitors **3** and **4** exhibited very potent enzyme inhibitory activity with K_i values of 12.9 pM and 8.9 pM, respectively.[22] Interestingly, carboxylic acid derived inhibitor **3** did not exhibit any appreciable antiviral activity in MT-2 cells. However, inhibitor **4** with a carboxamide derivative showed antiviral IC_{50} value of 93 nM in MT-2 cells.[22] This result is not surprising as it has been reported that intracellular concentration of PIs are a dynamic balance between influx, efflux, and sequestration.[23,24] Because DRV has been shown to serve as a transport substrate of P-glycoprotein, the carboxylic acid derivatives are expected to show similar efflux profiles.[25] Thus, we have assayed intracellular concentrations of inhibitor **3** and **4** in MT-2 and MT-4 cells. Interestingly, both cell lines showed very low uptake of carboxylic acid derived inhibitor **3**.[22] The sequestration and influx properties of PIs are important issues that need serious attention for PI design. We have determined X-ray crystal structures of inhibitor **3** and **4** with HIV-1 protease complexes. Our structural analysis revealed enhanced hydrogen bonding interactions in the S2′ subsite relative to DRV–HIV-1 protease complex.[22] A stereoview of the active site interactions of inhibitor-bound HIV-1 protease is shown in Figure 2. As can be seen, the P2′ 4-carboxylic acid ligand is involved in water-mediated hydrogen bonding interactions with Gly48′ backbone amide NH located in the flap region of HIV-1 protease.[17,22] More recently, Raines and co-workers reported the HIV-1 protease inhibitor **5** with a phenyl boronic acid as the P2′ ligand.[26] This inhibitor also exhibited very potent enzyme inhibitory activity relative to darunavir. Furthermore, X-ray crystal structure of inhibitor **5**-bound HIV-1 protease revealed that boronic acid is involved in enhanced hydrogen bonding interaction including a water-mediated hydrogen bonding interaction with Gly48′ similar to inhibitor **3** and **4**.[26] While boronic acid containing inhibitor **5** displayed potent enzyme inhibitory activity, assessment of its antiviral activity in cell-based assays was not carried out. Thus, the intriguing question is whether boronic acid derived inhibitors show better antiviral potency than carboxylic acid derived compounds. In the present studies, we have synthesized a selected group of HIV-1 protease inhibitors containing boronic acid as the P2′ ligand and evaluated their enzyme inhibitory activity as well as antiviral activity compared to darunavir. Fur-

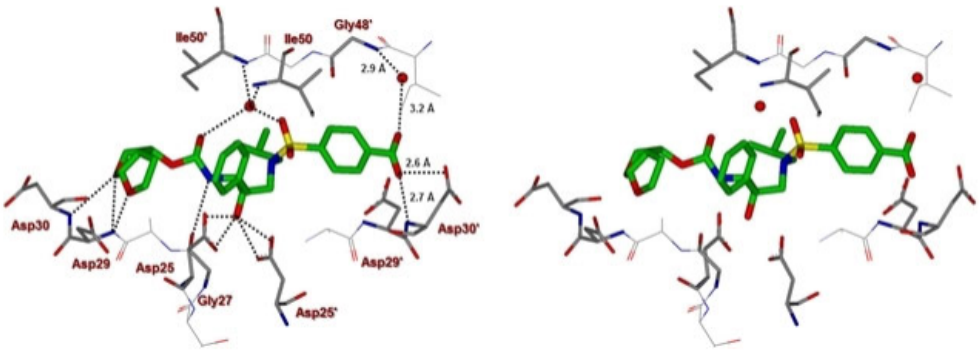


Figure 2. A stereoview of inhibitor **3**-bound X-ray structure of HIV-1 protease. The major orientation of the inhibitor is shown. The inhibitor carbon atoms are shown in green, water molecules are red spheres, and the hydrogen bonds are indicated by dotted lines (PDB ID: 4I8W).

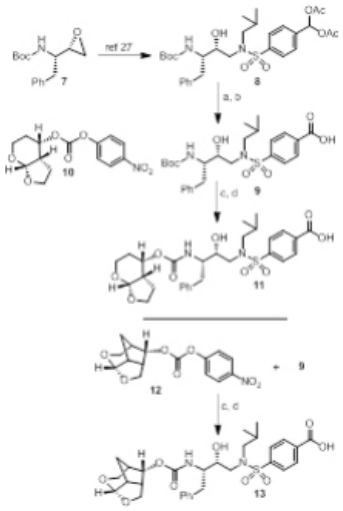
ChemPubSoc Europe

CHEMMEDCHEM Full Papers

thermore, we have evaluated antiviral activity against a panel of highly darunavir-resistant HIV-1 variants for compounds **5** and **6** with darunavir control. To assess ligand-binding site interactions, we have determined a high resolution X-ray structure of inhibitor **6**-bound HIV-1 protease.

Synthesis of carboxylic and boronic acid based inhibitors

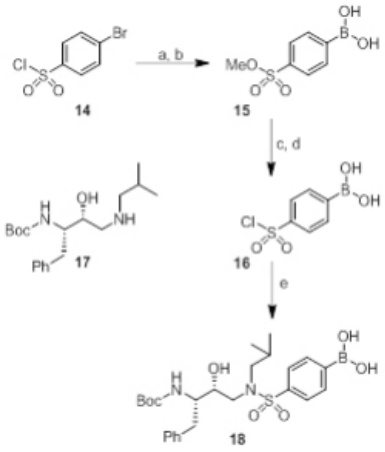
For our investigation regarding the effect on enzyme inhibitory and antiviral activity for inhibitors containing 4-phenylcarboxylic acid and 4-phenylboronic acid as the P2′ ligand, we have synthesized a set of selected inhibitors. The synthesis of 4-phenylcarboxylic acid derivatives **11** and **13** is shown in Scheme 1.



Scheme 1. Synthesis of PIs **11**, and **13**: (a) K_2CO_3, MeOH, 23 °C, 30 min; (b) $NaClO_2$, 2-methyl-2-butene, NaH_2PO_4, *t*BuOH-H_2O, 23 °C, 3 h (86% for two steps); (c) TFA, CH_2Cl_2, 0 °C to 23 °C, 3 h; (d) **10**, DIPEA, CH_3CN, 23 °C, 12 days (37–80% for two steps).

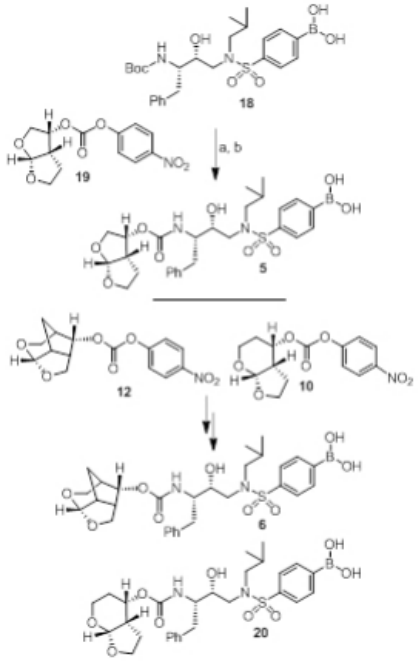
Commercially available (*S*,*S*)-3(*t*-Boc-amino)-1,2-epoxy-4-phenylbutane **7** was converted into hydroxyethylamine sulfonamide derivative **8** with a diacetoxymethyl group as P2′ ligand as described by us previously.[27] Exposure of the acetate derivative **8** to K_2CO_3 in methanol at 23 °C for 30 min provided the corresponding aldehyde. Pinnick oxidation[28] of the resulting aldehyde with sodium chlorite in *t*-butanol in the presence of 2-methyl-2-butene and phosphate buffer afforded the carboxylic acid derivative **9** in 86% yield over two steps. For the conversion of carboxylic acid derivative **9** to inhibitor **11** Boc-group was removed by exposure to trifluoroacetic acid (TFA) in CH_2Cl_2 at 0 °C to 23 °C for 3 h. Reaction of the resulting amine with known[29] activated carbonate of tetrahydropyranofuran **10** in the presence of diisopropylethylamine (DIPEA) in CH_3CN at 23 °C for 12 days afforded inhibitor **11** in 37% yield over two steps. For the synthesis of carboxylic acid derivative **13**, Boc derivative **9** was treated with TFA and the resulting amine was reacted with *crown*-THF carbonate **12**[30] as described above to provide **13** in 80% yield over two steps.

For convenient synthesis of inhibitors containing 4-phenylboronic acid, we first prepared 4-(chlorosulfonyl)phenyl boronic acid. As shown in Scheme 2, commercially available 4-bromophenyl sulfonyl chloride **14** was treated with NaOMe in methanol at 0 °C to 23 °C for 12 h to provide the corresponding methyl sulfonate derivative in 62% yield.[31] Reaction of the resulting 4-bromosulfonate derivative with *n*BuLi in THF at −78 °C for 10 min followed by addition of $B(OMe)_3$ at −78 °C for 30 min provided boronic acid derivative **15** in 45% yield. Treatment of **15** with triethylamine in methanol at 60 °C for 24 h followed by reaction of the resulting triethylammonium salt with thionyl chloride, heated at 80 °C for 24 h, furnished sulfonyl chloride **16** in 89% yield.[32] This sulfonyl chloride was then reacted with the known hydroxyethylamine derivative **17**[27,33] in CH_2Cl_2 in the presence of triethylamine at 0 °C to 23 °C for 1 h to furnish hydroxyethylsulfonamide dipeptide isostere **18** with a 4-phenylboronic acid as the P2′ ligand in 82% yield.



Scheme 2. Synthesis of boronic acid derivative **18**: (a) NaOMe, MeOH, 0 °C to 23 °C, 12 h (62%); (b) *n*BuLi, THF, −78 °C, $B(OMe)_3$, 30 min (45%); (c) Et_3N, MeOH, 60 °C for 24 h; (d) $SOCl_2$, 80 °C for 24 h (89% for two steps); (e) Et_3N, CH_2Cl_2, 0 °C to 23 °C, 1 h (82%).

Synthesis of inhibitors with boronic acid as the P2′ ligand is shown in Scheme 3. Boc-derivative **18** was treated with TFA in CH_2Cl_2 at 0 °C for 30 min to provide the corresponding amine. Reaction of the resulting amine with known activated *bis*-THF carbonate derivative **19**[29] in the presence of DIPEA in CH_3CN at 23 °C for 10 days furnished inhibitor **5** in 47% yield. Similarly, reaction of resulting amine with activated crown-THF carbonate **12** and activated Tp-THF carbonate **10** afforded inhibitors **6** and **20** in 55% and 51% yields, respectively.

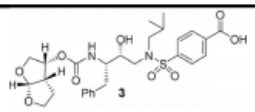
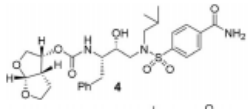
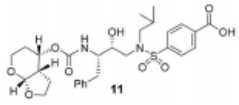
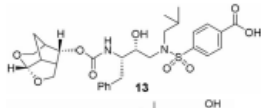
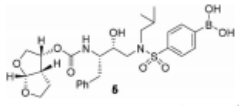
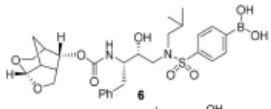
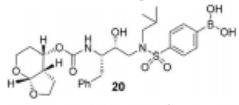
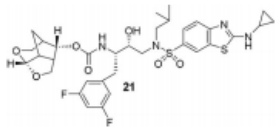


Scheme 3. Synthesis of PIs: (a) TFA, CH_2Cl_2, 0 °C, 30 min; (b) carbonate **19**, DIPEA, CH_3CN, 23 °C, 10 days, (47% yield for two steps).

Results and Discussion

Our structure based design strategies resulted in a range of potent PIs with intriguing structural features, particularly PIs with cyclic ether-derived P2 ligands.[3,14] Through X-ray structural analysis, we have shown that these PIs make extensive interactions particularly with backbone atoms in the active site. In general, these PIs exhibited exceptional enzyme affinity.[14] Even though various factors such as cell penetration, flux, and intracellular sequestration may affect antiviral activity, many of these PIs also showed very potent antiviral activity.[3] Furthermore, they maintained robust antiviral activity against a wide range of highly multidrug-resistant HIV-1 variants. For the current studies, we designed a selected set of PIs containing carboxylic acid and boronic acid functionalities in the P2′-ligand. These functional groups can form more hydrogen bonds in the S2′-subsite than the aromatic amine functionality of darunavir.[9] The structure and activity of these PIs are shown in Table 1. We first evaluated these PIs in an enzyme-inhibitory assay.[34] As shown, inhibitor **3** with a P2-carboxylic acid showed a K_i value of 12.9 pM. We subsequently determined antiviral activity of these PIs in MT-2 and MT-4 human-T-lymphoid cells exposed to HIV_{LAI}.[35] Interestingly, in this cell-based antiviral assay, inhibitor **3** did not exhibit appreciable antiviral activity with IC_{50} values of 1 μM and 1 mM, in MT-2 and MT-4 cells,

Table 1. Structures and activity of carboxylic and boronic acid containing inhibitors.

Entry	Inhibitor	K_i [pM][a]	IC_{50} [nM][c]
1		12.9	> 1000
2		8.9	93
3		7.6	> 1000
4		3.2	> 1000
5		0.5[b]	48.9
6		15.5	37.7
7		2.1	120
8		14	0.017

[a] K_i values represent at least five data points. Standard error in all cases was less than 7%. Darunavir (DRV) exhibited K_i = 16 pM. [b] Reported value. [c] Values are means of at least three experiments in MT-2 cells. Standard error in all cases was less than 5%. DRV exhibited IC_{50} = 1.6 nM.

respectively. Carboxamide derivative **4** showed very potent enzyme K_i and antiviral IC_{50} values of 93 nM and 38 nM in MT-2 and MT-4 cells, respectively. Both carboxylic acid containing inhibitors **11** and **13** showed excellent enzyme inhibitory activity (K_i 7.6 and 3.2 pM, respectively) however, they did not show antiviral activity. Boronic acid derivative **5** has been shown to have very potent enzyme inhibitory activity of 0.5 pM.[26] However, there is no report of its antiviral activity in a cell-based assay. We have evaluated the boronic acid PI **5** in an antiviral assay with MT-2 cells. It turned out that boronic acid derivative **5** showed improved antiviral IC_{50} value of 48.9 nM compared to carboxylic derivatives **3** and **4** (entry 2, Table 1). We prepared PI **6** containing a crown-like THF ligand as P2 ligand in

ChemPubSoc Europe

CHEMMEDCHEM Full Papers

combination with boronic acid P2′ ligand. This inhibitor exhibited an enzyme inhibitory K_i value of 15.5 pM and antiviral activity of 37.7 nM (entry 6). Inhibitor **20** with a (3*aS*,4*S*,7*aR*)-hexahydro-4*H*-furo[2,3-*b*]pyran as the P2 ligand also showed very potent enzyme K_i but much reduced antiviral IC_{50} value of (120 nM) compared to DRV (entry 7).

Clinically, DRV is used widely for the treatment of naïve and experienced HIV/AIDS patients. However, heavily ART-regimen-experienced patients have been showing treatment failure with approved PIs, including DRV.[21,35] Therefore, we have examined inhibitors **5** and **6** against DRV-resistant HIV-1 variants and compared their antiviral activity against DRV and one of our preclinical inhibitors, **21**. For these studies, MT-4 cells (1×10^4) are exposed to wild type HIV-1 and three highly DRV-resistant HIV-1 variants, $HIV_{DRV^R_{P20}}$, $HIV_{DRV^R_{P30}}$, and $HIV_{DRV^R_{PS1}}$.[35,36] These variants were selected by propagating cells with increasing concentrations of DRV. They are highly resistant to all currently approved PIs, including DRV and nucleoside-RT inhibitors, such as tenofovir. Antiviral activity of selected PIs was determined using p24 assays. The results are shown in Table 2. As can be seen, FDA approved PI, LPV was unable to block the replication of these highly DRV-resistant variants. Inhibitor **5**, resulting from incorporation of boronic acid in place of amine in DRV, showed relatively more potent antiviral activity (IC_{50} value of 219 nM) against DRV-resistant HIV-1 variants selected in the presence of DRV over 20 viral passages ($HIV_{DRV^R_{P20}}$). Compound **5** lost only 4-fold activity relative to its IC_{50} values against $HIV\text{-}1^{WT}$. Interestingly, DRV exhibited 164-fold loss of antiviral activity (IC_{50} 525.8 nM) compared with its IC_{50} against $HIV\text{-}1^{WT}$. Boronic acid derivative **5** displayed antiviral IC_{50} of 946 nM (19-fold loss over $HIV\text{-}1^{WT}$) compared with DRV (IC_{50} 601 nM, 187-fold over $HIV\text{-}1^{WT}$) against $HIV_{DRV^R_{P30}}$ variant. Both compound **5** and DRV did not exhibit any appreciable antiviral activity against the most highly DRV-resistant HIV-1 variants, $HIV_{DRV^R_{PS1}}$. Inhibitor **6** with a crown-like THF (*Crn*-THF) as the P2-ligand displayed improved antiviral activity relative to inhibitor **5** with a *bis*-THF P2 ligand. It showed antiviral IC_{50} value of 71 nM against $HIV_{DRV^R_{P20}}$ variants and IC_{50} value of 532 nM against $HIV_{DRV^R_{P30}}$ variants. Its fold-changes of antiviral activity relative to $HIV\text{-}1^{WT}$ are significantly lower than compound **5** or DRV. Compound **6**, however, exhibited antiviral $IC_{50} > 1$ μM (> 1000-fold change) against highly resistant HIV variants $HIV_{DRV^R_{PS1}}$. One of our recent inhibitors, **21**, showed exceptional antiviral activity against all highly DRV-resistant HIV-1 variants.[37,38] Both inhibitors **5** and **6** showed no appreciable cytotoxicity in MT-2 cells, with CC_{50} values of > 100 μM, similar to DRV. However, selectivity index (CC_{50}/IC_{50}) were > 2040 and > 2650 compared to 34480 for DRV and 2473684 for compound **21**.

X-ray description

To gain molecular insight into the ligand-binding site interactions of inhibitors **6** (GRL-008–19A) and **20** (GRL-031-19A), we co-crystallized these inhibitors with wild-type HIV-1 protease and the X-ray structures were refined at 1.33 and 1.13 Å resolution with R-factors of 13.7% and 12.5%, respectively. These structures contain a single PR dimer and the inhibitor binds in the active site cavity in two mutually exclusive orientations related by 180° rotation with major/minor occupancies of 60/40% and 65/35% for inhibitors **6** and **20**, respectively. The overall dimer structures are very similar to the HIV-1 protease with DRV (2IEN)[16] with root mean square difference (RMSD) for Cα of 0.20 and 0.12 Å for inhibitors **6** and **20**, respectively. Similarly, the two new structures of inhibitors **6** and **20** are alike (Cα RMSD of 0.14 Å). The urethane NH of both inhibitors forms hydrogen bonding interactions with the carbonyl oxygen of Gly27. In addition, these inhibitors show water-mediated interactions that connect the inhibitor carbonyl oxygen and sulfonamide oxygen with the amides of Ile50 and 50′ in the flaps. Similar interactions are shared by the protease-darunavir complex and other protease inhibitors.[4,16] The key interactions of inhibitor **6** and inhibitor **20** with HIV-1 protease are highlighted in Figure 3.

The hydrogen bond interactions between HIV-1 protease and the two inhibitors were compared with those for DRV. Inhibitors **6** and **20** differ only at the P2-group where the former boasts a *Crn*-THF while the latter contains a tetrahydropyranotetrahydrofuran (*Tp*-THF) ligand.[29,30] Both oxygen atoms of the *Crn*-THF in inhibitor **6** formed a network of hydrogen bonds with the backbone amide NHs of Asp30 and Asp29 in the S2 subsite. These interactions are very similar to what was observed in the X-ray structure of the *Crn*-THF containing inhibitor **21**, described recently.[37,38] For inhibitor **20**, both oxygens of the *Tp*-THF ligand also formed hydrogen bonds with the backbone amide NHs of Asp30 and Asp29. These interactions are comparable to the equivalent interactions with the *bis*-THF of DRV.[16,17]

The P2′ position of both inhibitors **6** and **20** harbors a phenylboronic acid moiety in place of the 4-aminosulfonamide of

Table 2. Antiviral activity of novel compounds against highly DRV-resistant HIV-1 variants.[a]

	IC_{50} [nM] (fold-change)				
	LPV	**1** (DRV)	**5**	**6**	**21** (GRL-14213)
$cHIV_{LAI}^{WT}$	13	3.2	48.9	37.7	0.017
$HIV_{DRV^R_{P20}}$	> 1000 (> 77)	525.8 (164)	219.3 (4)	71.2 (2)	0.0024
$HIV_{DRV^R_{P30}}$	> 1000 (> 77)	601 (187)	946.4 (19)	532.3 (14)	0.14
$HIV_{DRV^R_{PS1}}$	> 1000 (> 77)	5429.7 (1696)	> 1000	> 1000	1.3

[a] Numbers in parentheses represent fold changes in IC_{50} values for each isolate relative to the IC_{50} for wild-type $cHIV_{LAI}^{WT}$. All assays were conducted in triplicate, and the data are the mean values (± 1 standard deviation) derived from the results of three independent experiments. DRV, Darunavir; LPV, Lopinavir.

ChemPubSoc Europe

CHEMMEDCHEM Full Papers

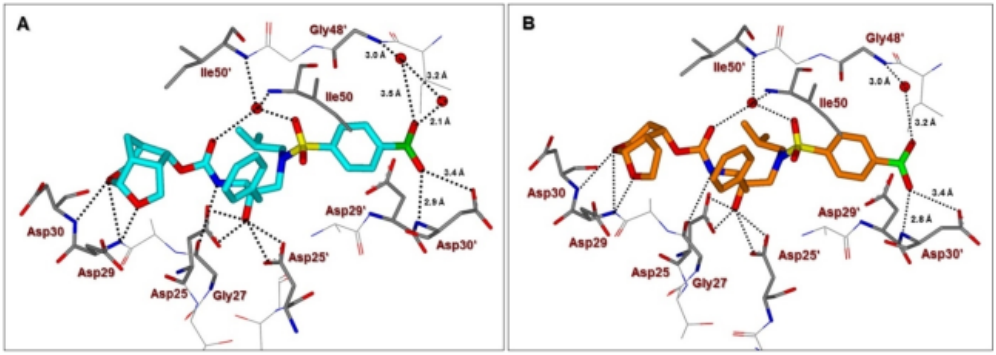


Figure 3. A) The X-ray structure of inhibitor **6**-bound HIV-1 protease (carbon atoms, turquoise; PDB ID: 6U7O). B) The X-ray structure of inhibitor **20**-bound HIV-1 protease (carbon atoms, orange; PDB ID: 6U7P). The major orientations of the inhibitors are shown. All water molecules are red spheres, the hydrogen bonds are shown by dotted lines and bond lengths are shown in the S2′-subsite.

DRV. In both crystal structures, the boronic acid is aligned in plane with the phenyl group with less than 2° difference in torsion angle. The boronic acid introduces an additional water-mediated hydrogen bond with the Gly48′ amide located on the flexible flap of HIV-1 protease. This interaction is similar to the carboxylic acid derivative **3** as shown in Figure 2. The other hydroxy group of the boronic acid is within hydrogen bonding distance to both main chain amide of Asp30′ (2.7–2.9 Å), although the bond angle departs from the ideal in the crystal structures. It appears that inhibitor **6** makes additional water-mediated interactions in the S2′-site of the active site. Overall, the inhibitor-HIV-1 protease hydrogen bonding interactions resemble those in DRV–HIV-1 protease with the exception of the new water-mediated interaction of boronic acid at P2′ and Gly48′, which is likely to stabilize the mobile flaps. These new interactions are likely to account for the improved enzyme affinity for inhibitors **6** and **20**. These inhibitors showed better antiviral activity than the corresponding carboxylic acid derivatives likely due to improved efflux profiles.

Conclusions

In conclusion, we have designed and synthesized a selected set of HIV-1 protease inhibitors containing phenylcarboxylic acid and phenylboronic acid as the P2′-ligands. These ligands are designed to form enhanced hydrogen bonding interactions with the backbone atoms in the S2′-subsite of HIV-1 protease. Inhibitor **3** with P2 *bis*-THF and P2′-phenylcarboxylic acid showed potent enzyme inhibitory K_i, but showed no antiviral activity due to poor cell penetration. However, carboxamide derivative **4** displayed excellent enzyme inhibitory activity and good antiviral activity (IC_{50} of 93 nM). The corresponding boronic acid derivative **5** showed excellent enzyme K_i and improvement of antiviral IC_{50} value (49 nM) over the carboxamide derivative. Carboxylic acid and boronic acid-containing inhibitors **6** and **13** in combination with a crown-like THF as the P2 ligand also showed very potent enzyme inhibitory activity. Carboxylic acid derivative **13**, however, showed poor antiviral activity. Boronic acid derived inhibitors displayed potent enzyme activity. They exerted much better antiviral activity in cell-based assay than the corresponding carboxylic acid derivatives possibly due to higher efflux profiles. However, DRV with a P2′-aniline ligand displayed significantly more potent antiviral activity than boronic acids. We have obtained high resolution X-ray structures of inhibitor **6**-HIV-1 protease and inhibitor **20**-HIV-1 protease complexes. Our structural studies established that both carboxylic acid and boronic acid-containing inhibitors are involved in significantly enhanced hydrogen bonding interactions and water-mediated hydrogen bonding interactions in the S2′-site. Inhibitor **5**, while less potent than DRV against HIV_{LAI}^{WT}, exerted similar antiviral activity against highly DRV-resistant HIV-1 variants, $HIV_{DRV\ P20}^{R}$ and HIV_{DRV}^{P30}. Inhibitor **6**, with a *Crn*-THF ligand as the P2-ligand, displayed improved antiviral activity compared to *bis*-THF derived inhibitor **5**. Interestingly, fold-change of antiviral activity for boronic acid derived inhibitors are significantly lower than DRV against these highly DRV-resistant HIV-1 variants. Only inhibitor **21** showed potent antiviral activity against $HIV_{DRV\ PS1}^{R}$. Further design of new PIs using the current molecular insight is in progress.

Experimental Section

General: All reactions were carried out under an argon atmosphere in either flame or oven-dried (120 °C) glassware. All reagents and chemicals were purchased from commercial suppliers and used without further purification unless otherwise noted. Anhydrous solvents were obtained as follows: Dichloromethane from calcium hydride, diethyl ether and tetrahydrofuran from Na/Benzophenone, methanol and ethanol from activated magnesium under argon. All purification procedures were carried out with reagent grade solvents (purchased form VWR) in air. TLC analysis was conducted using glass-backed Thin-Layer Silica Gel Chromatography Plates (60 Å, 250 μm thickness, F-254 indicator). Column chromatography was performed using 230–400 mesh, 60 Å pore diameter silica gel. ^1H, ^{13}C NMR spectra were recorded at room temperature on a Bruker AV800, DRX-500, ARX-400. Chemical shifts (δ values) are reported in parts per million, and are referenced to the deuter-

ChemPubSoc Europe

CHEMMEDCHEM Full Papers

ated residual solvent peak. NMR data is reported as: δ value (chemical shift, J-value (Hz), integration, where s = singlet, d = doublet, t = triplet, q = quartet, brs = broad singlet). LRMS and HRMS spectra were recorded at the Purdue University Department of Chemistry Mass Spectrometry Center.

4-(*N*-((2*R*,3*S*)-3-((*tert*-Butoxycarbonyl)amino)-2-hydroxy-4-phenylbutyl)-*N*-isobutylsulfamoyl)benzoic acid (9): To a stirred solution of acetal derivative **8** (100 mg, 0.165 mmol) in methanol (6 mL) at 23 °C, potassium carbonate (34 mg, 0.247 mmol) was added. The resulting mixture was stirred at 23 °C for 30 min. The solvent was evaporated under reduced pressure, and the product was extracted with ethyl acetate, dried over anhydrous sodium sulfate and concentrated. The crude product was purified by flash chromatography to provide the corresponding aldehyde (80 mg, 96%).

To a stirred solution of above aldehyde (80 mg, 0.158 mmol) at 23 °C in tBuOH (3.0 mL), were added NaH_2PO_4 (65 mg, 0.476 mmol), 2-methyl-2-butene (170 µL, 1.58 mmol), $NaClO_2$ (43 mg, 0.476 mmol) and H_2O (1.0 mL). The resulting reaction mixture was allowed stir for 3 h at 23 °C. Then the reaction mixture was extracted with EtOAc and washed with H_2O, brine solution, dried over Na_2SO_4, filtered, and concentrated under reduced pressure to yield crude residue which was purified by column chromatography over silica gel (5% MeOH/CH_2Cl_2) to afford acid inhibitor **9** (75 mg, 90%) as an amorphous solid. ^{1}H NMR (400 MHz, CD_3OD) δ = 8.17 (d, J = 8.5 Hz, 2H), 7.92 (d, J = 8.5 Hz, 2H), 7.27–7.18 (m, 4H), 7.15 (m, 1H), 3.71 (ddd, J = 9.3, 7.3, 2.5 Hz, 1H), 3.58 (ddd, J = 11.0, 7.3, 3.8 Hz, 1H), 3.43 (dd, J = 15.1, 2.6 Hz, 1H), 3.18–3.02 (m, 3H), 2.95 (dd, J = 13.7, 6.8 Hz, 1H), 2.54 (dd, J = 13.8, 10.6 Hz, 1H), 2.08–1.97 (m, 1H), 1.28 (s, 9H), 0.91 (d, J = 6.6 Hz, 3H), 0.87 ppm (d, J = 6.6 Hz, 3H); ^{13}C NMR (100 MHz, CD_3OD) δ = 166.7, 156.5, 143.3, 138.7, 134.2, 129.9, 128.9, 127.7, 127.1, 125.6, 78.5, 72.4, 56.7, 55.4, 52.0, 35.8, 27.2, 26.3, 18.9 ppm. LRMS-ESI (m/z): 521.2 $[M+H]^+$.

4-(*N*-((2*R*,3*S*)-3-(((((3a*S*,4*S*,7a*R*)-Hexahydro-4*H*-furo[2,3-*b*]pyran-4-yl)oxy)carbonyl)amino)-2-hydroxy-4-phenylbutyl)-*N*-isobutylsulfamoyl)benzoic acid (11): To a stirred solution of **9** (21 mg, 0.04 mmol) in CH_2Cl_2 (1.5 mL) was added TFA (0.5 mL) at 0 °C under argon atmosphere and the mixture was stirred at 23 °C for 3 h. Solvents were evaporated under reduced pressure, then CH_2Cl_2 (3 mL) was added and evaporated twice to give corresponding amine salt which was used for next step without further purification. The above amine salt (21 mg, 0.04 mmol) was dissolved in acetonitrile (1 mL) and *N*,*N*-diisopropylethylamine (28 µL, 0.16 mmol) was added. The resulting mixture was cooled to 0 °C and mixed carbonate **10** (10 mg, 0.03 mmol) was added to it. The resulting reaction mixture was allowed to warm to 23 °C and stir under argon for 12 days. After this period, the reaction mixture was diluted with a mixture of 10% MeOH in CH_2Cl_2 (10 mL), transferred to a separatory funnel, and washed with 1 M HCl (2 mL). The organic layer was dried over Na_2SO_4, filtered, and concentrated under reduced pressure to afford a crude solid which was purified by silica gel chromatography (5% MeOH/CH_2Cl_2) to yield the desired inhibitor **11** as an amorphous white solid (7.1 mg, 37%). R_f = 0.70 (silica plate, 20% MeOH/CH_2Cl_2). ^{1}H NMR (800 MHz, CD_3OD) δ = 8.17 (d, J = 8.1 Hz, 2H), 7.92 (d, J = 8.2 Hz, 2H), 7.30–7.14 (m, 5H), 6.97 (d, J = 8.6 Hz, 1H), 4.91 (s, 1H), 4.06 (t, J = 9.1 Hz, 1H), 3.96–3.68 (m, 4H), 3.44 (d, J = 15.1 Hz, 1H), 3.38 (t, J = 12.3 Hz, 1H), 3.16 (d, J = 14.0 Hz, 2H), 3.04 (dd, J = 15.4, 8.1 Hz, 1H), 2.97–2.91 (m, 1H), 2.53 (t, J = 12.2 Hz, 1H), 2.31 (s, 1H), 2.03 (s, 1H), 1.85 (p, J = 11.3, 10.5 Hz, 1H), 1.72 (q, J = 12.5 Hz, 1H), 1.65 (d, J = 13.4 Hz, 1H), 1.42 (d, J = 10.1 Hz, 1H), 1.37 (s, 1H), 0.93 (d, J = 5.6 Hz, 3H), 0.89 ppm (d, J = 6.0 Hz, 3H). ^{13}C NMR (200 MHz, CD_3OD) δ = 157.8, 144.4, 140.3, 131.4, 130.5, 129.2, 128.6, 127.1, 102.4, 74.1, 71.0, 69.3, 61.8, 58.4, 57.3, 53.6, 44.8, 37.3, 27.9, 27.4, 23.2, 20.3 ppm. HRMS (ESI) m/z: $[M+Na]^+$ Calcd for $C_{29}H_{38}N_2O_9S$ 613.2190; Found 613.2174.

4-(*N*-((2*R*,3*S*)-3-(((((3*S*,7a*S*,8*S*)-Hexahydro-4*H*-3,5-methanofuro[2,3-*b*]pyran-8-yl)oxy)carbonyl)amino)-2-hydroxy-4-phenylbutyl)-*N*-isobutylsulfamoyl)benzoic acid (13): To a stirred solution of **9** (25 mg, 0.048 mmol) in CH_2Cl_2 (1.5 mL) was added TFA (0.5 mL) at 0 °C under argon atmosphere and the mixture was stirred at 23 °C for 3 h. Solvents were evaporated under reduced pressure, then CH_2Cl_2 (3 mL) was added and evaporated twice to give corresponding amine which was used for next step without further purification. The above amine was dissolved in acetonitrile (1 mL) and then added DIPEA (33 µL, 0.187 mmol) at 0 °C under argon atmosphere. To this reaction mixture, activated *crown*-THF carbonate **12** (12 mg, 0.037 mmol) was added and stirred at 23 °C for 7 days. Volatiles were evaporated under reduced pressure and the crude residue was purified by column chromatography over silica gel to provide inhibitor **13** (23 mg) as an amorphous solid (80% yield over two steps). ^{1}H NMR (400 MHz, CD_3OD) δ = 8.16 (d, J = 8.5 Hz, 2H), 7.91 (d, J = 8.5 Hz, 2H), 7.28–7.20 (m, 5H), 7.16 (m, 1H), 5.33 (d, J = 6.2 Hz, 1H), 4.69 (dd, J = 9.1, 5.6 Hz, 1H), 3.95 (d, J = 11.3 Hz, 1H), 3.78–3.67 (m, 2H), 3.62–3.53 (m, 2H), 3.43 (dd, J = 15.1, 2.5 Hz, 1H), 3.38 (dd, J = 9.3, 5.9 Hz, 1H), 3.16 (ddd, J = 13.7, 5.6, 2.4 Hz, 2H), 3.03 (dd, J = 15.0, 8.4 Hz, 1H), 2.94 (dd, J = 13.6, 6.8 Hz, 1H), 2.68–2.60 (m, 2H), 2.54 (dd, J = 13.9, 10.8 Hz, 1H), 2.32 (q, J = 6.3 Hz, 1H), 2.02 (ddd, J = 12.7, 7.9, 6.3 Hz, 1H), 1.76 (d, J = 12.0 Hz, 1H), 1.49 (dt, J = 12.2, 3.9 Hz, 1H), 1.36 (m, 1H), 0.92 (d, J = 6.5 Hz, 3H), 0.87 ppm (d, J = 6.6 Hz, 3H); ^{13}C NMR (125 MHz, CD_3OD) δ = 166.7, 156.3, 143.0, 138.7, 134.3, 129.9, 128.9, 127.8, 127.1, 125.7, 104.2, 74.6, 72.6, 67.9, 59.5, 56.8, 56.0, 52.0, 44.6, 41.9, 37.3, 35.7, 26.4, 22.7, 18.9 ppm. LRMS-ESI (m/z): 603.2 $[M+H]^+$. HRMS-ESI (m/z): $[M+H]^+$ calcd for $C_{30}H_{39}N_2O_9S$, 603.2371; found 603.2360.

(4-(Methoxysulfonyl)phenyl)boronic acid (15): To a stirred solution of 4-bromophenylsulfonyl chloride (2.55 g, 10 mmol) in methanol (20 mL) was added NaOMe (810 mg, 15 mmol) at 0 °C. The reaction was stirred at 23 °C for 12 h and TLC showed the complete consumption of the starting material. The solvent was removed under vacuum and water was added to the residue. The mixture was extracted with a mixture (15:1) of CH_2Cl_2 and methanol. The combined organic layer was washed with brine, dried over Na_2SO_4 and filtered. The solvent was removed under reduced pressure and the residue was purified by silica gel column chromatography (ethyl acetate/hexane 1:20 to 1:15) to afford the corresponding methyl ester (1.56 g, 62% yield). ^{1}H NMR (400 MHz, CD_3OD) δ = 7.81 (s, 4H), 3.76 ppm (s, 3H). ^{13}C NMR (100 MHz, CD_3OD) δ = 134.3, 132.4, 129.3, 128.5, 56.1 ppm. MS (ESI) calcd for $C_7H_8BrO_3S$ $[M+H]^+$: 250.9, Found: 250.9, 252.9.

To a stirred solution of above ester (2.50 g, 10 mmol) in THF (20 mL) under argon at −78 °C, *n*BuLi (1.6 M in hexane, 6.9 mL, 11 mmol) was added. The reaction was stirred at −78 °C for 10 min and $B(OMe)_3$ (1.68 mL, 15 mmol) was added. The reaction was continued to stir at −78 °C for additional 30 min. The reaction was quenched with aqueous NH_4Cl solution. The mixture was extracted with ethyl acetate/hexane (ethyl acetate/hexane 1:3) and the organic layer was discarded. The aqueous layer was adjusted to pH 6 and the mixture was extracted with a mixture (15:1) of CH_2Cl_2 and methanol. The combined organic layer was washed with brine, dried over Na_2SO_4 and filtrated. The solvent was removed under reduced pressure to obtain the crude boronic acid derivative **15** (970 mg, 45% yield). This was used for the next step without purification. ^{1}H NMR (400 MHz, CD_3OD) δ = 7.91 (br s, 2H), 7.84 (d, J =

ChemPubSoc Europe

CHEMMEDCHEM Full Papers

8.0 Hz, 2H), 3.72 ppm (s, 3H). ^{13}C NMR (100 MHz, CD_3OD) δ = 136.1, 134.1, 126.4, 55.8 ppm. MS (ESI) calcd for $C_7H_{11}BO_6S$ $[M+H_2O]^+$: 234.0, Found: 234.0; calcd for $C_7H_9BNaO_5S$ $[M+Na]^+$: 239.0, Found: 239.0; calcd for $C_7H_{10}BO_5S$ $[M+H]^+$: 217.0, Found: 217.0.

(4-(Chlorosulfonyl)phenyl)boronic acid (16): To a stirred solution of boronic acid derivative **15** (1.08 g, 5 mmol) in methanol (10 mL) at 23 °C, was added TEA (2.8 mL, 20 mmol). The resulting reaction mixture was heated at 60 °C for 24 h. After this period, the solvent was removed completely under vacuum and the residue was used for the next step without purification.

To a stirred solution of the above residue at 0 °C, $SOCl_2$ (10 mL) was added. The reaction mixture was warmed to 23 °C and then heated at 80 °C for 24 h. After this period, the reaction mixture was concentrated under reduced pressure and the resulting residue was carefully quenched with crushed ice. The mixture was extracted with CH_2Cl_2. The combined organic layer was washed with brine, dried over Na_2SO_4 and filtrated. The solvent was removed under reduced pressure to give the title sulfonyl chloride **16** (980 mg slightly yellow solid, 89% yield over two steps). This sulfonyl chloride is sufficiently pure and was used for the next step without purification. An analytical sample was prepared by recrystallization in hexane; mp: decomposed at 110 °C. ^{1}H NMR (400 MHz, CD_3OD) δ = 8.06–8.01 (m, 2H), 7.91–7.97 ppm (m, 2H). ^{13}C NMR (100 MHz, CD_3OD) δ = 144.9, 134.5, 125.3 ppm. MS (EI) calcd for $C_6H_6ClO_2S$ $[M-B(OH)_2+2H]^+$: 177.0, Found: 177.0, 179.0.

(4-(*N*-((2R,3S)-3-((*tert*-Butoxycarbonyl)amino)-2-hydroxy-4-phenylbutyl)-*N*-isobutylsulfamoyl)phenyl)boronic acid (18): To a stirred solution of amine **17** (336 mg, 1 mmol) in CH_2Cl_2 (10 mL) at 0 °C, sulfonyl chloride **16** (264 mg, 1.2 mmol) followed by Et_3N (0.7 mL, 5 mmol) were added. The resulting reaction mixture was stirred at 23 °C for 1 h and TLC showed the complete consumption of the secondary amine. The reaction was quenched with ice water and adjust to pH 6 with 1 N HCl. The mixture was extracted with a mixture (15:1) of CH_2Cl_2 and methanol. The combined organic layer was washed with brine, dried over Na_2SO_4, and filtrated. After the solvent was removed under reduced pressure, the residue was purified by silica gel column chromatography (CH_2Cl_2 and methanol 50:1 to 15:1) to afford sulfonamide **18** (425 mg, 82% yield). ^{1}H NMR (400 MHz, CD_3OD, two rotamers) δ = 7.95–7.70 (m, 4H), 7.25–7.09 (m, 5H), 3.80–3.57 (m, 2H), 3.46–3.56 (m, 1H), 3.15–2.83 (m, 4H), 2.60–2.44 (m, 1H), 2.05–1.93 (m, 1H), 1.27, 1.15 (s, 9H), 0.88 (d, 6.4 Hz, 3H), 0.84 ppm (d, 6.4 Hz, 3H). ^{13}C NMR (100 MHz, CD_3OD, major rotamer) δ = 156.4, 140.2, 138.7, 133.9, 129.3, 129.0, 127.7, 126.0, 125.9, 125.6, 78.6, 72.7, 57.2, 55.3, 52.4, 35.8, 27.3, 26.5, 19.1 ppm. MS (ESI) calcd for $C_{21}H_{30}BN_2O_7S$ $[M-C_4H_9]^+$: 465.2, Found: 465.1; calcd for $C_{25}H_{38}BN_2O_7S$ $[M+H]^+$: 521.3, Found: 521.3.

(4-(*N*-((2*R*,3*S*)-3-(((((3*R*,3*aS*,6*aR*)-Hexahydrofuro[2,3-*b*]furan-3-yl)oxy)carbonyl)amino)-2-hydroxy-4-phenylbutyl)-*N*-isobutylsulfamoyl)phenyl)boronic acid (5): To a stirred solution of Boc-derivative **18** (21 mg, 0.04 mmol) in CH_2Cl_2 (1.5 mL) at 0 °C, TFA (0.5 mL) was added. The resulting reaction mixture was stirred at 0 °C for 30 min. The reaction mixture was concentrated under reduced pressure and ice water (2 mL) was added to the residue. The pH was adjusted to neutral with 2.5 N NaOH. The mixture was extracted with a mixture (15:1) of CH_2Cl_2 and methanol. The combined organic layer was washed with brine, dried over Na_2SO_4 and filtrated. The solvent was removed under reduced pressure, the resulting crude amine was used for the next step without purification.

To a stirred solution of the above amine in CH_3CN (1.0 mL), DIPEA (70 µL, 10 equiv) followed by mixed carbonate **19** (13 mg, 0.04 mmol) were added. The resulting reaction mixture was stirred at 23 °C for 10 days. After this period, the solvent was removed under reduced pressure and 1 N HCl was added to the residue and the pH was adjusted to about 6. The mixture was extracted with a mixture (15:1) of CH_2Cl_2 and methanol. The combined organic layer was washed with brine, dried over Na_2SO_4 and filtrated. The solvent was removed under reduced pressure and the residue was purified by silica gel column chromatography (CH_2Cl_2/methanol/HCO_2H, 100:2:1 to 100:5:1) to afford inhibitor **5** (11 mg, 47% yield) as an amorphous solid. The proton spectrum is identical with the reported one.[3] Analytical data: ^{1}H NMR (400 MHz, CD_3OD) δ = 7.87–7.77 (m, 4H), 7.26–7.16 (m, 5H), 5.59 (d, J = 5.2 Hz, 1H), 4.97–4.91 (m, 1H), 3.94 (dd, J = 9.6, 6.0 Hz, 1H), 3.83–3.64 (m, 5H), 3.45 (dd, J = 14.8, 2.4 Hz, 1H), 3.23–3.10 (m, 2H), 3.01–2.84 (m, 3H), 2.53 (dd, J = 14.0, 10.8 Hz, 1H), 2.09–1.99 (m, 1H), 1.56–1.47 (m, 1H), 1.38–1.32 (m, 1H), 0.94 (d, J = 6.4 Hz, 3H), 0.88 ppm (d, J = 6.4 Hz, 3H). ^{13}C NMR (200 MHz, CD_3OD) δ = 157.8, 157.7, 140.3, 135.1, 130.5, 129.3, 127.4, 127.2, 110.8, 74.6, 74.5, 72.1, 70.6, 58.9, 57.4, 53.9, 46.9, 37.1, 28.0, 27.0, 20.5, 20.4 ppm. MS (ESI) calcd for $C_{27}H_{38}BN_2O_9S$ $[M+H]^+$: 577.2, Found: 577.3.

(4-(*N*-((2*R*,3*S*)-3-(((((3*S*,7a*S*,8*S*)-Hexahydro-4*H*-3,5-methanofuro[2,3-*b*]pyran-8-yl)oxy)carbonyl)amino)-2-hydroxy-4-phenylbutyl)-*N*-isobutylsulfamoyl)phenyl)boronic acid (6): Following the procedure outlined for inhibitor **5**, reaction of amine derived from Boc-derivative **18** (20 mg, 0.04 mmol) and activated *crown*-THF carbonate **12** (10 mg, 0.031 mmol) afforded inhibitor **6** (12.5 mg, 55% for two steps) as an amorphous solid. ^{1}H NMR (400 MHz, CD_3OD) δ = 7.79 (s, 5H), 7.30–7.21 (m, 4H), 7.19–7.13 (m, 1H), 5.34 (d, J = 6.2 Hz, 1H), 4.71–4.64 (m, 1H), 3.95 (d, J = 11.2 Hz, 1H), 3.81–3.67 (m, 2H), 3.61–3.54 (m, 2H), 3.47–3.35 (m, 2H), 3.26–3.20 (m, 1H), 3.17 (dd, J = 14.0, 3.4 Hz, 1H), 3.10 (dd, J = 13.6, 8.1 Hz, 1H), 3.02–2.80 (m, 2H), 2.73–2.60 (m, 2H), 2.55 (dd, J = 13.8, 11.0 Hz, 1H), 2.32 (m, 1H), 2.06–1.96 (m, 1H), 1.76 (d, J = 12.0 Hz, 1H), 1.48 (dt, J = 12.1, 3.9 Hz, 1H), 0.91 (d, J = 6.6 Hz, 3H), 0.86 ppm (d, J = 6.9 Hz, 3H); ^{13}C NMR (200 MHz, CD_3OD) δ = 156.4, 138.8, 133.7, 129.1, 128.8, 127.9, 126.0, 125.8, 104.3, 74.7, 73.0, 72.1, 68.0, 59.7, 57.4, 56.1, 56.0, 55.0, 53.4, 52.5, 48.1, 44.7, 42.0, 37.4, 35.8, 26.6, 22.8, 19.1, 19.0 ppm. LRMS-ESI (*m/z*): 603.2 $[M+H]^+$. HRMS-ESI (*m/z*): $[M+Na]^+$ calcd for $C_{29}H_{39}BN_2O_9SNa$, 624.2398; found 624.2393.

(4-(*N*-((2*R*,3*S*)-3-(((((3a*S*,4*S*,7a*R*)-Hexahydro-4*H*-furo[2,3-*b*]pyran-4-yl)oxy)carbonyl)amino)-2-hydroxy-4-phenylbutyl)-*N*-isobutylsulfamoyl)phenyl)boronic acid (20): Following the procedure outlined for inhibitor **5**, reaction of amine salt derived from Boc-derivative **18** (21 mg, 0.04 mmol) and activated carbonate **10** (13.6 mg, 0.04 mmol) afforded inhibitor **20** (12 mg, 51%) as an amorphous solid. ^{1}H NMR (400 MHz, CD_3OD, two isomers) δ = 7.79 (br s, 4H), 7.31–7.09 (m, 5H), 4.92–4.86 (m, 2H), 4.05 (t, J = 7.6 Hz, 1H), 3.84–3.70 (m, 4H), 3.46–3.32 (m, 2H), 3.21–3.06 (m, 2H), 3.02–2.84 (m, 2H), 2.57–2.49 (m, 1H), 2.35–2.25 (m, 1H), 2.07–1.95 (m, 1H), 1.92–1.79 (m, 1H), 1.78–1.60 (m, 2H), 1.46–1.36 (m, 1H), 0.91 (d, 6.4 Hz, 3H), 0.86 ppm (d, 6.4 Hz, 3H). ^{13}C NMR (100 MHz, CD_3OD, major isomer) δ = 156.2, 139.7, 138.8, 133.6, 129.2, 129.0, 127.7, 125.9, 125.6, 100.9, 72.8, 69.4, 67.8, 60.2, 57.3, 55.7, 55.6, 52.4, 43.3, 35.7, 26.5, 25.8, 21.7, 18.9 ppm. HRMS (ESI) calcd for $C_{28}H_{39}BN_2O_9SNa$ $[M+Na]^+$: 612.2398, Found: 612.2391.

Determination of X-ray structures of HIV-1 protease–inhibitor complexes for inhibitors 6 and 20: HIV-1 PR optimized for bacterial expression was purified as described.[39] PR at 4.2 mg mL^{-1} was complexed with inhibitor at 5:1 molar ratio. Crystals of protease–inhibitor complexes were grown by hanging drop vapor diffusion with well solutions of 1.4 M sodium chloride, 0.1 M sodium acetate buffer pH 6.0 and 1.3 M sodium chloride, 0.1 M sodium acetate pH 6.2 reservoir solution for inhibitors **6** and **20**, respectively.

Streak-seeding was employed to induce crystallization for inhibitor **6**. Crystals were cryo-cooled to 90 K and 1.0 Å wavelength X-ray diffraction data was collected on beamline 22-ID at SER-CAT, Advanced Photon Source, Argonne National Lab (Chicago, IL, USA). Data were processed and scaled using HKL2000.[40] The structures were solved by molecular replacement in PHASER[41] in CCP4i suite,[42,43] using PR-APV (PDB ID: 3NU3)[44] as an isomorphous model and refined with COOT,[45,46] and Refmac5[47] in CCP4i, using anisotropic displacement parameters (B-factors). JLigand[48] in CCP4i, was used to generate refinement restraints for inhibitors. Crystallographic statistics are supplied in Table 1. Coordinates and structure factors for inhibitors **6** (GRL-008–19A) and **20** (GRL-031-19A) were deposited in the Protein Data Bank,[49] with accession codes 6U7O and 6U7P.

Acknowledgements

The present research was supported by grants from the National Institutes of Health AI150466 (A.K.G.) and AI150461 (I.T.W.). D.W.K. is supported by Georgia State University Molecular Basis of Disease Fellowship. This work was also supported in part by the Intramural Research Program of the Center for Cancer Research, National Cancer Institute, National Institutes of Health (H.M.: R01AI121315); a Grant-in-Aid for Scientific Research (Priority Areas) from the Ministry of Education, Culture, Sports, Science, and Technology of Japan (Monbu-Kagakusho), a Grant for Promotion of AIDS Research from the Ministry of Health, Welfare, and Labor of Japan; grants from Japan Agency for Medical Research and Development (AMED) (H.M.: JP15fk0410001 and JP18fk0410001); and grants from National Center for Global Health and Medicine Research Institute (H.M.). We thank the staff at the Southeast Regional-Collaborative Access Team (SER-CAT) at the Advanced Photon Source, Argonne National Laboratory, for assistance during X-ray data collection. Use of the Advanced Photon Source was supported by the US Department of Energy, Office of Science, Office of Basic Energy Sciences, under Contract No. W-31-109-Eng-38. We also thank the Purdue University Center for Cancer Research, which supports the shared NMR and mass spectrometry facilities.

Conflict of interest

The authors declare no conflict of interest.

Keywords: brain penetration · drug resistance · genetic barriers · HIV-1 protease inhibitors · structure-based design

[1] A. K. Ghosh, S. Gemma, *Structure-Based Design of Drugs and Other Bioactive Molecules: Tools and Strategies*, Wiley-VCH, **2014**, 337–354.
[2] R. E. Hubbard, *Structure-Based Drug Discovery, An Overview*, RSC Publishing, **2006**.
[3] A. K. Ghosh, H. L. Osswald, G. Prato, *J. Med. Chem.* **2016**, *59*, 5172–5208.
[4] A. Wlodawer, J. Vondrasek, *Annu. Rev. Biophys. Biomol. Struct.* **1998**, *27*, 249–284.
[5] S. Broder, *Antiviral Res.* **2010**, *85*, 1–18.
[6] A. Edmonds, M. Yotebieng, J. Lusiama, Y. Matumona, F. Kitetele, S. Napravnik, S. R. Cole, A. Van Rie, F. Behets, *PLoS Med.* **2011**, *8*, e1001044.
[7] C. Peng, B. K. Ho, T. W. Chang, N. T. Chang, *J. Virol.* **1989**, *63*, 2550–2556.
[8] N. E. Kohl, E. A. Emini, W. A. Schleif, L. J. Davis, J. C. Heimbach, R. A. Dixon, E. M. Scolnick, I. S. Sigal, *Proc. Natl. Acad. Sci. USA* **1988**, *85*, 4686–4696.
[9] A. K. Ghosh, Z. L. Dawson, H. Mitsuya, *Bioorg. Med. Chem.* **2007**, *15*, 7576–7580.
[10] M. P. de Bethune, V. Sekar, S. Spinosa-Guzman, M. Vanstockem, S. De Meyer, P. Wigerinck, E. Lefebvre, *Darunavir (Prezista, TMC114): from Bench to Clinic, Improving Treatment Options for HIV-infected Patients in Antiviral Drugs: from Basic Discovery through Clinical Trials*, John Wiley & Sons, Hoboken, **2011**, pp. 31e45.
[11] *Guidelines for the Use of Antiretroviral Agents in HIV-1-Infected Adults and Adolescents*: https://aidsinfo.nih.gov/contentfiles/lvguidelines/adultandadolescentgl.pdf (accessed March 31, 2018).
[12] S. Hué, R. J. Gifford, D. Dunn, E. Fernhill, D. Pillay, *J. Virol.* **2009**, *83*, 2645–2654.
[13] H. Hayashi, N. Takamune, T. Nirasawa, M. Aoki, Y. Morishita, D. Das, Y. Koh, A. K. Ghosh, S. Misumi, H. Mitsuya, *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 12234–12239.
[14] A. K. Ghosh, D. D. Anderson, I. T. Weber, H. Mitsuya, *Angew. Chem. Int. Ed.* **2012**, *51*, 1778–1802; *Angew. Chem.* **2012**, *124*, 1812–1838.
[15] A. K. Ghosh, B. Chapsal, I. T. Weber, H. Mitsuya, *Acc. Chem. Res.* **2008**, *41*, 78–86.
[16] Y. Tie, P. I. Boross, Y.-F. Wang, L. Gaddis, A. K. Hussian, S. Leshchenko, A. K. Ghosh, J. M. Louis, R. W. Harrison, I. T. Weber, *J. Mol. Biol.* **2004**, *338*, 341–352.
[17] A. Y. Kovalevsky, F. Liu, S. Leshchenko, A. K. Ghosh, J. M. Louis, R. W. Harrison, I. T. Weber, *J. Mol. Biol.* **2006**, *363*, 161–173.
[18] Y. Koh, H. Nakata, K. Maeda, H. Ogata, G. Bilcer, T. Devasamudram, J. F. Kincaid, P. Boross, Y.-F. Wang, Y. Tie, P. Volarath, L. Gaddis, R. W. Harrison, I. T. Weber, A. K. Ghosh, H. Mitsuya, *Antimicrob. Agents Chemother.* **2003**, *47*, 3123–3129.
[19] S. De Meyer, H. Azijn, D. Surleraux, D. Jochmans, A. Tahri, R. Pauwels, P. Wigerinck, M. P. de Bethune, *Antimicrob. Agents Chemother.* **2005**, *49*, 2314–2321.
[20] G. Sterrantino, M. Zaccarelli, G. Colao, F. Baldanti, S. Di Giambenedetto, T. Carli, F. Maggiolo, M. Zazzi, *Infection* **2012**, *40*, 311–318.
[21] S. De Meyer, E. Lathouwers, I. Dierynck, E. De Paepe, B. Van Baelen, T. Vangeneugden, S. Spinosa-Guzman, E. Lefebvre, G. Picchio, M.-P. de Béthune, *AIDS* **2009**, *23*, 1829–1840.
[22] R. S. Yedidi, K. Maeda, W. S. Fyvie, M. Steffey, D. A. Davis, I. Palmer, M. Aoki, J. D. Kaufman, S. J. Stahl, H. Garimella, D. Das, P. T. Wingfield, A. K. Ghosh, H. Mitsuya, *Antimicrob. Agents Chemother.* **2013**, *57*, 4920–4927.
[23] S.-F. Zhou, *Xenobiotica* **2008**, *38*, 802–832.
[24] P. G. Hoggard, A. Owen, *J. Antimicrob. Chemother.* **2003**, *51*, 493–496.
[25] H. Fujimoto, M. Higuchi, H. Watanabe, Y. Koh, A. K. Ghosh, H. Mitsuya, N. Tanoue, A. Hamada, H. Saito, *Biol. Pharm. Bull.* **2009**, *32*, 1588–1593.
[26] I. W. Windsor, M. J. Palte, J. C. Lukesh III, B. Gold, K. T. Forest, R. T. Raines, *J. Am. Chem. Soc.* **2018**, *140*, 14015–14018.
[27] A. K. Ghosh, P. R. Sridhar, S. Leshchenko, A. K. Hussain, J. Li, A. Y. Kovalevsky, D. E. Walters, J. Wedekind, V. Grum-Tokars, D. Das, Y. Koh, K. Maeda, H. Gatanaga, I. T. Weber, H. Mitsuya, *J. Med. Chem.* **2006**, *49*, 5252–5261.
[28] B. S. Bal, W. E. Childers, H. W. Pinnick, *Tetrahedron* **1981**, *37*, 2091–2096.
[29] A. K. Ghosh, B. D. Chapsal, G. L. Parham, M. Steffey, J. Agniswamy, Y.-F. Wang, M. Amano, I. T. Weber, H. Mitsuya, *J. Med. Chem.* **2011**, *54*, 5890–5901.
[30] A. K. Ghosh, K. V. Rao, P. R. Nyalapatla, H. L. Osswald, C. D. Martyr, M. Aoki, H. Hayashi, J. Agniswamy, Y. F. Wang, H. Bulut, D. Das, I. T. Weber, H. Mitsuya, *J. Med. Chem.* **2017**, *60*, 4267–4278.
[31] E. Kunst, F. Gallier, G. Dujardin, M. S. Yusubov, A. Kirschning, *Org. Lett.* **2007**, *9*, 5199–5202.
[32] Sulfonyl chloride **16** is quite stable. The preliminary X-ray structure shows mixed conformational isomers. Further refinement is in progress.
[33] A. K. Ghosh, W. S. Fyvie, M. Brindisi, M. Steffey, J. Agniswamy, Y.-F. Wang, M. Aoki, M. Amano, I. T. Weber, H. Mitsuya, *ChemMedChem* **2017**, *12*, 1942–1952.
[34] M. V. Toth, G. R. Marshall, *Int. J. Pept. Protein Res.* **1990**, *36*, 544–550.
[35] Y. Koh, M. Amano, T. Towata, M. Danish, S. Leshchenko-Yashchuk, D. Das, M. Nakayama, Y. Tojo, A. K. Ghosh, H. Mitsuya, *J. Virol.* **2010**, *84*, 11961–11969.
[36] B. E. Smart, *J. Fluorine Chem.* **2001**, *109*, 3–11.

[37] M. Aoki, H. Hayashi, K. V. Rao, D. Das, N. Higashi-Kuwata, H. Bulut, H. Aoki-Ogata, Y. Takamatsu, R. S. Yedidi, D. A. Davis, S.-i. Hattori, N. Nishida, K. Hasegawa, N. Takamune, P. R. Nyalapatla, H. L. Osswald, H. Jono, H. Saito, R. Yarchoan, S. Misumi, A. K. Ghosh, H. Mitsuya, *eLife* **2017**, *6*, e28020.

[38] A. K. Ghosh, K. V. Rao, P. R. Nyalapatla, S. Kovela, M. Brindisi, H. L. Osswald, B. Sekhara Reddy, J. Agniswamy, Y. F. Wang, M. Aoki, S. I. Hattori, I. T. Weber, H. Mitsuya, *ChemMedChem* **2018**, *13*, 803–815.

[39] J. M. Sayer, J. Agniswamy, I. T. Weber, J. M. Louis, *Protein Sci.* **2010**, *19*, 2055–2072.

[40] Z. Otwinowski, W. Minor, *Processing of X-ray Diffraction Data Collected in Oscillation Mode. Methods in Enzymology, 276: Macromolecular Crystallography, Part A* (Eds.: C. W. Carter, Jr., R. M. Sweet) Academic Press, New York, **1997**, pp. 307–326.

[41] A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni, R. J. Read, *J. Appl. Crystallogr.* **2007**, *40*, 658–674.

[42] M. D. Winn, C. C. Ballard, K. D. Cowtan, E. J. Dodson, P. Emsley, P. R. Evans, R. M. Keegan, E. B. Krissinel, A. G. W. Leslie, A. McCoy, S. J. McNicholas, G. N. Murshudov, N. S. Pannu, E. A. Potterton, H. R. Powell, R. J. Read, A. Vagin, K. S. Wilson, *Acta Crystallogr. Sect. D* **2011**, *67*, 235–242.

[43] E. Potterton, P. Briggs, M. Turkenburg, E. Dodson, *Acta Crystallogr. Sect. D* **2003**, *59*, 1131–1137.

[44] C.-H. Shen, Y.-F. Wang, A. Y. Kovalevsky, R. W. Harrison, I. T. Weber, *FEBS J.* **2010**, *277*, 3699–3714.

[45] P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, *Acta Crystallogr. Sect. D* **2010**, *66*, 486–501.

[46] P. Emsley, K. Cowtan, *Acta Crystallogr. Sect. D* **2004**, *60*, 2126–2132.

[47] G. N. Murshudov, A. A. Vagin, E. J. Dodson, *Acta. Crystallogr. Sect. D* **1997**, *53*, 240–255.

[48] A. A. Lebedev, P. Young, M. N. Isupov, O. V. Moroz, A. A. Vagin, G. N. Murshudov, *Acta. Crystallogr. Sect. D* **2012**, *68*, 431–440.

[49] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, P. E. Bourne, *Nucleic Acids Res.* **2000**, *28*, 235–242.

Manuscript received: September 3, 2019
Revised manuscript received: September 18, 2019
Accepted manuscript online: September 23, 2019
Version of record online: ■■ ■■, 0000