

PURIFICATION AND CHARACTERIZATION OF A PRE-THERAPY HIV-1 PROTEASE
VARIANT AND VARIANTS CONTAINING DRUG-PRESSURE SELECTED
MUTATIONS FOR ELECTRON PARAMAGNETIC RESONANCE (EPR) STUDIES

By

ESTRELLA GARLIT GONZALES

A THESIS PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE

UNIVERSITY OF FLORIDA

2011

© 2011 Estrella Garlit Gonzales

To my Mom Evie, whose love I will always remember,
to my ever supportive and loving Dad John, and to my siblings, especially Ate Emmy

ACKNOWLEDGEMENTS

I am forever grateful to God for his guidance and for whom I gather my strength and will to live. It is because of Him that we humans are born into this world having the gift of wisdom and knowledge to ceaselessly understand the processes that drive the world, which He created with His greatness.

I dedicate this to my parents, John and Evie Gonzales, for their unending love, guidance, and support. My Mom may not be in this world anymore but I will forever cherish the love and memories we shared. I thank Mom and Dad for having molded me into a disciplined and hard-working person not only for me to attain self-fulfillment but most importantly to strive for the benefit of the other.

I thank my sisters, Emmy and Grace, and brothers, Elmer and Ellsworth, for their unwavering love and support. I particularly thank Ate Emmy for her help in everything and just for being a sister and a friend at the same time. I am very grateful for Alex Aloy, for his unconditional love and for just always being there for me. I thank all my other friends for their care and support.

I sincerely thank my adviser, Dr. Gail Fanucci, for the opportunity to work in her research group. I thank her for providing knowledge and insights not just about technical aspects of research but also about dealing with shortcomings in work and life, in general. I am grateful for her encouragement and support.

I would like to thank my committee members, Dr. Ben Dunn of the Department of Biochemistry and Molecular Biology in the UF College of Medicine and Dr. Ben Smith. I thank Dr. Dunn for his collaboration with the HIV-1 protease project and for valuable insights about my research. I especially thank Dr. Smith for his patience and encouragement and for believing in my potential.

I would like to express my gratitude to Dr. Alex Angerhofer for help in the use of the EPR instrument. I likewise thank Dr. Steve Hagen of the Department of Physics for the use of his CD spectrometer. I also thank Dr. Cristina Dancel of the Department of Chemistry Mass Spectrometry Facility for helping me run and analyze my samples. I am also indebted to her because she has unconditionally helped me survive in graduate school.

I would like to thank my fellow Fanucci group members for their friendship and support. I especially thank Dr. Jamie Kear, Dr. Mandy Blackburn, Mike Veloro, Ian de Vera, and Rochelle Huang for helpful discussions on the HIV-1 protease project and because of them I learned so much about biochemistry and biophysical techniques. I also thank Dr. Luis Galiano for conceptualizing the HIV-1 PR project.

I acknowledge the National High Magnetic Field Lab (NHMFL) in-house research proposal (IHRP), National Institute of Health (NIH), National Science Foundation (NSF), and University of Florida (UF) Startup for funding.

TABLE OF CONTENTS

	<u>page</u>
ACKNOWLEDGEMENTS	4
LIST OF TABLES.....	9
LIST OF FIGURES.....	11
LIST OF ABBREVIATIONS.....	16
ABSTRACT	21
CHAPTER	
1 INTRODUCTION	23
Introduction to HIV	23
HIV-1 Structure and Genome	24
HIV-1 Life Cycle	25
Introduction to HIV-1 Protease.....	26
Structure and Function	26
Conformational Sampling of HIV-1 Protease.....	28
HIV-1 Protease Cleavage Sites and Substrate Recognition.....	30
HIV-1 Protease Inhibitors	32
HIV-1 Protease Variants and Subtype Polymorphisms	34
HIV-1 Protease Drug Resistant Variants	35
Mechanisms of drug resistance	38
Influence of protease cleavage site mutations on drug resistance.....	39
Scope of Work	40
Details of HIV-1 PR Constructs	40
Previous Work on HIV-1 PR Constructs	41
Objectives of Work	43
Summary	44
2 BACKGROUND ON TECHNIQUES	56
Cloning and Protein Expression Systems.....	56
Site-Directed Mutagenesis.....	59
Purification and Characterization of Proteins	61
Isolation of Proteins from Inclusion Bodies and Refolding.....	61
Chromatography.....	62
SDS-PAGE.....	64
Circular Dichroism (CD) Spectroscopy	65
Mass Spectrometry (MS).....	67
Electron Paramagnetic Resonance (EPR) Spectroscopy	67

	Site-Directed Spin Labeling	68
	Continuous-Wave EPR (CW-EPR).....	69
	Double Electron-Electron Resonance (DEER) Spectroscopy	71
	DEER Data Analysis	73
	Tikhonov Regularization	74
	Background Subtraction and Self-Consistent Analysis.....	76
	Gaussian Reconstruction	78
3	CLONING AND MUTAGENESIS	89
	Materials and Methods.....	89
	Cloning of HIV-1 PR PRE and POST Constructs	89
	Site-Directed Mutagenesis of HIV-1 PR Constructs	90
	Results and Discussion.....	92
	Cloning	92
	Mutagenesis	94
4	PURIFICATION AND CHARACTERIZATION OF HIV-1 PROTEASE CONSTRUCTS	104
	Materials and Methods.....	104
	Expression of HIV-1 PR.....	104
	Purification of HIV-1 PR.....	105
	Spin Labeling of HIV-1 PR.....	107
	SDS-PAGE	107
	Circular Dichroism (CD) Spectroscopy	108
	Mass Spectrometry (MS).....	108
	CW-EPR.....	109
	Results and Discussion.....	109
	SDS-PAGE	110
	CD Spectroscopy	111
	MS	112
	CW-EPR.....	113
5	DISTANCE MEASUREMENTS OF HIV-1 PROTEASE CONSTRUCTS	132
	Previous DEER Studies on HIV-1 PR	133
	Materials and Methods.....	136
	Sample Preparation for DEER Data Collection	136
	DEER Experiments	137
	DEER Data Analysis	138
	Results and Discussion.....	138
	Comparison of HIV-1 PR Apo Constructs.....	139
	Comparison of HIV-1 PR Constructs With and Without RTV.....	147
	Experimental Limitations	151
6	CONCLUSIONS AND FUTURE WORK	172

Conclusions 172
Future Work 174

APPENDIX

A PRIMER CHARACTERISTICS AND PCR PARAMETERS FOR
MUTAGENESIS EXPERIMENTS 175

B ADDITIONAL INFORMATION ON THE PURIFICATION AND
CHARACTERIZATION OF HIV-1 PR CONSTRUCTS..... 178

C POPULATION VALIDATION OF SELECTED HIV-1 PR CONSTRUCTS..... 184

LIST OF REFERENCES 189

BIOGRAPHICAL SKETCH..... 203

LIST OF TABLES

<u>Table</u>	<u>page</u>
1-1 Various proteins expressed by the HIV-1 genes.....	46
1-2 Cleavage sites in HIV-1 polyproteins.....	50
1-3 Protease inhibitor (PI) resistance-associated mutations.....	54
1-4 Sequence alignment of subtype B LAI, B _{si} , PRE _i , POST _i , V6 _i , and MDR769 _j	54
1-5 Amino acid substitutions made on post-therapy (POST) sequence.....	55
3-1 <i>E. coli</i> codon-optimized DNA and amino acid sequence of PRE.....	97
3-2 <i>E. coli</i> codon-optimized DNA and amino acid sequence of POST.....	97
3-3 Naming convention for HIV-1 PR constructs.	98
3-4 <i>E. coli</i> codon-optimized DNA and amino acid sequence of PMUT1.	98
3-5 <i>E. coli</i> codon-optimized DNA and amino acid sequence of PMUT3.	98
3-6 <i>E. coli</i> codon-optimized DNA and amino acid sequence of PMUT2.	99
3-7 <i>E. coli</i> codon-optimized DNA and amino acid sequence of PMUT5.	99
3-8 <i>E. coli</i> codon-optimized DNA and amino acid sequence of B _{si}	99
3-9 <i>E. coli</i> codon-optimized DNA and amino acid sequence of B _{si} -I63P.....	100
4-1 Theoretical isoelectric points (pI) of HIV-1 PR constructs.....	116
4-2 Standard parameters used for circular dichroism (CD) spectroscopy data collection.	116
4-3 Standard parameters used for CW-EPR data collection.....	116
4-4 Theoretical MW of HIV-1 PR constructs with and without MTSL and MW determined from ESI-MS analysis.	117
5-1 Standard parameters for DEER data collection.....	157
5-2 Parameters used for the pulse sequence.	157
5-3 Individual population profiles for PRE apo.....	158
5-4 Individual population profiles for POST apo.	159

5-5	Individual population profiles for PMUT1 apo.	160
5-6	Individual population profiles for B _{si} apo.	161
5-7	Individual population profiles for B _{si} -I63P apo.....	162
5-8	Summary of most probable distance of HIV-1 PR apo constructs.....	163
5-9	Comparison of relative percentages of the individual populations for HIV-1 PR PRE apo at different τ values.	165
5-10	Comparison of the relative percentages of the individual populations for HIV-1 PR apo constructs.	166
5-11	Individual population profiles for PRE RTV.....	168
5-12	Individual population profiles for POST RTV.	169
5-13	Comparison of relative percentages of the individual populations for PRE and POST HIV-1 PR constructs in the absence and presence of RTV.	170
A-1	Primers to produce the pET23_ PMUT1 construct.	175
A-2	Primers to produce the pET23_ PMUT3 construct.	175
A-3	Primers to produce the pET23_ PMUT2 construct.	175
A-4	Primers to produce the pET23_ PMUT5 construct.	176
A-5	Primers to produce the pET23_ B _{si} -I63P construct.....	176
A-6	Components of the PCR mixture.....	176
A-7	Thermal cycling parameters for HIV-1 protease site-directed mutagenesis reactions.....	177
B-1	Buffers used in the purification of HIV-1 PR.	178
B-2	Optical density at 280 nm and estimated protein concentration of HIV-1 PR samples used for CD data collection.	178
B-3	Optical density at 280 nm and estimated concentration of HIV-1 PR samples used for EPR data collection	179

LIST OF FIGURES

<u>Figure</u>	<u>page</u>
1-1 The assembly of viral proteins and other constituents in HIV-1.....	45
1-2 Schematic diagram of the HIV-1 genome.....	45
1-3 Viral life cycle of HIV-1.. ..	47
1-4 Structural assembly of HIV-1 from an immature to a mature virus.	47
1-5 Ribbon diagrams of HIV-1 PR highlighting the main regions.....	48
1-6 Illustration of the “fireman’s grip”.. ..	48
1-7 Crystal structures of HIV-1 PR in the closed conformation.....	49
1-8 Molecular dynamics (MD) simulation structures of HIV-1 PR.....	49
1-9 Schematic representation of HIV-1 <i>gag</i> and <i>gag-pol</i> polyproteins.....	50
1-10 Chemical structure of HIV-1 PR substrate.....	50
1-11 Substrate binding pocket of HIV-1 protease.	51
1-12 Various types of HIV-1inhibitors	51
1-13 Chemical structures of the FDA-approved HIV-1 protease inhibitors.	52
1-14 General classification of HIV.....	53
1-15 Sequence variation of protease inhibitor (PI) naïve and treated isolates of subtype B HIV-1 PR.	53
1-16 Ribbon diagrams of HIV-1 PR PRE, POST, V6, and MDR769.....	55
2-1 Example of vector map.....	80
2-2 Host and vector elements essential in protein overexpression.....	80
2-3 pET23a vector map.	81
2-4 Structure of isopropyl-β-D-1-thiogalactopyranoside (IPTG).....	81
2-5 Various steps in site-directed mutagenesis	81
2-6 Typical circular dichroism (CD) spectra.	82

2-7	Energy diagram representing the Zeeman Effect.	82
2-8	Chemical modification of the protein side chain.....	83
2-9	Typical EPR spectrum of a system with electron spin state of $m_s = \pm\frac{1}{2}$	83
2-10	Energy diagram and EPR spectrum of a system with an electron spin state of $m_l = 0, \pm 1$	83
2-11	EPR line shapes representing various modes of spin motion.....	84
2-12	Pulse sequence of the four-pulse DEER experiment.....	84
2-13	Illustration of the intramolecular and intermolecular interactions in doubly labeled HIV-1 protease.....	85
2-14	Plot of the experimental dipolar evolution function	85
2-15	Pake pattern obtained by Fourier transformation.....	86
2-16	Example of an L-curve.....	86
2-17	Flow-chart of the Self-Consistent Analysis (SCA) method.....	87
2-18	Example of population validation process.	88
3-1	Map of original vector containing the Pre _i -K55C and Post _i -K55C HIV-1 PR DNA inserts.	97
3-2	DNA gel of pET23a plasmid containing PRE and POST HIV-1 PR DNA.	100
3-3	Sample DNA gel after <i>DpnI</i> digestion.	101
3-4	DNA gel of pET23_PMUT1.	101
3-5	DNA gel of pET23_PMUT3.	102
3-6	DNA gel of pET23_PMUT2.	102
3-7	DNA gel of pET23_PMUT5.	103
4-1	SDS-PAGE gel of PRE.....	117
4-2	SDS-PAGE gel of POST.	118
4-3	SDS-PAGE gel of PMUT1.....	118
4-4	SDS-PAGE gel of PMUT3.....	119

4-5	SDS-PAGE gel of PMUT2.	119
4-6	SDS-PAGE gel of PMUT5.	120
4-7	SDS-PAGE gel of B _{Si}	120
4-8	SDS-PAGE gel of B _{Si} -I63P.	121
4-9	SDS-PAGE gel of spin labeled HIV-1 PR constructs.	121
4-10	Circular dichroism spectra for spin labeled HIV-1 PR constructs.	122
4-11	Mass spectra of spin labeled PRE.	123
4-12	Mass spectra of spin labeled POST.	124
4-13	Mass spectra of spin labeled PMUT1.	125
4-14	Mass spectra of spin labeled PMUT3.	126
4-15	Mass spectra of spin labeled PMUT2.	127
4-16	Mass spectra of spin labeled PMUT5.	128
4-17	Mass spectra of spin labeled B _{Si}	129
4-18	Mass spectra of spin labeled B _{Si} -I63P.	130
4-19	CW-EPR spectra for spin labeled HIV-1 PR constructs.	131
5-1	HIV-1 PR crystal structures with spin labels at the K55C/K55C' sites.	154
5-2	Dipolar evolution curves and distance distribution profiles for apo and RTV-bound subtype B HIV-1 PR.	154
5-3	Distance distribution profiles of subtype B HIV-1 PR apo and with inhibitors ...	155
5-4	Dipolar evolution curves and distance distribution profiles for subtype B apo, V6, and MDR769.	155
5-5	Distance distribution profiles of apo forms of HIV-1 PR subtype B, C, F, CRF01_A/E and multi-drug resistant constructs, V6 and MDR769.	156
5-6	Absorption spectra for a nitroxide spin-label.	156
5-7	DEER data analysis for PRE apo.	158
5-8	DEER data analysis for POST apo.	159

5-9	DEER data analysis for PMUT1 apo..	160
5-10	DEER data analysis for B _{si} apo..	161
5-11	DEER data analysis for B _{si} -I63P apo.....	162
5-12	Dipolar evolution curves and distance distribution profiles of HIV-1 PR PRE, POST, PMUT1, B _{si} , and B _{si} -I63P apo constructs.....	163
5-13	Overlay of distance distribution profiles of HIV-1 PR apo constructs with B _{si} ...	164
5-14	Overlay of distance distribution profiles between HIV-1 PR apo constructs	164
5-15	Dipolar evolution curves, individual Gaussian populations, and L-curves at different τ values.....	165
5-16	Relative percentage of individual populations for HIV-1 PR PRE, POST, PMUT1, B _{si} , and B _{si} -I63P apo constructs..	166
5-17	Population analysis of HIV-1 PR PRE, POST, PMUT1, B _{si} , and B _{si} -I63P apo constructs..	167
5-18	Ribbon diagrams of HIV-1 PR PRE and POST	167
5-19	DEER data analysis for PRE RTV.....	168
5-20	DEER data analysis for POST RTV.....	169
5-21	Overlay of dipolar evolution curves and distance distribution profiles of PRE apo and PRE RTV and POST apo and POST RTV.....	170
5-22	Relative percentage of individual populations for PRE and POST HIV-1 PR constructs in the absence and presence of RTV.	171
5-23	Change in relative percent of individual populations for PRE and POST HIV-1 PR constructs in the absence and presence of RTV.	171
B-1	Calibration curve of 4-oxo-TEMPO standard.....	179
B-2	CW-EPR spectra for spin labeled HIV-1 PR PRE, POST, and PMUT1	180
B-3	Mass spectra of spin labeled PRE.....	181
B-4	Mass spectra of spin labeled POST..	182
B-5	Mass spectra of spin labeled PMUT1.	183
C-1	Population validation of PRE apo.	184

C-2	Population validation of PRE RTV.....	185
C-3	Population validation of POST RTV.....	186
C-4	Population validation of PMUT1 apo.	187
C-5	Population validation of B _{si} -I63P apo.....	188

LIST OF ABBREVIATIONS

AIDS	Acquired immune deficiency syndrome
A	Adenine
Ala (A)	Alanine
<i>Amp</i>	Ampicillin
APV	Amprenavir
<i>Rev</i>	Anti-repression transactivator protein
APT	Approximate Pake transformation
Arg (R)	Arginine
Asn (N)	Asparagine
Asp (D)	Aspartic acid
ATV	Atazanavir
AZT	β -D-(+)-3'-azido-3'-deoxythymidine
<i>Bla</i>	β -lactamase
BME	β -mercaptoethanol
μ_B	Bohr magneton ($9.27400949 \times 10^{-24} \text{ J}\cdot\text{T}^{-1}$)
CA	Capsid
CD	Circular dichroism
CRF	Circular recombinant form
CD4	Cluster of differentiation 4
CW-EPR	Continuous wave-electron paramagnetic resonance
Cys (C)	Cysteine
C	Cytosine
Da	Dalton
DRV	Darunavir

dNTP	Deoxynucleotide triphosphate
DNA	Deoxyribonucleic acid
DSC	Differential scanning calorimetry
DMSO	Dimethylsulfoxide
DTT	Dithiothreitol
DEER	Double electron-electron resonance
dsDNA	Double stranded-DNA
EI	Electron impact
EPR	Electron paramagnetic resonance
ESE	Electron spin echo
ESI	Electrospray ionization
<i>Env</i>	Envelope
EDTA	Ethylenediaminetetraacetic acid
FPV	Fosamprenavir
FWHM	Full width at half maximum
Glu (E)	Glutamic acid
Gln (Q)	Glutamine
Gly (G)	Glycine
gp	Glycoprotein
diGly (Gly-Gly)	Glycylglycine (2-[(2-aminoacetyl)amino]acetic acid)
<i>Gag</i>	Group specific antigen
G	Guanine
<i>g</i>	Gyromagnetic ratio or spectroscopic factor
HAART	Highly-active antiretroviral therapy
His (H)	Histidine

HIV	Human immunodeficiency virus
IB	Inclusion body
IDV	Indinavir
IN	Integrase
pI	Isoelectric point
Ile (I)	Isoleucine
IPTG	Isopropyl- β -D-thiogalactoside
ITC	Isothermal titration calorimetry
kDa	Kilodalton
LSB	Laemmli sample buffer
Leu (L)	Leucine
LBF	Local backbone fluctuations
LPV	Lopinavir
LB	Luria-Bertani
Lys (K)	Lysine
B	Magnetic field
MHC	Major histocompatibility complex
m_e	Mass of electron (9.109×10^{-31} kg)
MS	Mass spectrometry
MA	Matrix
T_m	Melting temperature
mRNA	Messenger RNA
MTSL	(1-oxy-2,2,5,5-tetramethylpyrrolinyl-3-methyl) methanethiosulfonate
Met (M)	Methionine
MD	Molecular dynamics

MW	Molecular weight
MC	Monte Carlo
MCS	Multiple cloning site
<i>Nef</i>	Negative factor protein
NFV	Nelfinavir
NNRTI	Non-nucleoside reverse transcriptase inhibitor
NMR	Nuclear magnetic resonance
NOE	Nuclear Overhauser effect
NC	Nucleocapsid
NRTI	Nucleos(t)ide reverse transcriptase inhibitor
ORF	Open reading frame
OD	Optical density
<i>Ori</i>	Origin of replication
Phe (F)	Phenylalanine
<i>h</i>	Planck's constant ($6.626068 \times 10^{-34} \text{ m}^2 \cdot \text{kg/s}$)
PAGE	Polyacrylamide gel electrophoresis
PES	Polyethersulfone
<i>Pol</i>	Polymerase
PCR	Polymerase chain reaction
POST	Post-therapy
PRE	Pre-therapy
Pro (P)	Proline
PR	Protease
PI	Protease inhibitor
PDB	Protein Data Bank

<i>e</i>	Proton electric charge ($1.60217653 \times 10^{-19}$ C)
RT	Reverse transcriptase
RNA	Ribonucleic acid
RTV	Ritonavir
SQV	Saquinavir
SCA	Self-consistent analysis
Ser (S)	Serine
SNR	Signal-to-noise ratio
SDSL	Site-directed spin labeling
SEC	Size exclusion chromatography
SDS	Sodium dodecyl sulfate
SU	Surface
Thr (T)	Threonine
T	Thymine
TKR	Tikhonov regularization
TPV	Tipranavir
<i>Tat</i>	Transactivating regulatory protein
TM	Transmembrane
Trp (W)	Tryptophan
Tyr (Y)	Tyrosine
UV-Vis	Ultraviolet-Visible
Val (V)	Valine
<i>Vpr</i>	Viral protein R
<i>Vpu</i>	Viral protein U
<i>Vif</i>	Virion infectivity factor

Abstract of Thesis Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Master of Science

PURIFICATION AND CHARACTERIZATION OF A PRE-THERAPY HIV-1 PROTEASE
VARIANT AND VARIANTS CONTAINING DRUG-PRESSURE SELECTED
MUTATIONS FOR ELECTRON PARAMAGNETIC RESONANCE (EPR) STUDIES

By

Estrella Garlit Gonzales

August 2011

Chair: Gail E. Fanucci
Major: Chemistry

Human Immunodeficiency Virus (HIV) is the infectious agent that causes AIDS (Acquired Immune Deficiency Syndrome). HIV Type 1 protease (HIV-1 PR) is the enzyme essential in the maturation of the virus. Many anti-retroviral drugs target HIV-1 PR to prevent formation of the mature, infectious particle. Recently, protease inhibitors (PI) have been ineffective in the treatment of AIDS because of the emergence of drug-resistant variants of HIV-1 PR. Thus, the structure of HIV-1 protease and its interaction with substrates and inhibitors have been continually studied. Of particular interest is the flap region of HIV-1 PR, in which its flexibility allows it to adopt a variety of conformations and therefore control access of substrate or inhibitor. Moreover, naturally-occurring polymorphisms and drug-pressure selected mutations may have some impact on the flap conformation and flexibility. Several techniques have been used to sample the conformations of HIV-1 PR including pulsed electron paramagnetic resonance (EPR) methods. In this work, double electron-electron resonance (DEER) spectroscopy is used to characterize the flap conformation and flexibility of a

pre-therapy variant of HIV-1 protease and several variants containing drug-pressure selected mutations.

The primary HIV-1 PR variants in this work are referred to as pre-therapy (PRE_i) and post-therapy (POST_i). These constructs were cloned into a vector suitable for bacterial expression of the protease. Mutagenesis was carried out on the POST_i variant to obtain several constructs containing drug-pressure selected mutations. The HIV-1 PR constructs were purified, spin-labeled, and subsequently characterized by sodium dodecylsulfate-polyacrylamide gel electrophoresis (SDS-PAGE), circular dichroism (CD) spectroscopy, mass spectrometry (MS), and continuous wave-electron paramagnetic resonance (CW-EPR) spectroscopy.

DEER analysis was performed on inactive PRE_i, POST_i, POST_i-A82V, B_{si} (stabilized, inactive form of subtype B LAI), and B_{si}-I63P apo constructs. Analysis of DEER data reveals the conformational states of HIV-1 PR corresponding to the tucked/curled, closed, semi-open, and wide-open conformations. Results have shown variations in the most probable distance and predominant conformations of the different HIV-1 PR variants. These results have demonstrated that drug-pressure selected mutations can alter the flap conformation and flexibility of HIV-1 PR. DEER analysis was also done on the PRE_i and POST_i constructs after addition of a protease inhibitor, ritonavir (RTV). Overall, DEER is a powerful technique for monitoring the effect of drug-pressure selected mutations and addition of inhibitor on the flap conformation and flexibility of HIV-1 protease.

CHAPTER 1 INTRODUCTION

Introduction to HIV

Human immunodeficiency virus (HIV) is the infectious particle that causes AIDS (Acquired Immune Deficiency Syndrome), a clinical condition that involves the suppression of the human immune system and progressively leads to opportunistic infections and rare cancers. Since its discovery, the number of people infected with HIV has increased in tremendous rates and AIDS has evolved into a global epidemic. In the last decade, the annual number of new HIV infections has decreased and because of the development of antiretroviral therapy, the number of AIDS-related deaths has been reduced. Although the figures have been steadily declining since the late 1990s, a considerable percentage of the world population is presently living with HIV. In 2009, about 33.3 million people are HIV-infected, 68% of which come from the Sub-Saharan Africa (1).

HIV belongs to the Lentivirus family, a subfamily of the retroviruses, which possess a positive single-stranded messenger RNA (2). Retroviruses rely on a host cell for reproduction of their genomic RNA. Upon infection of a host cell, viral RNA is reverse-transcribed into DNA, integrated into the host cell genome, and replicated by the host cell machinery. Lentiviruses are associated with a disease having a long incubation period and leading to immune suppression.

HIV exists as HIV-1 and HIV-2. Although these two types of HIV share only 40% sequence homology, they have identical morphological and genomic organizations (3). The replication cycle is similar and both viruses lead to immunological failure and similar clinical manifestations upon infection. However, based on epidemiologic, clinical, and

virologic data, HIV-2 is shown to be less virulent than HIV-1. HIV-2 infection is characterized by a long asymptomatic period and lower transmission rate (4).

HIV-1 Structure and Genome

The HIV type 1 (HIV-1) virus, shown in Figure 1-1, has a cone-shaped core containing the viral capsid (CA) protein. The capsid contains two identical RNA strands that are closely associated with the viral RNA-dependent DNA polymerase (Pol), also called reverse transcriptase (RT), and the nucleocapsid (NC) proteins (2). A gp120 external surface (SU) protein connected noncovalently to a gp41 transmembrane (TM) protein and embedded in the lipid bilayer, constitute the envelope glycoproteins, which surround the viral surface. The inner membrane of the virus consists of the matrix (MA) protein that provides integrity to the virus.

The HIV-1 genome, which is shown in Figure 1-2, is about 9.8 kb and consists of open reading frames (ORF) that code for several viral proteins (2). It is a full-length viral mRNA, which is translated into the *gag* and *gag-pol* proteins (2, 5). Table 1-1 summarizes the various viral proteins and their function and localization (6). The three major structural proteins, MA (p17), CA (p24), and NC (p7), are encoded by the *gag* gene while the *pol* gene encodes the viral enzymes. The *gag* precursor p55 produces the smaller proteins, p17, p24, p7, and p6, by proteolytic cleavage. The *gag-pol* precursor protein is cleaved into products consisting of the reverse transcriptase (RT), protease (PR), and integrase (IN) proteins. PR processes the *gag* and *gag-pol* polyproteins and IN is involved in virus integration. The *env* gene produces the envelope proteins, which are vital for the recognition and binding of virus to the host cell. The envelope gp120 and gp41 are made from precursor gp160, which is a mono-spliced RNA from the full-length viral mRNA. *Tat* and *rev* code for viral regulatory

proteins that are involved in upregulating HIV replication and enhancing expression of viral proteins, respectively. *Nef* encodes a negative factor protein, which serves multiple functions including viral suppression and downregulation of CD4 and MHC Class I molecules. The genes, *vif*, *vpr*, and *vpu* produce accessory proteins, which promote viral replication, assembly, and budding; viral infectivity; degradation of CD4 cells; and the overall suppression of the cell immune response (2).

HIV-1 Life Cycle

Figure 1-3 illustrates the steps involved in viral replication and maturation of HIV-1 (7). The cycle begins with the attachment of the mature virion to the host cell by binding of gp120 to the CD4 receptor (Step 1). After fusion (Step 2), the capsid is uncoated and the virus releases the RNA, reverse transcriptase, integrase, and other viral proteins into the host cell cytoplasm. The genomic RNA is then reverse transcribed by RT into a double-stranded DNA (Step 3). The viral DNA is transported across the nucleus and integrated into the host cell genome by IN (Step 4). The HIV DNA formed is referred to as the provirus, which can be inactive for several years. Upon activation of proviral DNA, RNA polymerase in the host cell produces copies of the HIV genetic material and shorter messenger RNA serve as the template for translation into the viral proteins (Step 5). The viral proteins and RNA then assemble into an immature virus. In step 6, a budding event occurs, wherein the virus exits and uses the cell surface as an envelope to form its outer membrane. Eventually, the polyproteins within the non-infectious virus are cleaved by HIV protease and reorganized to form the mature virus (Step 7). Figure 1-4 shows the structural assembly of the immature and mature HIV-1.

Introduction to HIV-1 Protease

The role of HIV-1 protease (PR) in the viral life cycle is to form a mature, infectious virus by post-translational processing of the *gag* and *gag-pol* polyproteins. Anti-retroviral drugs have been used against HIV-1 to prevent formation of the mature, infectious particle. In particular, HIV-1 protease inhibitors (PI) have been shown to inhibit post-translational processing, block HIV-1 replication, and reduce viral load (8). Despite their potency, anti-retroviral drugs have become ineffective against emerging drug-resistant variants of HIV-1 PR (9). Thus, there is the continuing need to examine the structure of HIV-1 PR and its interactions with various substrates and inhibitors in order to design more effective drugs.

Structure and Function

Human immunodeficiency virus type 1 (HIV-1) protease (EC 3.4.23.16) is a member of the aspartic protease family (10). It is composed of two noncovalently associated, structurally identical monomers, where each monomer consists of 99 amino acids (11). In retroviral proteases, each monomer is generally composed of four structural elements: A1 and A2 loop, B1 and B2 loop, C1 and C2 α -helix, and the D1 and D2 loop, as indicated in Figure 1-5A (12-15). In HIV-1 protease, all of these elements are present except that a loop replaces the α helix C1. The B1 loop contains the catalytic aspartic acid. The D1 loops, also referred to as “flaps”, consist of β -hairpins, which are functionally important because of their flexibility and involvement in the binding of substrate or inhibitor. The “hinge” corresponds to the region that links the helix C1 and the flap and the region between the flap and A2. The dimerization domain consists of a four-stranded β -sheet interface that is formed by the amino and

carboxyl termini of both monomers. A typical X-ray crystal structure of HIV-1 protease is shown in Figure 1-5B, in which the key regions are labeled.

To date, there are over 400 structures deposited in the Protein Data Bank (PDB) (<http://www.pdb.org>) including structures of the apoenzyme and substrate- or inhibitor-bound HIV-1 PR. The active site region of the protease is primarily formed by residues 25-32, 47-53, and 80-84, in which Asp25 and Asp25' are the major catalytic residues indicated in Figure 1-5B (16). The conserved active site residues include the catalytic triad Asp25-Thr26-Gly27, which form a loop and stabilized by a network of hydrogen bonds (18). The O δ 1 atoms in the Asp25/25' carboxylate groups interact closely and form a nearly co-planar arrangement. The hydrogen bonding network forming the interaction between Asp25, Thr26, and Gly27 is referred to as the “fireman’s grip,” as illustrated in Figure 1-6. The O γ 1 atom of each Thr26 is protonated by the amide group of the opposite Thr26 residue and deprotonated by the carboxyl group of residue 24 in the opposite loop (14).

Structures of HIV-1 protease were determined crystallographically in various conformations. The first X-ray crystal structures of the apoenzymes were reported at 2.7 to 3.0 Å resolution (19-22). In most of the crystal structures without ligand, HIV-1 protease is usually found in an open conformation, in which the flaps are slightly oriented away from the active site (15). The binding of substrate or inhibitor requires a substantial movement of the flaps to form into a closed conformation (23, 24). The flaps shift by as much as 7 Å and the subunits rotate by about 2° around the axis along the subunit β -sheet interface (14, 25-27). In the closed conformation, the flap residues are interacting strongly with substrate or inhibitor and therefore access to the active site is

blocked. The closed conformation of HIV-1 PR, shown in Figure 1-7, illustrates the inaccessibility of the active site to substrate or inhibitor.

Conformational Sampling of HIV-1 Protease

HIV-1 protease undergoes conformational changes due to the intrinsic flexibility of its flaps (23, 24). The flaps of unliganded HIV-1 PR are described as having an ensemble of various conformational states in dynamic equilibrium, which include predominantly semi-open states and small populations of closed and wide-open conformations (28). Studies on the crystal structures of HIV-1 PR and other retroviral proteases have revealed that the semi-open conformation is the most thermodynamically stable (14). Various other techniques have been used to describe the different HIV-1 PR flap conformations including nuclear magnetic resonance (NMR) spectroscopy, isothermal titration calorimetry (ITC), and molecular dynamics (MD) simulations (28-43).

NMR studies have been performed on free and bound HIV-1 protease to study flap dynamics (28-32). NMR relaxation experiments have shown that transverse relaxation rates of backbone amides of HIV-1 protease are relatively higher in the flap region of the free protease (28). This indicates that without substrate/inhibitor, the HIV-1 PR flaps are relatively flexible in the millisecond-microsecond timescale corresponding to conformational exchange. In the inhibitor-bound protease, the flaps, with the exception of residues 50 and 51, are rigid in solution (43). A more recent study by Freedberg et al. has shown relaxation and nuclear Overhauser effect (NOE) data indicating that flaps are flexible in the sub-nanosecond (ns) timescale, which is faster than overall tumbling of protein (29). The sub-ns fluctuations of the flap tips (residues 49-53) reflect a

dynamic equilibrium between the ensemble of semi-open conformational states and possibly including minor populations of the closed and wide-open conformations.

Differential scanning calorimetry (DSC) and isothermal titration calorimetry (ITC) have been used to study the stability of HIV-1 protease and monitor structural changes upon binding of ligand. The combination of DSC and ITC provides an accurate means of determining thermodynamic binding constants of protease-inhibitor complexes. DSC and ITC studies have revealed that the dimeric structure of HIV-1 PR is stabilized by inhibitor binding in an entropically-driven process (33-36). Structure-based analysis of binding energetics have shown that residues in the active site and the flap region have the greatest van der Waals contacts with the inhibitor and therefore contribute more to the energetics of binding (34). Todd et al. examined the extent of stabilization upon inhibitor binding by determining the residue stability constant, which is a statistical measure of the probability of a residue to be either in a folded or unfolded state (33, 34). It has been established that the central part of the protease, which includes the dimer interface and adjacent areas, defines the most stable region of the protease. Meanwhile, the flap region corresponds to the most unstable part of the protein and upon inhibitor binding, the flaps are forced into a conformation which allows direct interaction with the inhibitor.

Unlike X-ray, which reveals only a single protein conformation, and NMR, which detects the average conformation, molecular dynamics (MD) simulation experiments sample the flap closing and opening events in HIV-1 PR (37). MD simulation studies were the first to provide a structure for the wide-open conformation (38, 39). Hornak et al. performed simulations on ligand-free HIV-1 PR using a low-viscosity implicit solvent

to characterize the transition from a closed to semi-open conformation (37). The simulations reveal that the unliganded HIV-1 PR adopts a semi-open flap conformation and transiently forms into the closed conformation with reversal of flap handedness, as seen in Figure 1-8. Several flap opening events were observed, including the wide-open conformation, in which the distance between Ile50/50' reached to about 30 Å. Such open state exists transiently and reverts back to the predominant semi-open form. Hornak et al. also studied the effect of manual docking of inhibitor into the open state of HIV-1 PR and they observed spontaneous conversion of the protease into the closed form (40). X-ray crystal structures and other MD simulation studies have shown that flaps can adopt a curled or tucked conformation, which constitute only a small fraction of the conformational ensemble of HIV-1 PR (41, 42, 44).

Recently, Double Electron-Electron Resonance (DEER) spectroscopy has been utilized by our group to examine flap conformations of different HIV-1 PR variants (45-51). DEER results obtained by our group will be described in further detail in Chapter 5.

HIV-1 Protease Cleavage Sites and Substrate Recognition

HIV-1 protease binds to at least ten natural substrates, whose amino acid sequences are given in Table 1-2. The substrates do not have sequence homology and are asymmetric around the cleavage sites, in both size and charge distribution (52). According to the nomenclature of Schechter and Berger (53), the subsites within the substrate/inhibitor are represented as P4-P3-P2-P1/P1'-P2'-P3'-P4', where the slash represents the scissile bond or cleavage site. The residues on the N-terminal side of the scissile bond are unprimed while the C-terminal residues are primed (14). The corresponding binding pockets in the enzyme are labeled S1 and S1', S2 and S2', S3 and S3', and S4 and S4'.

Figure 1-9 shows an illustration of *gag* and *gag-pol* polyproteins and the corresponding cleavage sites. The order in which HIV-1 protease recognizes and cleaves *gag* and *gag-pol* polyproteins is not well-established (14). However, it is known that the initial cleavage occurs between p6^{*gag-pol*} and PR (site VIII), which is essential for the release of PR (54). There is some evidence that the first target of HIV-1 PR is located on the p2/NC (site III) (55, 56). The last and slowest step is the cleavage of the CA/p2 (site II) (57, 58). Sequence variability, within the recognition sites, may affect rates of cleavage and impact protease activity (59).

The mechanism of enzyme action of HIV-1 PR, commonly called the “push–pull” mechanism, is described by a classic acid–base catalysis (60). It involves a water molecule that protonates the carboxyl groups of Asp25 in both monomers. A proton from the carboxyl dyad is then transferred to the P1 carbonyl oxygen of the substrate, where the peptide bond is eventually cleaved. As shown in Figure 1-10, the mechanism involves a tetrahedral intermediate, which is mimicked in most of the currently used licensed protease inhibitors.

In HIV-1 protease, residues 8, 23, 25, 27-30, 32, 47-50, 53, 76, 80-82 and 84 form the substrate binding pocket, as depicted in Figure 1-11 (61). Figure 1-11B shows the specific residues in the protease that interact with the substrate subsites. Substrate binds to the protease in an extended β -strand conformation, wherein several hydrogen bonds form between the backbones of substrate and protease to facilitate binding (52). The sites between the P2–P2' positions on the substrate directly bridge the enzyme active site. The S2/S2' interior pocket, which comprises residues 28-30, 32, 47, 49-50, 76, and 84, is smaller compared to the other binding sites and is known to be specific

(14). This limits the size and type of residue at P2/P2' that can bind the pocket (62, 63). Like substrates, protease inhibitors were designed so that the P2–P2' residues form strong interactions with the enzyme binding site (14). The emergence of a mutation in the active site can disrupt inhibitor binding at independent sites between the P2–P2' region (52). This led Schiffer et al. to structurally examine substrate recognition by HIV-1 protease. Their structural studies have shown that most of the substrates within the active site pocket adopt a toroid shape on the unprimed side (P1 to P3 region) and an extended conformation on the primed side, which is referred to as the “substrate envelope” (52). From this finding, they have suggested that the enzyme recognizes substrates based on shape and accessibility rather than the amino acid sequence. They also proposed that inhibitors should adopt the “substrate envelope” to fit well within the active site region.

HIV-1 Protease Inhibitors

There is no definitive cure or vaccine against AIDS; however, HIV infection can be treated using clinically licensed drugs (16, 61). Currently, there are 25 anti-retroviral drugs in the market that target different steps in the HIV-1 life cycle (Figure 1-12). These are 1) coreceptor antagonists and fusion inhibitors, which target viral entry, 2) nucleos(t)ide and non-nucleoside reverse transcriptase (RT) inhibitors (NRTI and NNRTI), which block reverse transcription, 3) integrase inhibitors, which prevent integration of viral DNA into the host-cell genome, and 4) protease inhibitors (PI), which hinder viral maturation (61). Zidovudine or AZT (β -D-(+)-3'-azido-3'-deoxythymidine), an RT inhibitor released in 1989, was the first drug to be approved for the treatment of AIDS and HIV infection. The current regimen for HIV infection uses highly-active

anti-retroviral therapy (HAART), which consists of a combination of licensed NRTI/NNRTI and/or PI.

HIV-1 PR inhibitors were originally designed based on classical substrates or transition-state analogs and are modeled from the peptidomimetic inhibitors for the related aspartic protease, renin. However, mechanism-based screening of renin inhibitors and peptide mixtures has resulted in limited success in identifying HIV-1 PR inhibitors. “Structure-based drug design” emerged as a better strategy, and HIV-1 protease was the first therapeutic target (64). The protease inhibitors are generally designed such that the P2–P2’ region fits the active site. Most of them are hydrophobic, with a cyclical side chain at P1 and a fairly bulky functional group at P1’ (60).

To date, there are ten protease inhibitors approved by the United States Food and Drug Administration (FDA), namely saquinavir (SQV), ritonavir (RTV), indinavir (IDV), nelfinavir (NFV), amprenavir (APV), fosamprenavir (FPV), lopinavir (LPV), atazanavir (ATV), tipranavir (TPV), and darunavir (DRV) (60, 65, 66). The chemical structures are shown in Figure 1-13. All are competitive active site inhibitors that bind with affinities to the purified enzyme ranging from low nanomolar to picomolar (60). With the exception of tipranivir and darunavir, protease inhibitors are peptidomimetics, which were designed based on the transition state formed as a result of protease-substrate interaction (Figure 1-10). They are composed of non-cleavable, dipeptide isosteres as core scaffolds and commonly contain a secondary hydroxyl group, which substitutes the P1 carbonyl moiety of substrates (16). The hydroxyl at the P1 site of the inhibitor is essential for tight binding with the protease by forming interactions with the catalytic Asp25/25’ residues. Another common feature in the protease-inhibitor complex is a

conserved water molecule that mediates contacts between the P2 and P1' carbonyl oxygen atoms of the inhibitors and the Ile50/50' amide groups of the enzyme (27).

HIV-1 Protease Variants and Subtype Polymorphisms

Given the relatively small size of HIV-1 protease, the sequence variability in HIV-1 protease is presumably limited (67). However, recent data from the Los Alamos HIV database reveals an abundance of gene sequences derived from clinical isolates (68). The variability in the genetic sequence of HIV-1 protease is primarily due to the high mutation rate resulting from the lack of proofreading ability of viral reverse transcriptase. In reverse transcription, at least one nucleotide substitution occurs for every three rounds of replication ($\sim 10^{-4}$ to 10^{-5} mutations per nucleotide per cycle of replication), on average (69). Other factors of genetic variation include the high *in vivo* rate of HIV replication, high frequency of recombination, and the accumulation of variants during the course of infection (11). The HIV-1 virus may be subjected to additional selective pressures, such as anti-retroviral therapy, that contribute to genetic differences between isolates within an individual and between individuals (15). There is no prototype or consensus wild-type HIV-1 protease but rather a complex mixture of related sequences (55).

HIV-1 protease sequences that arise from naturally occurring polymorphisms are classified into groups, subtypes (or clades), and circulating recombinant forms (CRFs), as shown in Figure 1-14 (70). Groups refer to the viral lineage. The most common is Group M (major), which comprises at least nine distinct subtypes: A-D, F-H, J, and K, and several CRFs. The subtypes refer to the taxonomic groups within a particular lineage whereas CRFs refer to various recombinant forms of the virus that arise from

the genetic combination of two viral subtypes (68). Group O viral strains can be found in west-central Africa and groups N and P, which were discovered in Cameroon, are extremely rare (11).

Historically, the most common HIV-1 subtype is B, which is prevalent in North America, Western Europe, and Australia (11). The subtype B isolates have been widely studied for drug screening and susceptibility testing. However, they account for only a small portion of HIV-1 isolates worldwide. The majority of HIV-1 infection-related cases, which occur in the sub-Saharan Africa, is primarily caused by subtype C HIV-1 (1, 11). Recently, other subtypes have become more frequent and account for about 25% of HIV-infected patients, particularly in Europe (71). The Los Alamos HIV database contains several sequences of the subtype B viral genome (68). The most widely studied sequence is HXB2 (Accession Number K03455), a specific clone from the French isolate LAI, which does not have any identical match to sequences of other patient-derived isolates (72). The use of HXB2 (or LAI) as a consensus reference sequence is based on historical precedence (73).

HIV-1 Protease Drug Resistant Variants

HIV-1 protease drug-resistant variants evolve from the selective pressure of protease inhibitor (PI) therapy (15). The emergence of these variants has led to the failure of most anti-retroviral drugs (9). Almost half of the amino acid residues of these drug-resistant variants arise from the accumulation and persistence of multiple mutations, which lead to drug resistance (16, 74-76). Drug-resistant variants have reduced sensitivity and affinity to protease inhibitors (PI), but still maintain their ability to bind and process natural substrates (16).

The Stanford HIV database maintains a collection of the sequences of various HIV-1 protease isolates (74). A comparison of the HIV-1 PR subtype B sequences from PI-treated and untreated (PI naïve) patients reveals variations in several conserved regions of the protease (Figure 1-15). This suggests that exposure to inhibitor contributes to the prevalence of mutations in HIV-1 PR in sites that are functionally important in substrate binding and catalysis, such as the active site and flap regions. There is also a noticeable increase in the mutation prevalence in sites outside of the functional regions, particularly at positions 10, 36, 62, 63, 71, and 90.

The International AIDS Society-USA group has reported data that have been published or presented at scientific meetings on mutations in HIV-1 PR associated with PI resistance (65). The mutations are classified as either primary or secondary mutations. A recent review lists 15 primary mutations (30, 32, 46, 47, 48, 50, 54, 58, 74, 76, 82, 83, 84, 88, and 90), shown in Table 1-3, that are associated with PI resistance (77). Based on the list, the most common mutations that arise from the majority of inhibitors can be found at positions 10, 46, 54, 82, and 90.

Primary or active site mutations, which are the main contact residues for drug binding, cause resistance by reducing susceptibility of protease to an inhibitor (16, 61, 65, 77). These mutations lead to inefficient processing of the enzyme and subsequent decrease in the replicative capacity of the virus (61). The primary mutations, which appear frequently in HIV-1 PR drug-resistant variants, include D30N, G48V, I50V/L, V82A, I84V, and L90M (16, 78). They usually emerge in combination with each other and other sites within and outside the enzyme active site. Studies have revealed that the D30N/L90M mutation is associated with nelfinavir resistance while the G48V/L90M

mutation arises as a result of saquinavir therapy (61, 79, 80). Mutations at I50L/V and I84V are usually common in atazanavir, fosamprenavir, and darunavir (77). I84V is also found as a major mutation in indinavir, ritonavir, saquinavir, and amprenavir therapy and may develop with the mutation L90M (58, 77). The most common V82A mutation is observed predominantly in HIV-1 isolates from patients receiving treatment with ritonavir as well as in combination with indinavir and lopinavir (58, 81). In cases where ritonavir is used as a pharmacological booster for other PIs, the mutations that are selected are the same as with the unboosted protease inhibitor (Table 1-3) (77).

Secondary or compensatory mutations, which can be present before therapy or appear in response to inhibitors, are typically found outside the active site. They do not have a substantial effect on the resistance phenotype but they help restore the replicative capacity of the virus (61, 65, 77). For instance, the N88D mutation arises as a result of the nelfinavir resistance-associated D30N/L90M mutation to compensate for the impaired viral replication (79). They also compensate for impaired protease function by increasing the catalytic efficiency and thermodynamic stability of the enzyme (82-86). The study by Clemente et al. has shown that the combination of non-active site mutations M36I and A71V increased the enzymatic efficiency of a mutant containing the D30N primary mutation (84). Mahalingam et al. has demonstrated that mutants having the K45I and N88D secondary mutations increased the stability of protease, which may compensate for the lower activity and result in enhanced viral replication (85, 86).

Several secondary mutations exist as common polymorphisms in drug-naïve patients. A number of studies have revealed that the sequences of isolates obtained from untreated patients possess amino acid substitutions associated with inhibitor

resistance (87). Some of these natural polymorphisms are found at positions 10, 36, 46, 63, 71, 77 and 82 (67). Based on the Stanford HIV database, the mutation at position 63 is a common amino acid polymorphism, which is found in over 50% of the HIV-1 PR subtype B isolates (Figure 1-14) (74). The L63P mutation occurs in multiple combinations with other resistance mutations to compensate for reduced viral fitness (16, 61, 67, 87). It has been demonstrated to improve viral replication in the presence of double mutations at positions 82 and 84 and therefore compensate for the deleterious effects of these mutations. The L63P mutation may stabilize the protease and influence flap motion (87, 88).

Mechanisms of drug resistance

Mutations in HIV-1 protease can decrease the incorporation and binding affinity of an inhibitor (16, 61). The altered binding is primarily due to the inherent plasticity of HIV-1 PR, which means that the protease can adjust by rearranging interactions not only within the region of the mutated residue but throughout the enzyme (16, 65, 89-92). Active site mutations alter the shape of the binding pocket and therefore directly impact the protease-inhibitor interaction (23). Schiffer et al. suggested that an inhibitor, which protrudes beyond the substrate envelope, causes mutations in specific HIV-1 protease residues that it directly contacts (93). The primary mutations that have been shown to interact with inhibitor atoms and therefore confer drug resistance are D30, I47, G48, I50, V82, and I84.

The mechanism by which non-active site mutations alter inhibitor binding remains elusive although it has been suggested that these mutations confer drug resistance through an indirect mechanism (16, 93). Rose et al. proposed that non-active site mutations affect protein flexibility by either preventing the flaps from adopting the wide-

open conformation or decreasing the rate of flap closure so that the protease preferentially binds the substrate over the inhibitor (23). The study by Schiffer et al. concurred that the change in conformational flexibility as a result of the mutations can be explained by the hydrophobic sliding mechanism (92). Each monomer of HIV-1 protease consists of 19 hydrophobic residues outside the active site, which comprise the hydrophobic core. Molecular dynamics simulation of unliganded HIV-1 protease demonstrated that the hydrophobic residues alter their side-chain conformations and exchange van der Waal contacts with other residues. The exchange of hydrophobic interactions resulting from the sliding of hydrophobic residues past one another is referred to as the hydrophobic sliding mechanism. This process, which requires only minimal energy expenditure, can be correlated with changes in the dynamics of the protease. Thus, the conformational changes associated with non-active site secondary mutations can be explained by the extensive rearrangement of the hydrophobic core, which detrimentally impacts inhibitor binding (92).

Influence of protease cleavage site mutations on drug resistance

The sequence of substrate cleavage sites can coevolve with the protease to attain drug resistance without compromising enzyme function (93). A number of studies have reported the association between specific mutations in the protease and cleavage sites, such as in NC/p1 and p1/p6, which alters the susceptibility of HIV-1 protease to various protease inhibitors (16, 94-98). One study reported that p1/p6 cleavage site mutations are associated with the NFV-resistant D30N/N88D protease mutations (16, 99). The cleavage site amino acid changes have been demonstrated to compensate for the negative effects of the D30N/N88D by improving the replicative capacity and processing of substrate. In another study, it has been determined that substitutions in the *gag*

NC/p1 cleavage site (A431V, K436E and/or I437V/T), without any alterations in the protease sequence, were selected during PI exposure (65, 99). Additionally, Goodenow et al. (59) studied the association of *gag-pol* mutations with protease activity and Ho et al. (100) established a correlation between drug-associated changes in the *gag-pol* amino acid residues and viral replication and drug response. The findings from these studies suggest that the sequence of the cleavage site can undergo changes to compensate for the decreased interactions between mutant protease and substrate and therefore improve substrate binding and enzyme catalytic activity (65). In addition to the emergence of protease resistance to protease inhibitors, prolonged exposure to PI may lead to evolution of protease cleavage sites to maintain viral fitness (16).

Scope of Work

Details of HIV-1 PR Constructs

Naming and abbreviations of constructs (e.g. B_{si}) used in this work are described as follows. A subscript “s” following the subtype or variant of HIV-1 protease stands for stabilization against autoproteolysis. In subtype B HIV-1 PR, autoproteolytic cleavage has been found to occur between positions 6 and 7, 33 and 34, and 63 and 64. The Q7K mutation carried out by Rose et al. reduced the rate of autoproteolysis by more than a 100-fold (101). Meanwhile, the mutant protease constructed by Mildner et al., which consists of L33I and L63I substitutions, increased the stability of the HIV-1 protease with specificity and kinetic properties similar to the wild-type enzyme (102). In this work, the Q7K, L33I and L63I mutations were engineered into the subtype B HIV-1 PR construct to increase the stability of the enzyme and slow down autoproteolysis of active protease. The subscript “i” refers to inactive protease, in which a D25N mutation was incorporated into the sequence. The naturally occurring cysteine residues (C67A

and C95A) were substituted with alanine for the purpose of site-directed spin-labeling and to prevent non-specific disulfide bonding. All constructs for electron paramagnetic resonance (EPR) spectroscopy studies were engineered with a K55C mutation, which will be described in further detail in Chapter 5. In this work, the B_{SI} construct, which is the inactive form of subtype B LAI and contains the stabilizing mutations (Q7K, L33I, and L63I), is used as the reference construct.

The primary constructs used in this work are variants of subtype B HIV-1 PR. The sequences were derived from gag-pol alleles obtained from a pediatric subject (D1) infected with HIV by maternal transmission (59). The gag-pol alleles were isolated from serial blood samples obtained over 7 years before therapy initiation (pre-therapy) and after the development of multiple drug resistance following 77 weeks of initial combination protease inhibitor (PI) therapy including ritonavir (RTV) and an additional 16 weeks of treatment with indinavir (IDV) (post-therapy) (100). The pre-therapy and post-therapy protease sequences will be referred to as PRE and POST, respectively. Table 1-4 shows the sequence alignment of the PRE_i and POST_i protease constructs with the reference construct B_{SI} and multi-drug resistant constructs, V6_i (90) and MDR769_i (103), which were also obtained from patient isolates. The crystal structures in Figure 1-16 show that PRE_i contains four amino acid substitutions relative to subtype B LAI, excluding the D25N, C67A, and C95A mutations. POST_i consists of mutations in the active site (V82A), near-active site (L10I and I15V), flap (I54A and Q58E), and hinge (E34Q, M36I, and T37N) regions, some of which are also found in V6 and MDR769.

Previous Work on HIV-1 PR Constructs

In the presence of combination protease inhibitor (PI) anti-retroviral therapy, mutations in *gag* and *gag-pol* develop in viruses *in vivo* during replication. Ho et al.

investigated the pre-therapy and post-therapy *gag-pol* polyprotein and protease constructs to assess the influence of amino acid changes in *gag* and PR on viral fitness and drug resistance (100). To reveal which particular amino acid substitutions contributes to a change in viral replicative capacity and PI sensitivity, specific sites on the post-therapy *gag* and protease sequence were back-mutated into the pre-therapy (PRE) sequence. Each mutant consists of mutations at specific sites in the PR sequence and *gag* mutations at p2 (V376I) and p7 (V398E). The specific mutations are summarized in Table 1-5. The amino acids at position 487 (p6^{*gag*}), 494 (p6^{*gag*}) and 437 (p6^{*gag-pol*}) remain the same as in the post-therapy (POST) sequence.

The first mutation was carried out in the active site of the PR sequence wherein post-therapy alanine was converted to pre-therapy valine (A82V) (100). This resulted to a recovery of about 50% replicative capacity and an increase in sensitivity to either PI inhibitors, RTV or IDV, about half relative to pre-therapy virus. Mutation of the flap region, particularly at residues 54 and 58 (A54I, E58Q), improved viral replication and increased sensitivity to RTV. However, this virus was more resistant to IDV compared with the post-therapy virus. The third set of mutations involves a combination of active site (A82V) and near-active site (I10L, V15I) mutations. The final mutant consists of a combination of mutations in the flap (A54I, E58Q) and hinge (Q34E, I36M, N37T) region. The viruses containing the latter sets of mutations failed to restore replicative capacity to the post-therapy virus. However, these viruses showed reduced resistance to PI with levels similar to the pre-therapy virus. These results indicate that replication fitness does not necessarily correlate with susceptibility to PI (100). The findings of the study also reveal that *gag* can be a dominant modulator of protease resistance

phenotype by varying susceptibility of PR to inhibitors. Mutations in the *gag* p2/p7 cleavage site and p7 failed to fully recover viral replicative capacity but changes in these positions modulated response to PI.

Objectives of Work

The functionally important region of the HIV-1 protease that has drawn much attention is the flap region. The flexibility of the flaps allows control of substrate or inhibitor access to the active site (23, 24). It is believed that either naturally occurring polymorphisms or drug-pressure selected mutations in the protease can alter flap conformation and flexibility and therefore influence the accessibility of a substrate or inhibitor (83-85, 90, 104, 105). Thus, further study on the effect of mutations on flap conformation and flexibility is necessary. To characterize the flap conformations, spin labels were incorporated on specific sites on each flap using the site-directed spin labeling (SDSL) technique and the distance between the spin labels was measured using a pulsed electron paramagnetic resonance (EPR) method, called Double Electron-Electron Resonance (DEER) spectroscopy.

The overall goal of this work is to purify and characterize inactive subtype B HIV-1 PR variants for analysis by DEER spectroscopy. The previously described constructs, PRE_i, POST_i, and POST_i mutants containing the amino acid substitutions listed in Table 1-5, were studied. These constructs were characterized using the following techniques: SDS-PAGE, circular dichroism (CD) spectroscopy, mass spectrometry (MS), and continuous wave-electron paramagnetic resonance (CW-EPR) spectroscopy. DEER analysis was carried out on PRE_i, POST_i, POST_i-A82V, B_{si}, and B_{si}-I63P apo constructs to compare the flap conformations and determine whether the conformational changes are associated with drug-pressure selected mutations. The B_{si}-I63P mutant was

constructed to detect a conformational change resulting from a single amino acid substitution. The effect of the inhibitor, ritonavir (RTV), on the flap conformations of PRE_i and POST_i was also investigated.

Summary

This chapter has provided an introduction to HIV and HIV-1 protease. The structure and function of HIV-1 PR was described in detail, particularly focusing on the various conformations of the protease and its correlation with enzyme function. The occurrence of different variants of HIV-1 PR and the basis for the existence of naturally occurring polymorphisms and drug-pressure selected mutations was presented. The HIV-1 PR constructs studied in this work has been described.

The succeeding chapters will present the techniques and methods used and the results obtained in this work. Chapter 2 will provide a background on the various techniques used. In Chapter 3, the methodology and results for cloning and mutagenesis of all HIV-1 PR constructs will be presented. Chapter 4 will present the methods and results for the purification and characterization of all HIV-1 PR constructs. Chapter 5 will provide a description of the experimental design and methodology used for analyzing HIV-1 PR by DEER spectroscopy. In addition, previous DEER work on HIV-1 PR done by our group will be described. Results of DEER analysis on particular HIV-1 PR constructs will also be presented.

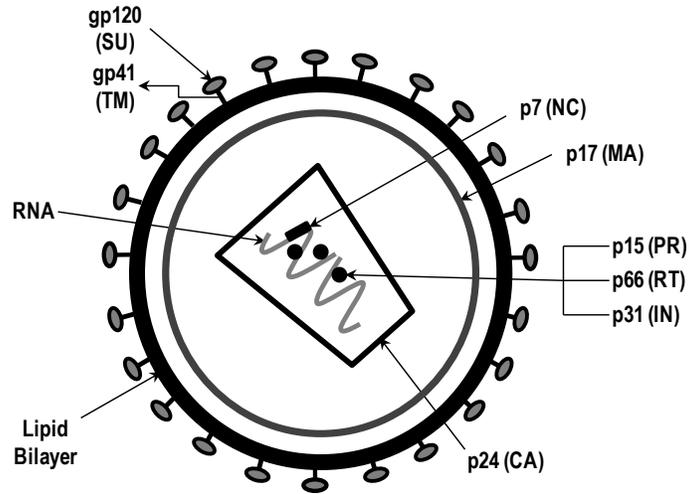


Figure 1-1. The assembly of viral proteins and other constituents in HIV-1. Figure adapted from Levy (2).

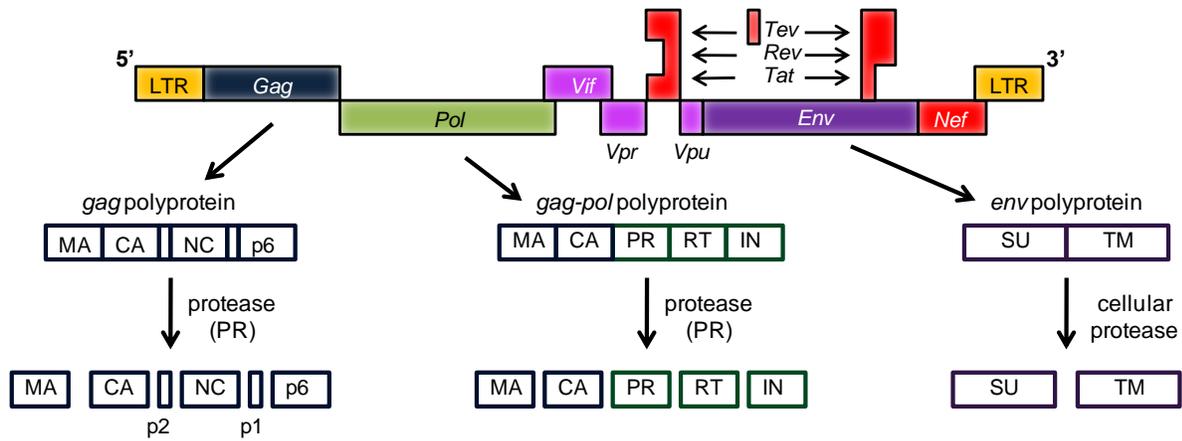


Figure 1-2. Schematic diagram of the HIV-1 genome containing the genes that code for several polyproteins including *gag* and *gag-pol*. Figure adapted from Levy (5).

Table 1-1. Various proteins expressed by the HIV-1 genes.

NAME	SIZE	FUNCTION	LOCALIZATION
<i>Gag</i>			
MA	p17	Membrane anchoring; <i>Env</i> interaction; nuclear transport of viral core (myristylated protein)	Virion
CA	p24	Core capsid	Virion
NC	p7	Nucleocapsid, binds RNA	Virion
<i>Pol</i>			
Protease (PR)	p15	<i>Gag</i> and <i>gag-pol</i> cleavage and maturation	Virion
Reverse Transcriptase (RT)	p66	Reverse transcription	Virion
RNase H	p51	RNase H activity	Virion
Integrase (IN)	p31	DNA provirus integration	Virion
<i>Env</i>	gp120/gp41	External viral glycoproteins bind to CD4 and secondary receptors	Plasma membrane, virion envelope
<i>Tat</i>	p16/p14	Viral transcriptional transactivator	Primarily in nucleolus/nucleus
<i>Rev</i>	p19	Regulates viral mRNA expression; RNA transport, stability and utilization factor (phosphoprotein)	Primarily in nucleolus/nucleus shuttling between nucleolus and cytoplasm
<i>Vif</i>	p23	Promotes virion maturation and infectivity	Cytoplasm (cytosol, membranes), virion
<i>Vpr</i>	p10-15	Virus replication; Promotes nuclear localization of pre-integration complex, inhibits cell division, arrests infected cells at G2/M stages in cell cycle	Virion/nucleus (possibly nuclear membrane)
<i>Vpu</i>	p16	Promotes extracellular release of viral particles; degrades CD4 in the ER	Integral membrane protein
<i>Nef</i>	p25-p27	Virus suppression; CD4 and MHC class I down-regulation (myristylated protein)	Plasma membrane, cytoplasm (possibly virion)

Table adapted from HIV Sequence Compendium 2010 (6).

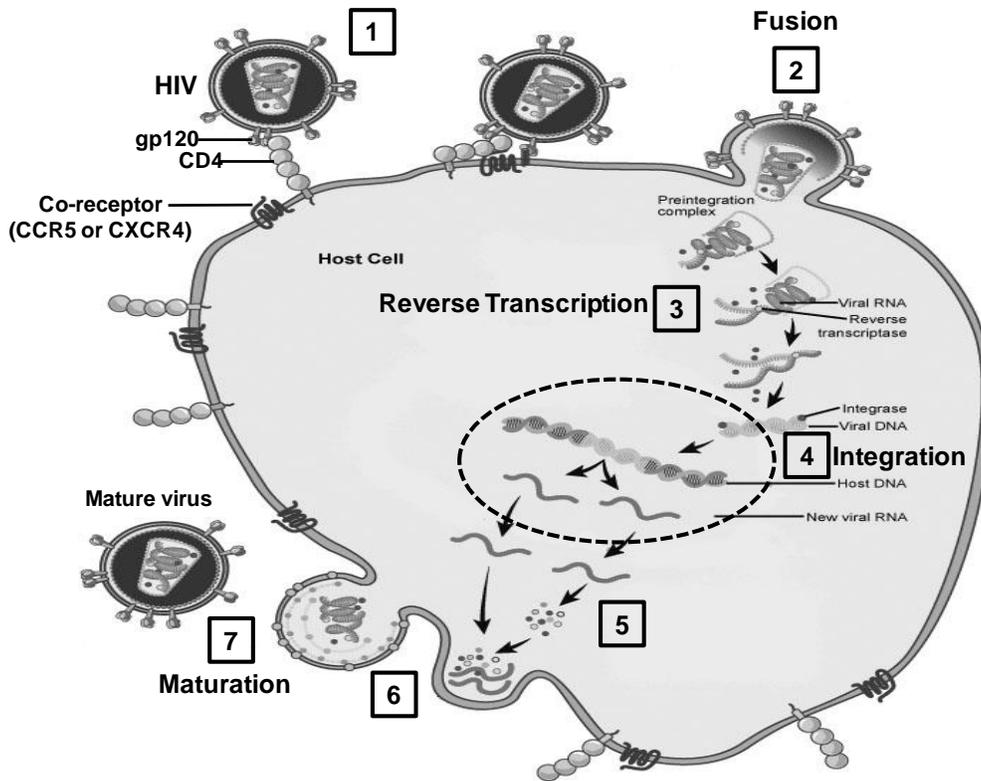


Figure 1-3. Viral life cycle of HIV-1. Numbered steps are described in the text in detail. Figure adapted from National Institute of Allergy and Infectious Diseases (<http://www.niaid.nih.gov>).

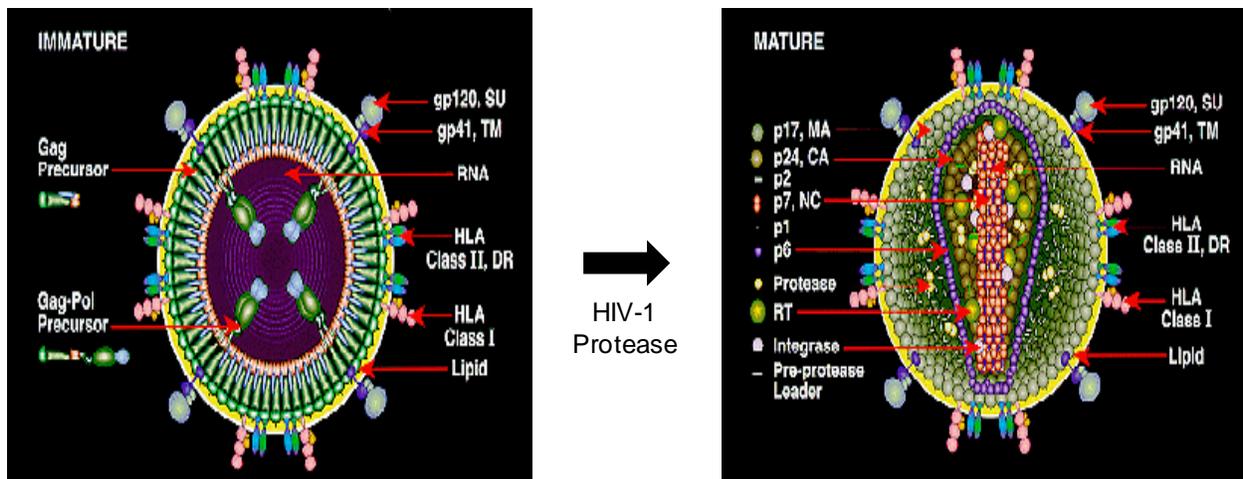


Figure 1-4. Structural assembly of HIV-1 from an immature to a mature virus. Image courtesy of National Institute of Allergy and Infectious Diseases (<http://www.niaid.nih.gov>).

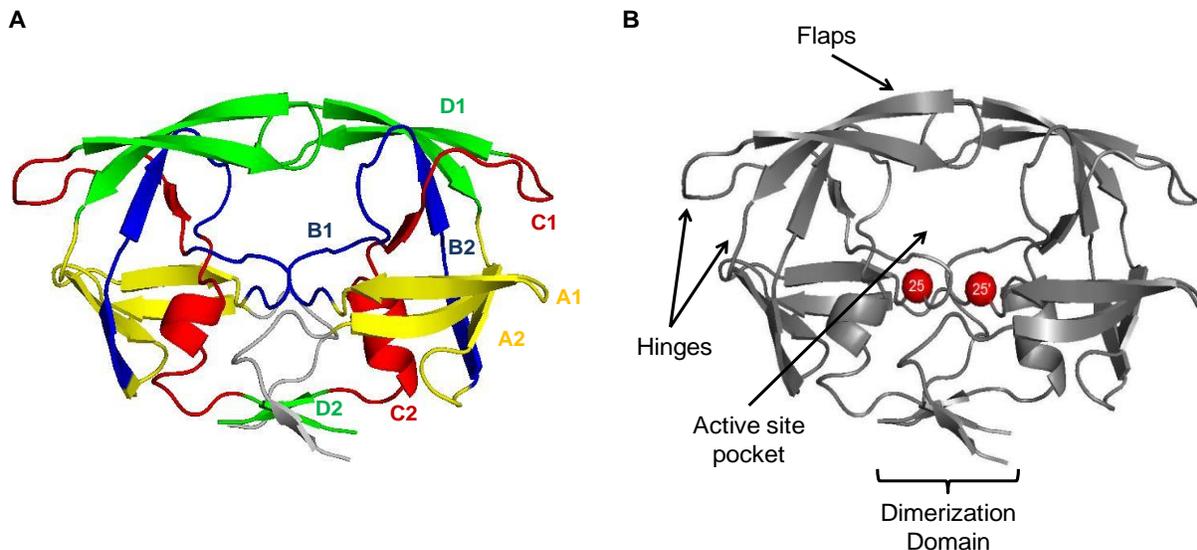


Figure 1-5. Ribbon diagrams of HIV-1 PR (PDB ID 2BPX) highlighting the A) four structural elements, A1 and A2 loop (yellow), B1 and B2 loop (blue), C1 loop and C2 α -helix (red), and D1 and D2 loop (green) and B) main regions. Structures rendered by PyMOL (17).

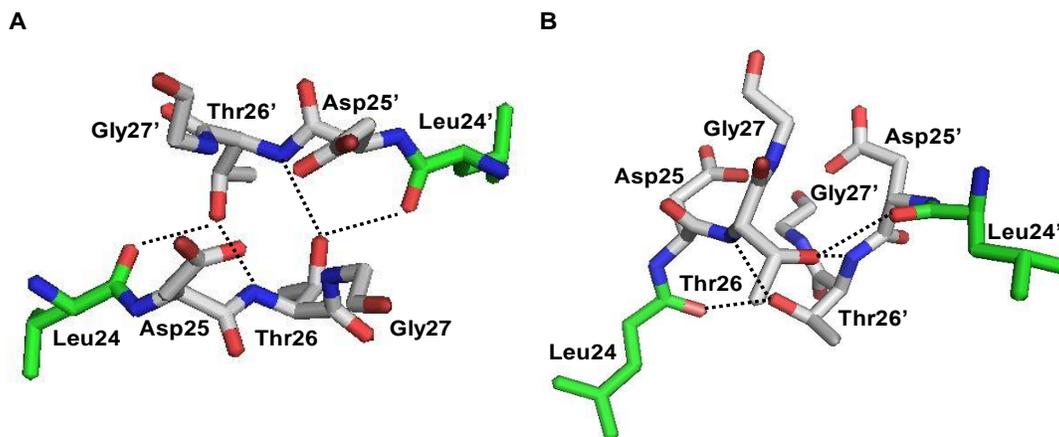


Figure 1-6. Illustration of the “fireman’s grip”. A) Top view and B) side view, in which the hydrogen bonds between the active site residues are shown. Atoms are color-coded as follows: carbon (gray), oxygen (red), and nitrogen (blue). Carbon atoms of Leu24 are shaded green to distinguish it from the Asp25-Thr26-Gly27 catalytic triad. Structures rendered by PyMOL (17).

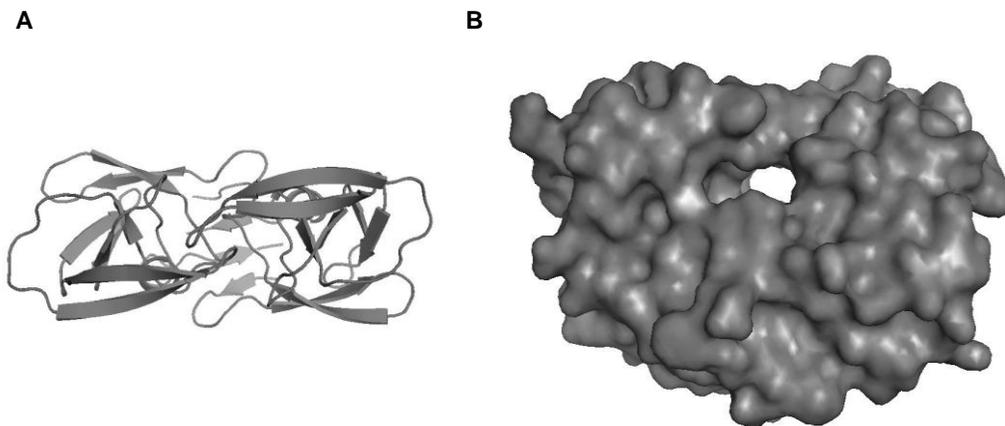


Figure 1-7. Crystal structures of HIV-1 PR (PDB ID 2BPX) A) top view and B) space filling model in the closed conformation. Structures rendered by PyMOL (17).

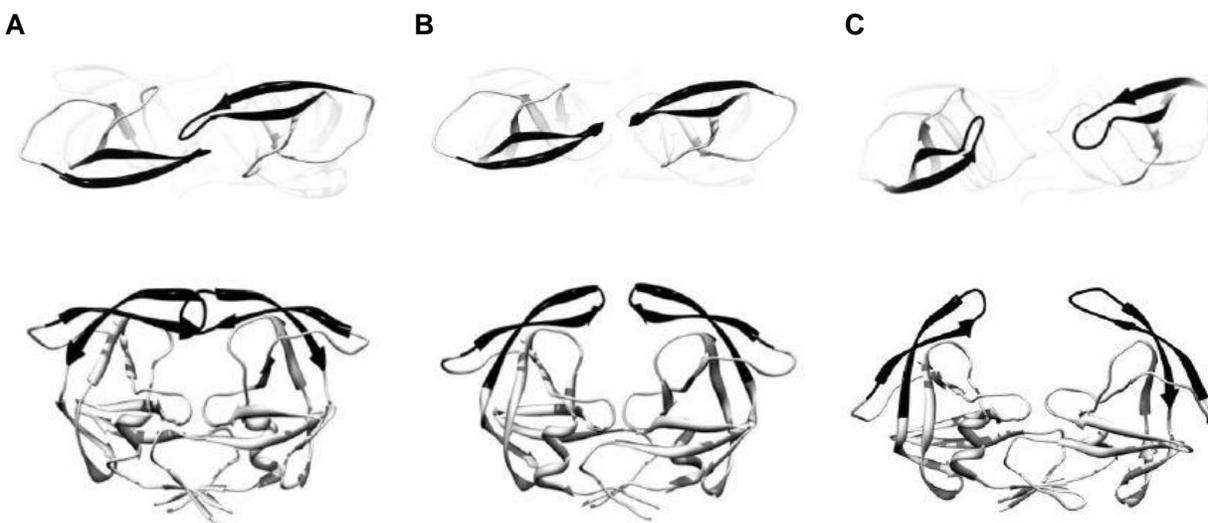


Figure 1-8. Molecular dynamics (MD) simulation structures of HIV-1 PR. Top view (top) and side view (bottom) of the A) closed, B) semi-open, and C) wide-open flap conformations. Figure modified from Simmerling et al. (37).

Table 1-2. Cleavage sites in HIV-1 polyproteins.

Site	Location	HXB2 ^a Consensus Sequence
<i>gag</i> polyprotein		
I	MA/CA	VSQNY / PIVQN
II	CA/p2	KARVL / AEAMS
III	p2/NC	SATIM / MQRGN
IV	NC/p1	ERQAN / FLGKI
V	p1/p6 ^{gag}	RPGNF / LQSRP
<i>pol</i> polyprotein		
VI	NC/TFP	ERQAN / FLREN
VII	TFP/p6 ^{gag-pol}	EDLAF / LQGKA
VIII	p6 ^{gag-pol} /PR	TSFSF / PQITC
IX	PR/RT	CTLNF / PISPI
X	RT/RN	GAETF / YVDGA
XI	RN/IN	IRKVL / FLDGI
<i>Nef</i> polyprotein		
Nef		AACAW / LEAQE

^aPDB accession ID of HIV-1 PR subtype B LAI consensus reference sequence.

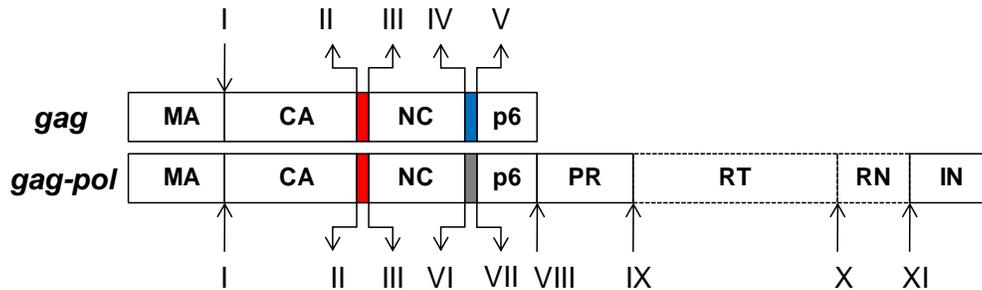


Figure 1-9. Schematic representation of HIV-1 *gag* and *gag-pol* polyproteins. Arrows and numbers designate cleavage sites (refer to Table 1-2 for the location of sites). The representations of the labels are as follows: MA (matrix), CA (capsid), NC (nucleocapsid), PR (protease), RT (reverse transcriptase), RN (RNase H), IN (integrase), p2 (red-shaded box), p1 (blue-shaded box), TFP (transmembrane protein, gray-shaded box).

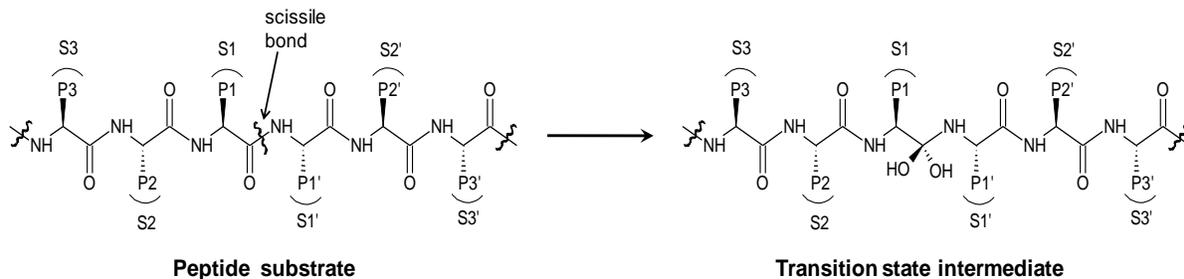


Figure 1-10. Chemical structure of HIV-1 PR substrate showing the scissile bond between the P1/P1' site (left) and the transition state intermediate (right) formed during substrate cleavage.

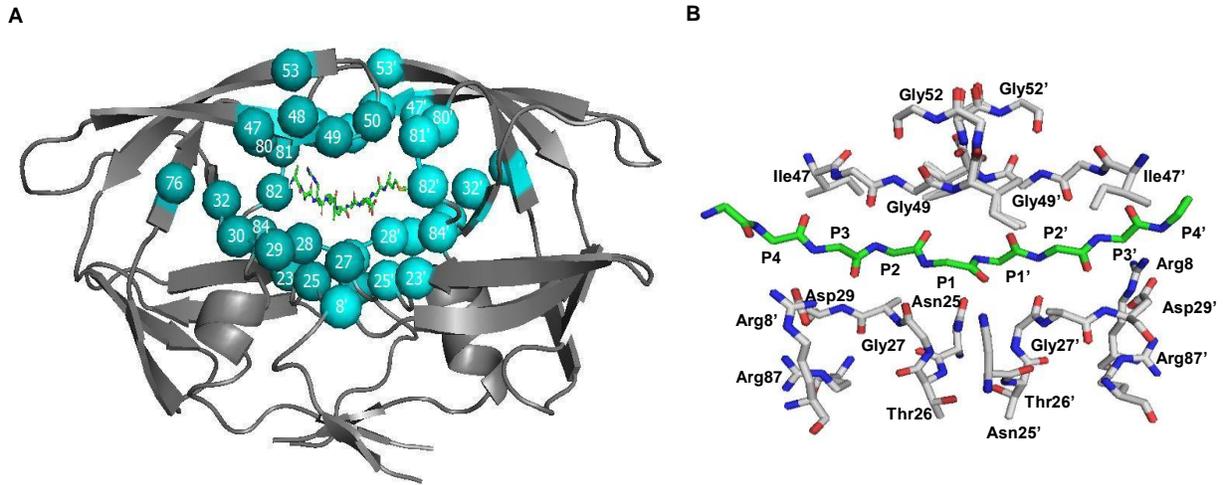


Figure 1-11. Substrate binding pocket of HIV-1 protease. A) Ribbon diagram (PDB ID 2BPX) showing the primary residues involved in substrate binding. Residues from the two subunits of the dimer are distinguished by a prime. B) Enlarged view of the substrate binding pocket (PDB 1F7A). Structures rendered by PyMOL (17).

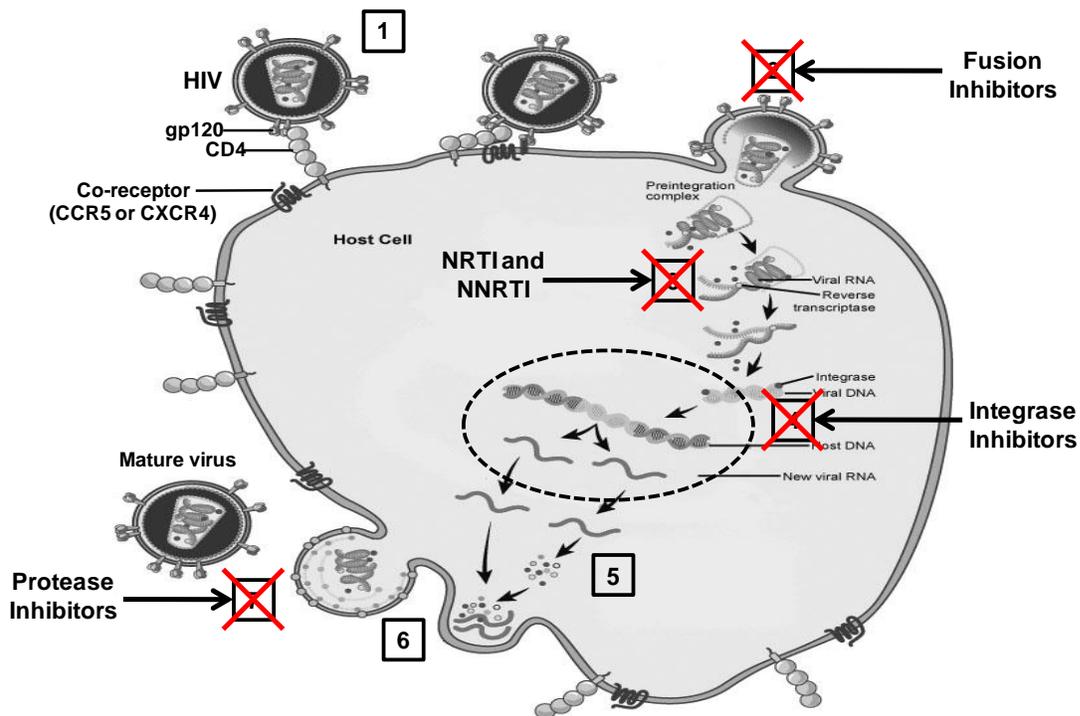


Figure 1-12. Various types of inhibitors targeting different stages in the HIV-1 viral life cycle. Figure adapted from National Institute of Allergy and Infectious Diseases (<http://www.niaid.nih.gov>).

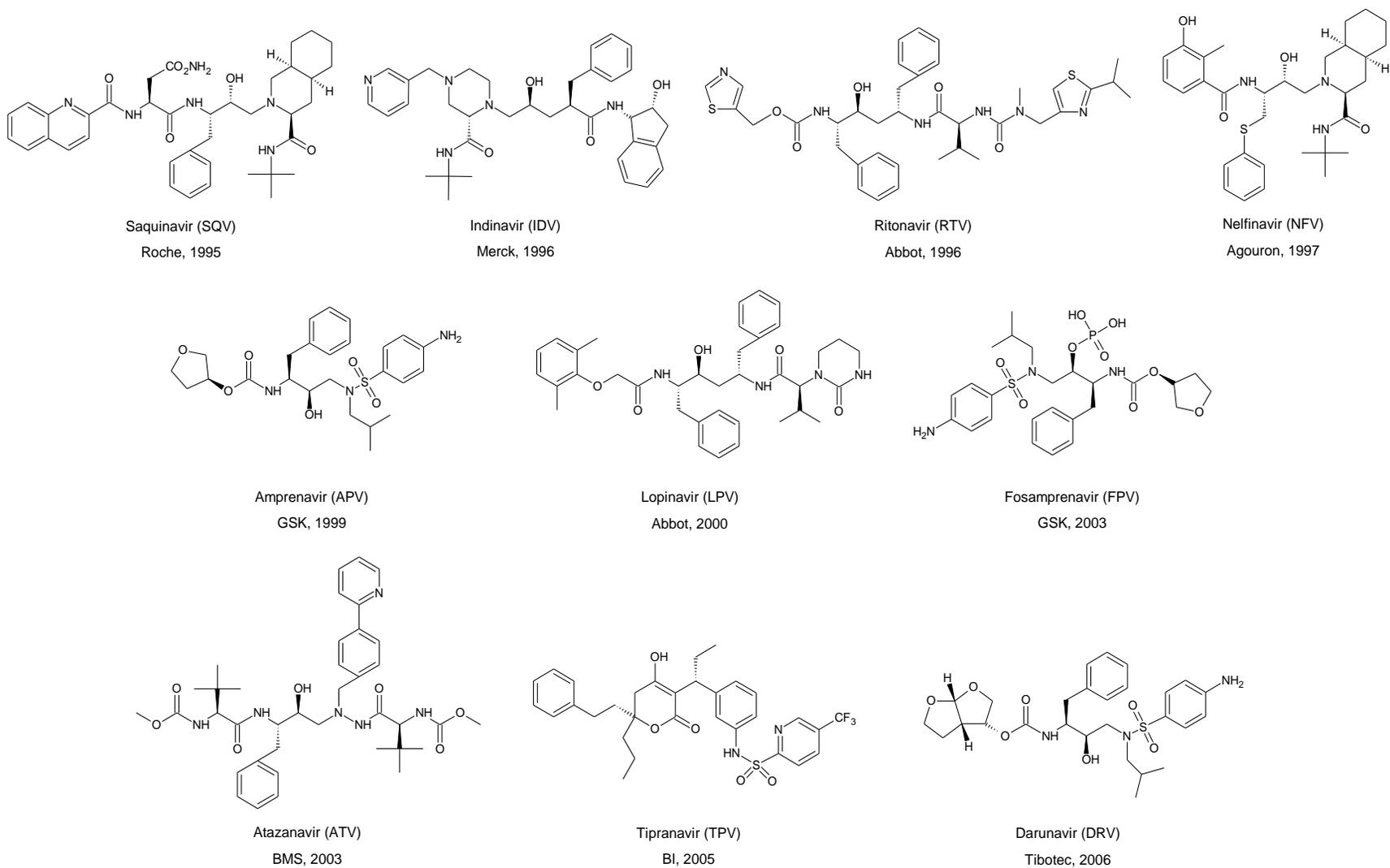


Figure 1-13. Chemical structures of the FDA-approved HIV-1 protease inhibitors (PI).

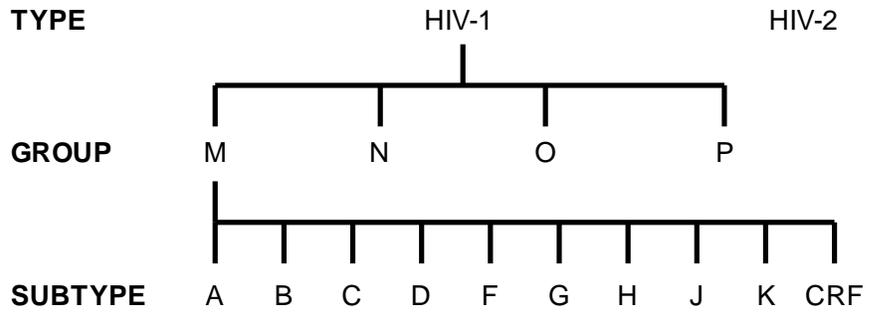


Figure 1-14. General classification of HIV.

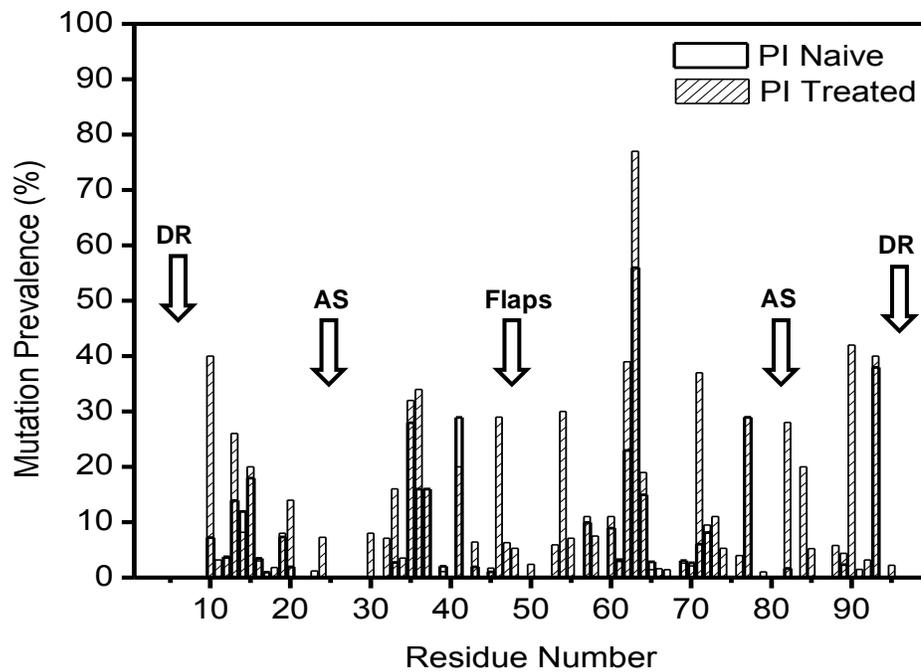


Figure 1-15. Sequence variation of protease inhibitor (PI) naïve and treated isolates of subtype B HIV-1 PR. The mutation prevalence is relative to the LAI consensus sequence. (DR = dimerization region, AS = active site)

Table 1-3. Protease inhibitor (PI) resistance-associated mutations.

Inhibitor	Primary Mutations	Secondary Mutations
Atazanavir/ +/- Ritonavir	I50L, I84V, N88S	L10I/F/V/C, G16E, K20R/M/I/T/V, L24I, V32I, L33I/F/V, E34Q, M36I/L/V, M46I/L, G48V, F53L/Y, I54L/V/M/T/A, D60E, I62V, I64L/M/V, A71V/I/T/L, G73C/S/T/A, V82A/T/F/I, I85V, L90M, I93L/M, V11I, V32I, L33F, T74P, L89V
Darunavir/ Ritonavir	I47V, I50V, I54L/M, L76V, I84V	L10F/I/R/V, V32I, M46I/L, I47V, I54L/V/M, G73S, L76V, V82A/F/S/T, L90M
Fosamprenavir/ Ritonavir	I50V, I84V	L10I/R/V, K20M/R, L24I, V32I, M36I, I54V, A71V/T, G73S/A, L76V, V77I, L90M
Indinavir/ Ritonavir	M46I/L, V82A/F/T, I84V	L10F/I/R/V, K20M/R, L24I, L33F, M46I/L, I50V, F53L, I54V/L/A/M/T/S, L63P, A71V/T, G73S, I84V, L90M
Lopinavir/ Ritonavir	V32I, I47V/A, L76V, V82A/F/T/S	L10F/I, M36I, M46I/L, A71V/T, V77I, V82A/F/T/S, I84V, N88D/S
Nelfinavir	D30N, L90M	L10I/R/V, L24I, I54V/L, I62V, A71V/T, G73S, V77I, V82A/F/T/S, I84V
Saquinavir/ Ritonavir	G48V, L90M	L10V, L33F, M36I/L/V, K43T, M46L, I54A/M/V, H69K/R, L89I/M/V
Tipranavir/ Ritonavir	I47V, Q58E, T74P, V82L/T, N83D, I84V	

Table adapted from Johnson et al. (77).

Table 1-4. Sequence alignment of subtype B LAI, B_{si}, PRE_i, POST_i, V6_i, and MDR769_i. Amino acid residues in blue represent differences relative to LAI. D25N and K55C mutations are labeled red and green, respectively. Q7K, L33I, and L63I stabilizing mutations are highlighted in yellow. C67A and C95A mutations are highlighted in cyan.

Construct	Sequence				
	10	20	30	40	50
B (LAI)	PQITLWQRPL	VTIKIGGQLK	EALLDTGADD	TVLEEMSLPG	RWKPKMIGGI
B _{si}	----- K -----	-----	----- N -----	----- I -----	-----
PRE _i	-----	-----	----- N -----	----- T-T -----	-----
POST _i	----- I -----	----- V -----	----- N -----	----- Q-IN-T -----	-----
V6 _i	-----	----- R -----	----- N -----	----- I-F-I -----	-----
MDR769 _i	----- I -----	-----	----- N -----	----- VN -----	----- L -----
	60	70	80	90	
B (LAI)	GGFIKVRQYD	QILIEICGHK	AIGTVLVGPT	PVNIIGRNLL	TQIGCTLNF
B _{si}	----- C -----	----- I ----- A -----	-----	-----	----- A -----
PRE _i	----- C -----	----- VP ----- A -----	-----	-----	----- A -----
POST _i	----- AC ----- E -----	----- VP ----- A -----	-----	----- A -----	----- A -----
V6 _i	----- C -----	----- P ----- A -----	V -----	----- A ----- M -----	----- A -----
MDR769 _i	----- VC -----	----- VP ----- A -----	V -----	----- A-V ----- M -----	----- A -----

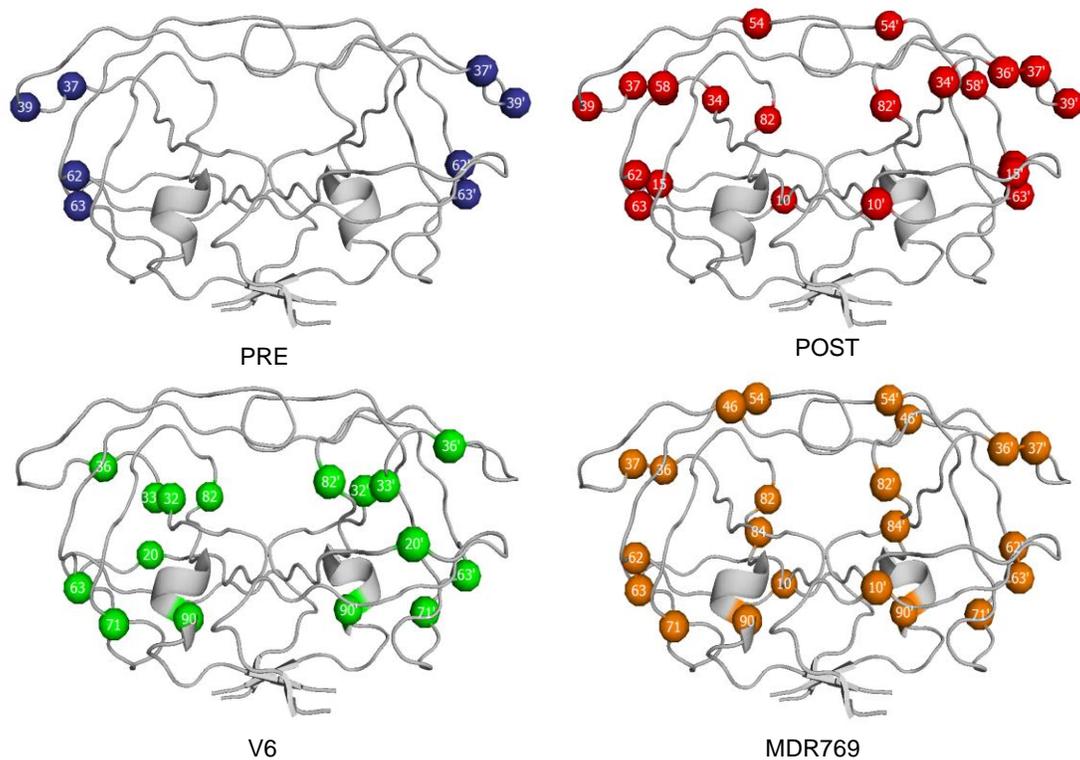


Figure 1-16. Ribbon diagrams of HIV-1 PR PRE, POST, V6, and MDR769, in which the amino acid differences relative to subtype B LAI are highlighted and labeled.

Table 1-5. Amino acid substitutions made on post-therapy (POST) sequence.

MUTANT	<i>gag</i>	PR
PMUT1	V376I, V398E	A82V
PMUT2	V376I, V398E	A54I, E58Q
PMUT3	V376I, V398E	I10L, V15I, A82V
PMUT5	V376I, V398E	Q34E, I36M, N37T, A54I, E58Q

CHAPTER 2 BACKGROUND ON TECHNIQUES

Cloning and Protein Expression Systems

Molecular cloning, also referred to as recombinant DNA technology, is a modern approach for obtaining sufficient quantities of a protein of interest. Cloning involves the insertion of a DNA segment of interest into an autonomously replicating DNA molecule or vector, which is used for the propagation of the target gene in a host bacterial cell (106, 107). The widely used bacterial system is *Escherichia coli* (*E. coli*). The vector, which contains suitable transcription and translation control elements, is also used for high-level production or overexpression of the target protein in the host cell (108). Protein expression using recombinant DNA technology requires careful selection of the DNA for insertion, vector, and type of *E. coli* cells.

Synthetic DNA can be purchased from DNA 2.0 (<https://www.dna20.com>) or other companies with DNA synthesizers. The advantages of using synthetic genes include the optimization for codon usage bias and enhancement of protein expression (109). Codon optimization is important particularly if the target protein is not naturally expressed in *E. coli*. The difference in the codon usage of prokaryotic and eukaryotic systems results in a poor level of protein expression (107). In addition, the DNA insert can be constructed to prevent the formation of secondary structures in mRNA, such as pause or stop loops, which may block translation (107, 109).

A typical expression vector, also known as plasmid, is shown in Figure 2-1. It is a circular, double-stranded DNA (dsDNA), which can replicate independently from the chromosome. The gene of interest can be inserted into the vector if restriction sites found in the plasmid are incorporated at the ends of the DNA insert. The chosen

restriction sites should not be naturally present within the gene to avoid excision of the gene by restriction enzymes. Restriction enzymes are endonucleases that recognize and cut a particular site producing a “staggered cut” or short overhanging single-stranded ends, referred to as sticky ends (108). The restriction sites are usually palindromic, that is, the sequence is read the same on both strands. The DNA of interest is then inserted into the plasmid by ligation, the process of aligning the ends of two double-stranded DNA molecules to form a covalent linkage or a phosphodiester bond (106).

Vectors consist of several features that are essential for its propagation in a host cell (108). The plasmid generally possesses an origin of replication (*ori*) so that DNA can replicate. It also contains a selectable marker or a resistance gene to distinguish cells that have taken up the vector from those that have not. A commonly used resistance gene is *bla* or *amp^R*, which encodes β -lactamase to inactivate ampicillin, an antibiotic that hinders growth of bacteria by preventing the polysaccharide chains in the bacterial cell walls to cross-link. In effect, cells with the *amp* gene survive in a medium containing ampicillin whereas those without the resistance gene die. Expression vectors possess a multiple cloning site (MCS) or polylinker region, which contains several restriction sites for insertion of the gene of interest. Plasmids also consist of a promoter region, where RNA transcription begins. Promoters such as the T7 promoter are often inducible and not constitutive.

The pET vector system, which is based on the T7 promoter-driven system, has been used to express thousands of different proteins (110-112). These vectors are often transformed into a host bearing the λ DE3 lysogen, which is the T7 RNA polymerase

gene, for expression of target proteins (113). Figure 2-2 illustrates the host and vector elements available for control of T7 RNA polymerase levels using the pET system (114). A number of pET vectors are commercially available including the pET23a vector (Figure 2-3), which is chosen for this work.

The *E. coli* expression system has many advantages including its ease of growth and manipulation and the availability of a variety of vectors and host strains that have been developed for maximizing protein expression (107). However, the challenge is in choosing the appropriate host strain for a particular expression vector. The pET vector system requires a host strain containing the bacteriophage λ DE3 gene, which encodes for T7 RNA polymerase (114). The benefit of having λ DE3 lysogen with a *lacUV5* promoter is that target genes are expressed only upon induction therefore basal expression levels and target gene expression can be controlled (112, 113). As illustrated in Figure 2-2, transcription of T7 RNA polymerase and expression of target gene proceeds by induction using lactose or its analog. Isopropyl- β -D-1-thiogalactopyranoside (IPTG) is the commonly used lactose analog (Figure 2-4). The use of the T7 expression system results in the production of large amounts of the desired protein after induction.

The most widely used hosts are BL21 and its derivatives, which lack both *lon* and *ompT* proteases (115, 116). In this work, BL21(DE3) cells with genotype *E. coli* B F⁻ *dcm ompT hsdS*(r_B⁻ m_B⁻) *gal* λ (DE3) is used as the expression strain because it contains the T7 system, which allows for high-level protein expression and easy induction (117). The BL21(DE3)pLysS cells of the genotype *E. coli* B F⁻ *dcm ompT hsdS*(r_B⁻ m_B⁻) *gal* λ (DE3) [pLysS Cam^r] are similar to the BL21(DE3) strain, but contains the pLysS

plasmid, which provides tighter control of protein expression (117). This strain carries the gene encoding for T7 lysozyme, which binds to T7 RNA polymerase to reduce basal transcription of target genes in the uninduced state. T7 lysozyme lowers the background expression level of target genes but does not interfere with levels of expression after induction with IPTG. The presence of target protein before induction can be potentially toxic to the host cell and results in low level of protein expression upon induction. Thus, BL21(DE3)pLysS cells are primarily used to minimize the production of toxic proteins.

Site-Directed Mutagenesis

Site-directed mutagenesis is an important technique in molecular biology for modifying DNA sequences, particularly in studying protein structure-function relationships. Most of the mutagenesis methods have been developed based upon the polymerase chain reaction (PCR) (108). The most common and simplest method is the QuickChange™ Site-Directed Mutagenesis System developed by Stratagene (118). In this method, a dsDNA vector is used as a template and primers containing the site of mutation are allowed to anneal to the DNA followed by extension using PCR (119). For each round of PCR, the mutation of interest is introduced using a pair of complementary primers.

Site-directed mutagenesis imposes certain requirements on the primers to favor primer-template annealing rather than primer-dimer formation (118, 119). Primers have a typical length of about 25–45 bases. Longer oligonucleotides are more specific and stable. The site of mutation is normally situated in the middle of the oligonucleotide such that about 10-15 base pairs are present on each side. The requirement for primer melting temperature (T_m) is at least 78°C, which should be higher than the annealing

temperature. Oligonucleotides with higher GC content have a greater T_m and are therefore more stable; however GC base pairs at the ends should be minimized to avoid the primer from annealing to itself.

The typical strategy for site-directed mutagenesis is shown in Figure 2-5 (119). The first step involves annealing of the primers to the template plasmid DNA. Then, DNA is linearly amplified using PCR in the presence of deoxynucleotide triphosphates (dNTPs) and DNA polymerase. A high-fidelity polymerase is used to minimize the chances of introducing unwanted mutations when extending an entire plasmid. *Pfu* DNA polymerase is especially used for this purpose because it is thermostable, non-strand displacing, and has the 3'-5' proofreading ability (118). Also, addition of dimethylsulfoxide (DMSO) into the reaction mixture prevents primers and template DNA to anneal to itself (119). DMSO disrupts base pairing, facilitating strand separation in GC rich regions of DNA, and reducing the propensity of the DNA to form secondary structure. After the PCR reaction, the product is digested with *DpnI*. This endonuclease preferentially nicks *dam* methylated and hemimethylated DNA (5'-Gm⁶↓ATC-3', where ↓ represents the cleavage site) and thus digests the parental plasmid but not the PCR product. DNA isolated from almost all *E. coli* strains is *dam* methylated and therefore susceptible to *DpnI* digestion. Also, the template plasmid must not come from a methylation deficient (*dam*⁻) *E. coli* strain such as JM101. Following *DpnI* digestion, the PCR product is transformed into a suitable strain of *E. coli* cells, e.g. XL1-blue cells, which is capable of sealing nicks in the PCR product. Without the *DpnI* digest, a large number of colonies will contain the parental DNA and not the plasmid DNA having the

mutation of interest because the transformation efficiency of the template plasmid is several orders of magnitude better than the linear PCR product.

Purification and Characterization of Proteins

The initial step in the isolation of a target protein involves cell lysis (106). Methods for breaking open cells involve mechanical disruption using a homogenizer, which uses a closely-fitting piston and sleeve to crush tissue; a French press, which is a device that shears open cells by pushing through a small orifice at high pressure; or a sonicator, which uses ultrasonic vibrations. After lysis, cells are filtered or centrifuged to remove particulate cell debris. The protein of interest is usually present in the supernatant solution. However, the water-soluble protein may also settle into inclusion bodies in the form of a cell pellet.

Isolation of Proteins from Inclusion Bodies and Refolding

Recombinant proteins are most often expressed in the intracellular space of the bacteria, but expression can be controlled so that the target protein is secreted into the periplasmic space or into the culture medium (107). Some water-soluble proteins are also produced in the bacterial cell in insoluble form, which lack biological activity as a result of protein misfolding and aggregation. Such proteins are sequestered into inclusion bodies.

After cell lysis, inclusion bodies can be separated by centrifugation from soluble proteins and other cellular components. It is then solubilized using denaturants, such as urea or guanidine hydrochloride, and reducing agents to break the disulfide bonds. Removal of the denaturant and the process of refolding the protein are carried out either by dilution or dialysis (107). The conditions of refolding and reformation of disulfide

bonds, which varies from protein to protein, must be optimized by determining the suitable pH and ionic strength of buffer, detergent, dilute acid or base to be used.

Chromatography

Protein purification requires chromatography to separate the target protein from all other proteins. The basic principle of chromatography is to bind a protein of interest to a solid support followed by elution using a suitable solvent (120). Separation can be based on ionic interactions, size exclusion, or affinity.

In ion exchange chromatography, ions bound electrostatically to an insoluble and chemically inert matrix, called cation or anion exchangers, are reversibly replaced by ions in solution (106). Proteins, in particular, are polyelectrolytes that possess both positive and negative charges and as such bind to either cation or anion exchangers. The strength of binding depends on the number of charges in the protein. The electrical charge of proteins depends on the amino acid composition and the pH of the medium. If the pH is lower than the isoelectric point (pI), the protein will have a positive net charge and a pH above the pI will lead to a net negative charge. In purifying protein, the pH and salt concentration of the buffer are chosen so that the desired protein is strongly bound to the selected ion exchanger (120). Weakly bound protein can be eluted with salt solutions of low concentrations whereas strongly bound proteins require high salt concentrations of about 10 mM to 1 M. Cellulosic ion exchangers, such as DEAE (diethylaminoethyl), are typically employed in the initial stage of purification to separate the target protein from the rest of the components in the cell lysate, which bind to the column.

Gel filtration, also called gel chromatography or size exclusion chromatography (SEC), is based on the different sizes and shapes of protein molecules (106). The

matrix is composed of beads made of hydrated, sponge-like material containing pores of different sizes. If a solution containing molecules of varying size is passed through a gel column, smaller molecules diffuse into the beads and prolong column passage. The largest proteins, however, cannot penetrate the pores and therefore traverse the column more rapidly. Gel filtration is a gentle method because proteins are not bound to the gel; however, the drawback is that it requires small volumes of concentrated samples (120). The method is typically carried out as the last step in the purification process.

Affinity chromatography is based on the ability of proteins to interact with specific molecules tightly but noncovalently, e.g. enzymes with substrates, receptors with ligands, antibodies to antigens, or glycoproteins with lectins (120). In this technique, protein in solution is separated from other substances by interacting with a ligand that is covalently attached to an inert and porous matrix. The desired protein is then recovered in highly purified form by optimized elution conditions, i.e. changing the pH or ionic strength to reduce protein-ligand interactions.

Fusion tags can be incorporated into vectors to allow expression of proteins with a short peptide sequence attached to the N- or C-terminus and thereby facilitate detection and purification of the target protein (107, 114). The presence of tags in the protein improves purification because the tagged protein can bind specifically to a particular resin. Commercially available tags include His₆ (six tandem histidine residues, which bind to Ni-NTA), GST (glutathione S-transferase, which binds to glutathione-Sepharose), thioredoxin (binds to ThioBond resin) and maltose binding protein (binds to amylose resin) among others (107). The His-Tag sequence is particularly useful for purifying proteins that are initially expressed in inclusion bodies because affinity

purification can be accomplished in denaturing conditions that solubilize the protein (114).

SDS-PAGE

Electrophoresis is the process of moving charged molecules by applying an electric field (121). Electrophoresis of macromolecules, such as proteins and nucleic acids, is carried out in a porous matrix of either polyacrylamide or agarose gel. When an electric field is applied to a protein solution, it migrates at a rate that depends on its net charge, size, and shape (120, 122).

Sodium dodecylsulfate-polyacrylamide gel electrophoresis (SDS-PAGE) uses a highly cross-linked gel of polyacrylamide as the inert matrix through which the proteins migrate (122). The matrix serves as a size-selective sieve in the separation. The gel is prepared by polymerization of acrylamide and cross-linked by methylene bis-acrylamide. The pore size of the gel can be adjusted so that it is small enough to retard the migration of the protein molecules of interest.

SDS-PAGE is often employed under denaturing conditions so that proteins are separated only according to size (120). Before electrophoresis, protein samples are dissolved in a sample buffer, which contains a reducing agent such as dithiothreitol (DTT) or β -mercaptoethanol (BME) to break any disulfide linkages in the proteins. The samples are then boiled to disrupt the tertiary structure of the protein such that all of the constituent polypeptides of a multi-subunit can be analyzed separately. The sample buffer also includes a negatively charged detergent, sodium dodecylsulfate (SDS), which binds to the hydrophobic regions of the protein molecules, causing them to unfold into extended polypeptide chains and release from their associations with other proteins or lipid molecules (122). SDS binds to the protein at a fairly constant ratio of about 1

molecule of SDS per 3 amino acids, resulting in an almost constant charge to weight ratio and a net negative charge (120). The proteins then migrate to the anode or positive end of the electrophoresis reservoir with a constant acceleration, independent of their composition. The size of a protein can be estimated by comparing the distance of migration with standard proteins of known molecular weight.

Circular Dichroism (CD) Spectroscopy

Circular dichroism (CD) spectroscopy is a technique often used for monitoring protein folding. Similar to ultraviolet-visible (UV-Vis) spectroscopy, it is based on the principle of light absorption of a certain material as a function of wavelength (106). A solution containing a solute absorbs light according to the Beer-Lambert law (Equation 2-1),

$$A = \log\left(\frac{I_0}{I}\right) = \epsilon cl \quad (2-1)$$

where A is the solute's absorbance or optical density, I_0 is the incident intensity of light at a given wavelength λ , I is the transmitted intensity at λ , ϵ is the molar extinction coefficient of the solute at λ , c refers to the molar concentration, and l is the length of the light path in cm. Because the value of ϵ varies with λ , A or ϵ can be plotted against λ to yield the absorbance spectrum. The wavelength of absorption depends on the functional groups (chromophores) or arrangement of atoms in the sample (123). Polypeptides absorb strongly in the ultraviolet (UV) region of the spectrum ($\lambda = 100$ to 400 nm) because of the aromatic side chains of Phe, Trp, and Tyr (124). In general, the wavelength at 280 nm is used to monitor the optical density of proteins.

Chromophores in chiral environments generate circular dichroism as a consequence of light absorption (124). A CD spectrometer measures the intensity of absorption of left-circularly polarized light relative to that of right-circularly polarized light over a continuous range of wavelengths. Proteins are chiral molecules with different values for left and right circularly polarized light, ε_L and ε_R . The difference in these quantities, $\Delta\varepsilon = \varepsilon_L - \varepsilon_R$, as a function of λ constitutes the CD spectrum. Different structural elements exhibit certain characteristic spectra, shown in Figure 2-6. Spectra of α -helical proteins produce very intense CD signals, which correspond to negative bands at 222 nm and 208 nm and a positive band at 193 nm, whereas spectra for proteins with β -sheet structure are variable, with a negative band appearing at 218 nm and positive band at 195 nm (126, 127). Disordered proteins, which have a “random” conformation, have very low ellipticity above 210 nm and negative bands near 195 nm (126). The CD spectrum can be analyzed to determine the percentage of secondary structure present in a single polypeptide or protein molecule (127). Various software for analyzing CD data are available at the following websites: CDPro (<http://lamar.colostate.edu/~sreeram/CDPro/main.html>), Circular Dichroism at UMDNJ (http://rwjms.umdj.edu/research/cdf/about_cd/applications.html), CONTIN (<http://s-provencher.com/pages/contin-cd.shtml>), CCA + the CD Spectrum Analyser System (<http://www.chem.elte.hu/departments/protnmr/cca/>), DICROPOT (<http://dicropot-pbil.ibcp.fr/>), DICHROWEB (<http://dichroweb.cryst.bbk.ac.uk/html/home.shtml>), K2D (<http://www.embl.de/~andrade/k2d.html>), and SOMCD (<http://geneura.ugr.es/cgi-bin/somcd/index.cgi>) (126).

Mass Spectrometry (MS)

Mass spectrometry (MS) is widely used for the identification and characterization of biological macromolecules, particularly proteins. It is a technique that measures the molecular mass by analysis of the ions produced upon ionization of a sample (128). Samples are traditionally ionized by electron-impact (EI), where molecules are bombarded with high energy particles, producing numerous fragments. Recent MS techniques use “soft” ionization methods to produce molecular ions and to overcome the problems associated with the thermal instability and involatility of macromolecular analytes (129).

The electrospray ionization (ESI) technique is a soft ionization method, which is capable of generating non-fragmented molecular ions from biological macromolecules in aqueous solution (128). In this method, the sample is passed through a low voltage needle, in which the electric field at the needle tip disperses the solution into a fine spray of charged droplets. These droplets, which form smaller droplets as they traverse through a tube, are eventually ionized and propelled into the mass analyzer. The output, referred to as the mass spectrum, consists of ions with different intensities and mass/charge (m/z) ratio. Mass spectra of proteins are usually complicated because of the presence of multiply-charged ions.

Electron Paramagnetic Resonance (EPR) Spectroscopy

Electron paramagnetic resonance (EPR) spectroscopy is a magnetic resonance technique that studies paramagnetic species or systems of unpaired electrons and their interactions in the presence of an external magnetic field. In the simplest case, a free electron possesses a magnetic moment that aligns parallel or antiparallel in a magnetic field. This induces the Zeeman Effect, a process whereby the electron spin states ($m_s =$

$\pm\frac{1}{2}$) break their spin degeneracy and split into levels, as illustrated in the energy diagram in Figure 2-7. In the case of EPR, the difference in energy levels (ΔE) is proportional to the applied magnetic field, mathematically expressed as the Zeeman equation (Equation 2-2) (130, 131). The change in energy is proportional to the Planck's constant ($h = 6.626068 \times 10^{-34} \text{ m}^2 \cdot \text{kg/s}$) and frequency, which is proportional to the strength of the applied magnetic field (B), g or the spectroscopic g-factor (approximately equal to 2 for most samples), and the Bohr magneton, μ_B ($9.2740154 \times 10^{-24} \text{ J/T}$). μ_B is a proportionality constant defined in Equation 2-3, where e is the electric charge, \hbar is $\frac{h}{2\pi}$ ($1.054 \times 10^{-34} \text{ J}\cdot\text{s}$), and m_e is the mass of the electron ($9.109 \times 10^{-31} \text{ kg}$). The resonant condition occurs when the applied energy is equal to the splitting between the levels, E_α and E_β .

$$\Delta E = h\nu = g\mu_B B \quad (2-2)$$

$$\mu_B = \frac{e\hbar}{2m_e} \quad (2-3)$$

Site-Directed Spin Labeling

Site-directed spin labeling (SDSL), in conjunction with EPR, is a common tool for analyzing protein structure and dynamics, particularly in measuring the mobility of proteins (132-135). This technique involves the introduction of a spin-labeled side chain into protein sequences, usually through the use of site-directed mutagenesis by cysteine substitution followed by reaction with a sulfhydryl-specific nitroxide reagent. The general reaction scheme is shown in Figure 2-8. The commonly used spin label is (1-oxy-2,2,5,5-tetramethylpyrrolinyl-3-methyl) methanethiosulfonate (MTSL) due to its

high sulfhydryl selectivity and reactivity, relatively smaller volume of the modified amino acid side chain, and sharp sensitivity of EPR spectra to structural changes (132).

SDSL-EPR is a useful technique for monitoring conformational changes in biomacromolecules because the motional dynamics of the side chain, as reflected in the EPR spectral line shape, correlates with the general feature of the protein fold. Changes in the rate of rotation of bonds results in changes in spectral shape. The EPR spectra are very sensitive to changes in secondary structure and local dynamics, thus can be used in the detection of protein conformational changes (132-140).

Continuous-Wave EPR (CW-EPR)

As previously described, EPR is a phenomenon that arises from the Zeeman Effect, wherein resonant condition is achieved when the energy applied is equal to the splitting of the energy levels in a magnetic field. In continuous wave-EPR (CW-EPR), the resonant condition is attained by varying the magnetic field at a constant frequency. The resultant spectrum, shown in Figure 2-9, is a derivative of an absorption peak or a dispersive peak, which results when the field is modulated (130, 131).

When analyzing proteins by SDSL-EPR, the resultant spectrum consists of three dispersive peaks, which correspond to the splitting of the energy levels as a result of the interaction of the nitroxide spin label with the nitrogen nucleus. The interaction of the electron in the nitroxide spin label and the nuclear spin of nitrogen is referred to as the hyperfine interaction. This coupling splits each of the Zeeman interaction levels into three different energy levels, based on the $2I+1$ splitting rule, corresponding to $m_I = 0, \pm 1$. The energy diagram and derivative absorption spectrum are given in Figure 2-10 (130, 131).

The line shape of an EPR spectrum of a spin-labeled side chain contains information about the dynamics or motion of the nitroxide ring on the nanosecond timescale (132). The motion of the nitroxide ring originates from rotation around bonds within the nitroxide side chain as well as local backbone fluctuations (LBFs), which are rigid-body motions of secondary structure elements or oscillations about backbone dihedral angles (135, 137, 139). These motions are reflected in the EPR spectra of the nitroxide side chain and provide information regarding three different correlation times of motion: τ_R or the rotational correlation time, which represents the overall tumbling of the protein in solution, τ_B or effective correlation time due to rotational isomerizations about the bonds that link the nitroxide to the backbone (spin label mobility), and τ_S or the effective correlation time for segmental motion of the backbone relative to the average protein structure (flexibility of backbone) (138). Figure 2-11 shows EPR spectral line shapes of nitroxides with varying correlation times. An EPR line shape characterized by very distinct, sharp peaks corresponds to a spin with fast isotropic motion (Figure 2-11A). When motion is restricted to some extent, the peak is broadened slightly (Figure 2-11B). As motion becomes more restricted and correlation time increases, significant line broadening is observed, as seen in the slow and rigid spectra (Figure 2-11C and D, respectively). Some spectra possess multiple dynamic states, which represent a rich variety of complex dynamic modes that are fingerprints of the local structure (141). For example, spins located in buried sites possess a dominant immobile component as a consequence of tertiary interactions while in most solvent exposed sites, a dominant mobile component is observed because of partial flexibility of the spin label (141).

Double Electron-Electron Resonance (DEER) Spectroscopy

The conformational changes that a protein molecule undergoes reveal information about its biological function. The relative movement in protein domains can be measured by determining the distance between two sites in the molecule. Site-directed spin labeling with CW-EPR can access distances of up to 25 Å while pulsed-EPR techniques can be used to measure distances between 20–80 Å (142). The four-pulse double electron-electron resonance (DEER) experiment has been the most widely applied method for distance measurements on spin-labeled biomacromolecules (143).

Two unpaired electrons separated by a distance r are coupled to each other through electron-electron dipolar and other short-range interactions such as J coupling (144). In general, the coupling between two spins S_A and S_B , is described by the Hamiltonian:

$$\hat{H}_{ee} = \hat{S}_A D \hat{S}_B + J \hat{S}_A \hat{S}_B \quad (2-4)$$

where D is the dipole-dipole tensor and J is the exchange coupling (142). For distances longer than 10 Å, the following are assumed: exchange coupling is isotropic, g anisotropy is small, and electron spins are localized at the center of the N–O bond (142). The dipole-dipole tensor can then be described by the point-dipole approximation:

$$\hat{H}_{ee} = [J + \omega_{dd}(1 - 3\cos^2 \theta)] \hat{S}_A \hat{S}_B \quad (2-5)$$

where θ is the angle between the spin-to-spin vector and the magnetic field axis and ω_{dd} is the dipole-dipole coupling between spins described by the equation:

$$\omega_{dd} = \frac{1}{r_{AB}^3} \frac{\mu_0}{4\pi\hbar} g_A g_B \mu_B^2 \quad (2-6)$$

Thus, distance measurements rely on ω_{dd} , which is inversely proportional to the cube of the distance r . μ_0 is the magnetic permeability of vacuum (1.256637×10^{-6} T·m/A), μ_B is the Bohr magneton and the g values can be approximated by the isotropic value, $g_A = g_B \approx 2.006$. At distances larger than 15 Å, J couplings are an order of magnitude smaller than the dipole-dipole coupling and can thus be neglected in Equation 2-5.

An echo experiment is typically employed to obtain data, which provides information about the interactions between an observer spin A and a second electron spin B. The pulse pattern for the four-pulse DEER echo experiment is illustrated in Figure 2-12 (143). At the observer frequency, a refocused Hahn-echo sequence is applied. Based on the vector model, a first $\frac{\pi}{2}$ pulse is applied to turn the magnetization of the A spins into the xy plane. After evolving for time τ_1 , a π pulse is applied to produce the electron spin echo (ESE), which corresponds to the inverted dotted line on Figure 2-12. After ESE, the final π pulse on the observer sequence is applied after another evolution time τ_2 , which eventually refocuses the spin echo. The π pulse at the pump frequency, in between the two π pulses on the observer frequency, is applied for detection of the dipolar interactions. The π pulse on the second irradiation frequency only affects the B spins, but the flipping of magnetization changes the local magnetic field around the A spins. By varying the delay time t , the intensity of the refocused echo varies as a result of the dipolar evolution. In effect, the refocused ESE corresponds to a coherence transfer echo, the intensity of which is modulated as a function of the strength of the coupling between spins A and B.

The echo signal $V(t)$ as a function of the dipolar evolution time t is a product of the contributions of $F(t)$ and $B(t)$, described by Equation 2-7.

$$V(t) = F(t)B(t) \quad (2-7)$$

As illustrated in Figure 2-13, $F(t)$ refers to the interactions of spins within the same molecule whereas $B(t)$ is background signal due to interactions with spins from neighboring molecules (142, 145). Figure 2-14 shows the plot of $V(t)$ as a function of delay time t . The experimentally determined dipolar evolution function $V(t)$ has to be separated into $F(t)$ and $B(t)$ before further data analysis. Separation of the form factor and background factor is carried out by background subtraction, which will be further discussed later in a separate section. The plot of the form factor, which corresponds to the background subtracted echo, is shown in Figure 2-14.

DEER Data Analysis

The dipolar evolution curve, which manifests the dipolar-dipolar coupling between two spins, can be converted into a distance distribution to determine the spin-to-spin distance r . The simplest method of conversion is by Fourier transformation into a frequency domain spectrum, represented by the Pake pattern shown in Figure 2-15, where the splitting is proportional to $\frac{1}{r^3}$. The distance distribution $P(r)$ can be computed exactly from the form factor $F(t)$ (Equation 2-8), where the kernel function $K(t,r)$ is expressed by Equation 2-9 (145). However, the solution is an inverse problem, which entails finding the distance profile that satisfies the experimental data.

$$F(t) = K(t,r)P(r) \quad (2-8)$$

$$K(t,r) = \int_0^1 \cos[(3 \cos^2 \theta - 1)\omega_{dd}t] dx \quad (2-9)$$

A few methods have been developed to solve the distance distribution $P(r)$. One approach to finding the best solution for the distance distribution $P(r)$ uses curve fitting methods. The method involves simulating distance distribution profiles to generate a theoretical dipolar evolution curve that adequately fits the experimental data. The Monte Carlo (MC) method is one variation to the curve fitting approach, which uses computational algorithms to generate a distance profile with an assumed Gaussian or Lorentzian form. Another method called Tikhonov regularization (TKR) is discussed in further detail in the next section.

Tikhonov Regularization

The conversion of the form factor $F(t)$ into a distance distribution $P(r)$ is a moderately ill-posed problem (142, 145). This means that there is no unique solution to finding the distance profile that best fits the experimental data because different distance profiles can produce similar dipolar evolution curves. In addition, small amounts of noise in the experimental data and slight errors in background subtraction can result in large changes in the distance distribution. Special mathematical algorithms have to be applied to solve such an ill-posed problem, which requires a certain smoothness of the distribution $P(r)$ (145). Smoothness can be achieved by minimizing the target function:

$$G_{\alpha}(P) = \|KP - F\|^2 + \alpha \left\| \frac{d^2}{dr^2} P \right\|^2 \quad (2-10)$$

This method of minimizing G_{α} is referred to as Tikhonov regularization. Equation 2-10 can also be expressed as:

$$G_{\alpha}(P) = \rho + \alpha\eta \quad (2-11)$$

where $\rho = \|KP - F\|^2$ is the criterion for minimum root mean square (RMS) deviation

and $\eta = \left\| \frac{d^2}{dr^2} P \right\|^2$ represents the second derivative of the distance distribution

corresponding to the smoothness criterion. The regularization parameter α (also referred to as λ in some literature) largely determines the solution to $P(r)$. The value of α must be chosen to compromise the smoothness of $P(r)$ and the deviation between experimental and simulated form factors. This can be achieved using the L-curve criterion as a guide. It is based on a plot of $\log \eta(\alpha)$ vs. $\log \rho(\alpha)$ with α being varied on a logarithmic scale as shown in Figure 2-16A.

As seen in Figure 2-16B, small α values correspond to undersmoothing, in which the distance distribution profile contains several sharp, narrow peaks. The corresponding dipolar evolution curve in Figure 2-16C is over-fit such that some of the noise is included in the TKR fit. Increasing α leads to a large decrease in η and smoothing does not affect the deviation. With large values of α , the deviation increases significantly with exceedingly small η values, thereby resulting to oversmoothing and exceptionally broad peaks appear in the distance distribution profile as shown in Figure 2-16H. The corresponding dipolar evolution curve in Figure 2-16I is under-fit, that is, the TKR fit neglects some of the oscillations that contribute to the signal and clearly does not adequately fit the data. The optimum regularization parameter refers to the transition between undersmoothing and oversmoothing, which often corresponds to the corner of the L-shaped curve, as indicated on Figure 2-16A. An optimum α value that results in a distance distribution profile and dipolar evolution curve shown in Figures 2-16D and

Figure 2-16E, respectively, is a compromise between smoothness of $P(r)$ and the quality of the fit of $F(t)$.

Background Subtraction and Self-Consistent Analysis

As previously mentioned, it is imperative to separate the form factor $F(t)$ from the background $B(t)$ in order to obtain an echo signal that accurately represents the intra-molecular spin-spin coupling and excludes signal contributed by intermolecular interactions. The intra-molecular spin-spin interactions have shorter distances, which correspond to the high frequency oscillations that last until decay time, T_{dd} (Figure 2-14). Good fits are typically obtained if the maximum dipolar evolution time fulfills the condition $t_{max} = 2T_{dd}$ (142). The background can then be determined from the portion of the curve at $t > \frac{t_{max}}{2}$. The DeerAnalysis2008 software (available online at <http://www.epr.ethz.ch/software/index>) used for DEER data analysis provides modules for subtracting background (145). One module fits the background to a simple polynomial of various degrees while the other fits the background to an exponential function corresponding to a homogeneous background with variable dimensions. The approximate Pake transformation (APT) module, provided by the software, gives a small range but good estimate of the correct background subtraction level.

Small variations in background subtraction can lead to small differences in the distance distribution profile (142). If background is fit before the dipolar modulation or evolution curve decays (overcorrection), longer distance distributions may be suppressed. If the background is assumed to decay slowly thus fitting it at a longer time (undercorrection), some spurious contributions at longer distances are introduced. Incorrect background subtraction can also lead to suppression of small peaks in the

distance profile that correspond to lowly populated states in the protein sample. Hence, it is necessary to identify the best level of background subtraction so that the most accurate distance distribution profile can be obtained.

The self-consistent analysis (SCA) method for obtaining the correct level of background subtraction was developed by our group. The steps involved in the analysis are presented in Figure 2-17. The first step is to generate the dipolar modulation or evolution curve and the distance distribution profile by Tikhonov regularization (TKR) using the recommended level of background subtraction provided by the DeerAnalysis2008 software. The distance distribution profile corresponds to the sum of the sub-populations or protein conformational states. The Gaussian reconstruction process determines the peak center, full width at half maximum (FWHM), and relative percentages of each sub-population. The values obtained are entered into the DeerSim software, which uses the Monte Carlo (MC) method to generate a theoretical dipolar modulation. The theoretical curve is then overlaid with the background subtracted echo curve or TKR fit. If the two curves do not match exactly, another background subtraction level has to be selected. The process is repeated until the DeerSim and TKR-generated dipolar evolution curves match within a certain error. The method for statistical analysis is currently being developed by another group member.

The process can also be applied to validate fine features present in the distance distribution profile. Minor populations in the distance profile can be suppressed using the DeerAnalysis software. The theoretical dipolar evolution curve generated is then overlaid with the experimental dipolar evolution curve. If the theoretical curve overlays with the experimental curve within the noise of the signal, the suppressed peak can be

regarded as unnecessary. If, however, the curves do not exactly match, then the suppressed peak contributes to the overall output and is considered as a sub-population. This population validation process is illustrated in Figure 2-18. Figures 2-18A and B show a sample distance distribution profile and the corresponding fit to Gaussian functions, respectively. Figure 2-18C corresponds to the overlaid experimental and theoretical modulations. Suppressing the peak at 30.5 Å results in a change in the theoretical dipolar evolution curve, which does not overlay exactly with the experimental curve (Figure 2-18D). The theoretical dipolar modulation that arises from suppressing the wide-open peak also does not overlay well with the experimental dipolar evolution curve (Figure 2-18E). In certain cases, suppressing two peaks results in cancellation of the dipolar modulations, that is, the theoretical and experimental curves match as a result of suppressing both peaks. The example in Figure 2-18F shows that suppressing the tucked/curled peak at 30.5 Å and the wide open peak leads to a significant alteration in the dipolar evolution curves. This implies that both tucked/curled and wide-open conformations contribute to the overall distance distribution profile.

Gaussian Reconstruction

The process of Gaussian reconstruction involves generating the sub-populations from the distance distribution profile. In this method, the TKR-generated distance profile is fit to a series of Gaussian functions that sum up to regenerate the original profile. Originally, brute force manual fitting was used to obtain the individual sub-populations. This was done by estimating values of distance, FWHM, and relative percentages to input into DeerSim and thus obtaining a distance profile that fits the original profile. Another method uses the Origin 8.0 software and its later versions to generate

Gaussian functions that fit into the original profile through the peak-finding capabilities of the software. The advantage of this method is that hidden or overlapping peaks can be identified by taking the second-derivative of the distance profile. The peak local minima correspond to the individual Gaussian populations.

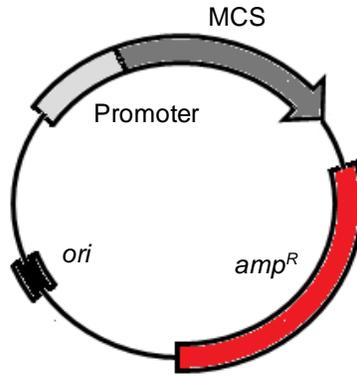


Figure 2-1. Example of vector map containing the necessary components of a plasmid including the origin of replication or *ori* (black); *amp^R*, an antibiotic resistance gene (red); multiple cloning site or MCS (dark gray); and promoter (light gray).

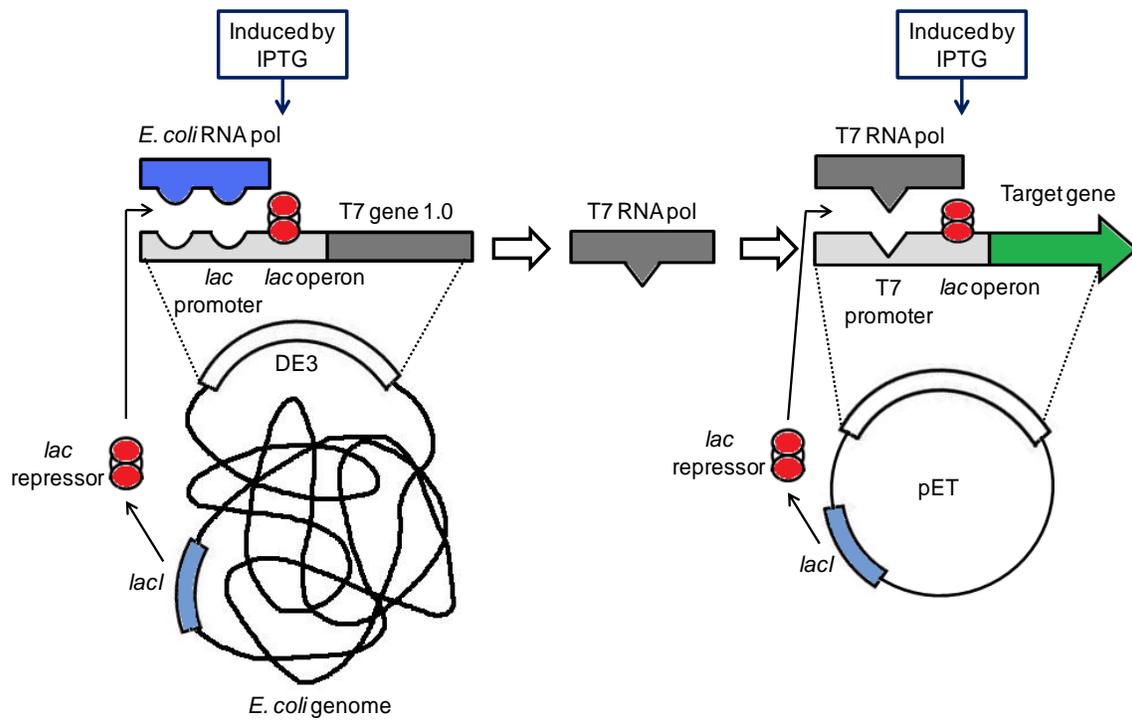


Figure 2-2. Host and vector elements essential in the overexpression of target protein. Figure adapted from Novagen (114).

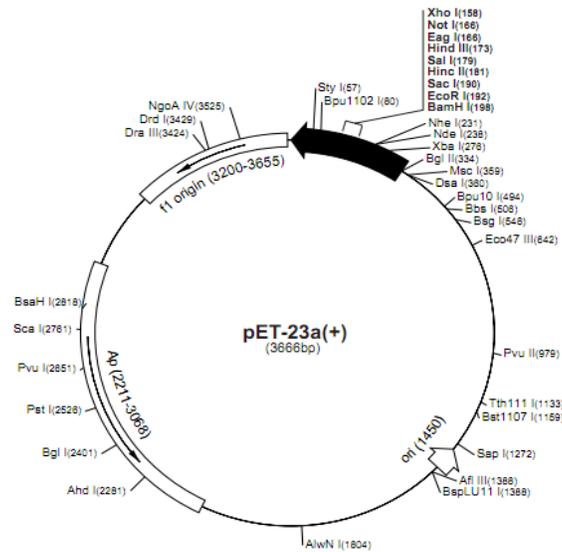


Figure 2-3. pET23a vector map. The heavily shaded black arrow refers to the T7 promoter region, which also contains the site for insertion of gene of interest. Figure courtesy of Novagen (<http://www.emdchemicals.com/life-science-research/>).

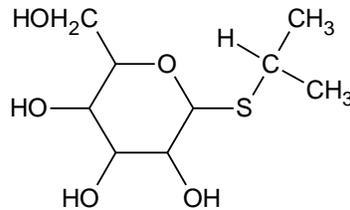


Figure 2-4. Structure of isopropyl-β-D-1-thiogalactopyranoside (IPTG).

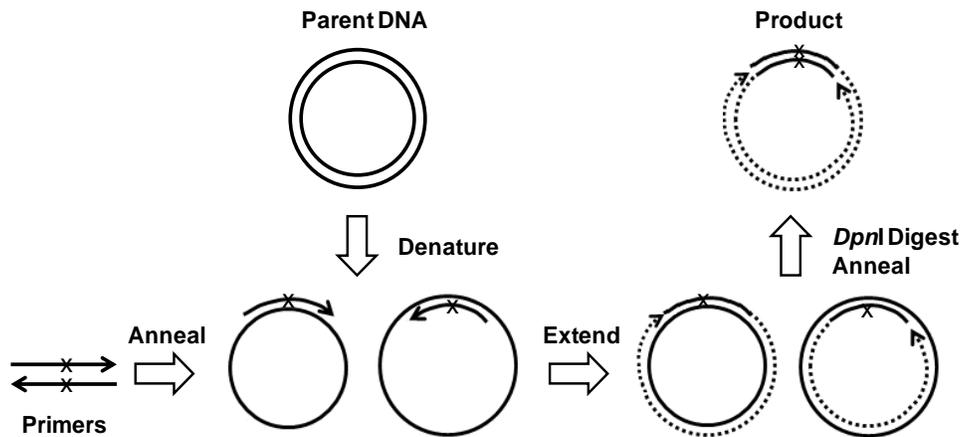


Figure 2-5. Various steps in site-directed mutagenesis, where x represents the mutation. Figure adapted from Zheng et al. (119).

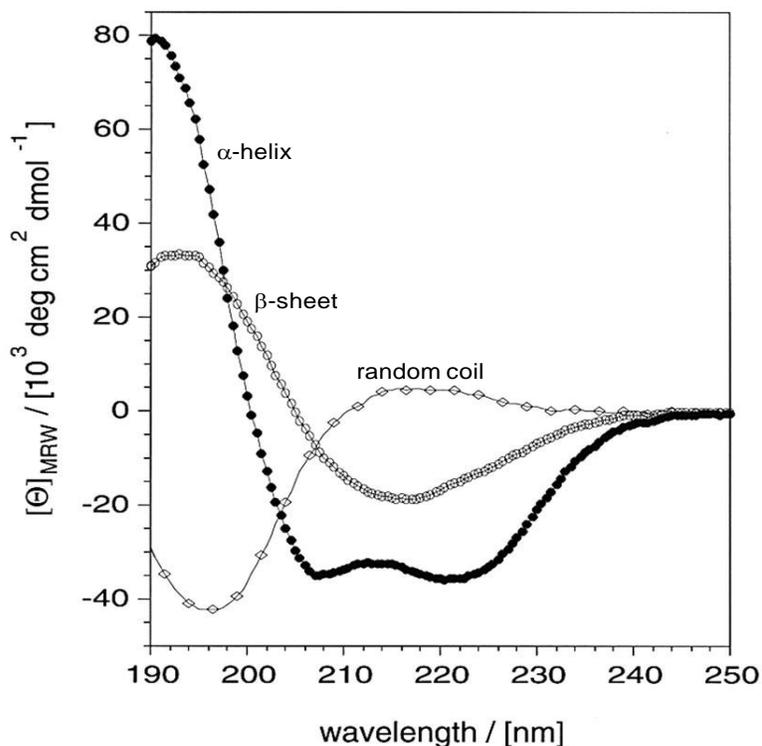


Figure 2-6. Typical circular dichroism (CD) spectra of α -helix (filled circles), β -sheet (open circles), and random coil (open diamonds) structures of protein. Reprinted by permission from Macmillan Publishers Ltd: *The EMBO Journal* Fändrich (125). Copyright 2002.

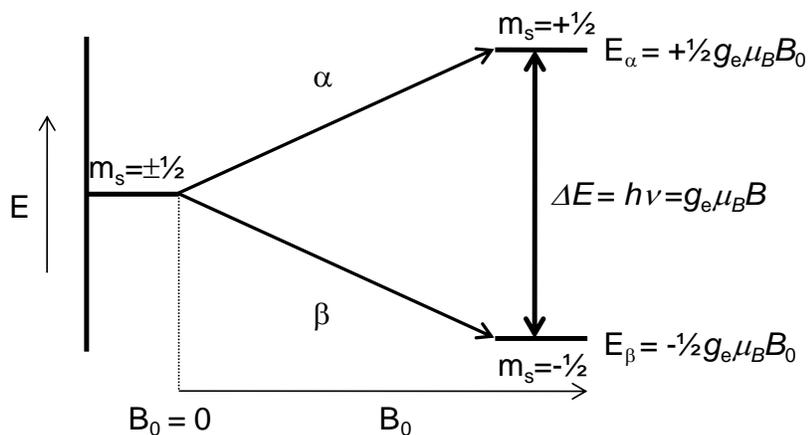


Figure 2-7. Energy diagram representing the Zeeman Effect.

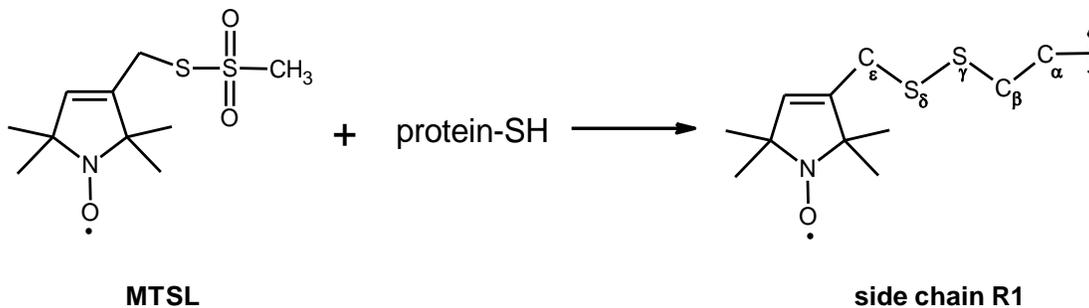


Figure 2-8. Chemical modification of the protein side chain (cysteine) using MTSL as spin label.

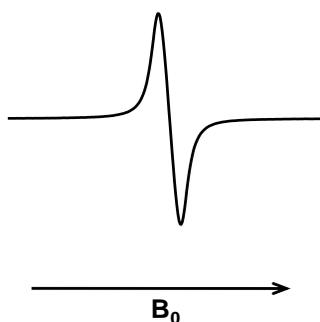


Figure 2-9. Typical EPR spectrum of a system with an electron spin state of $m_s = \pm 1/2$.

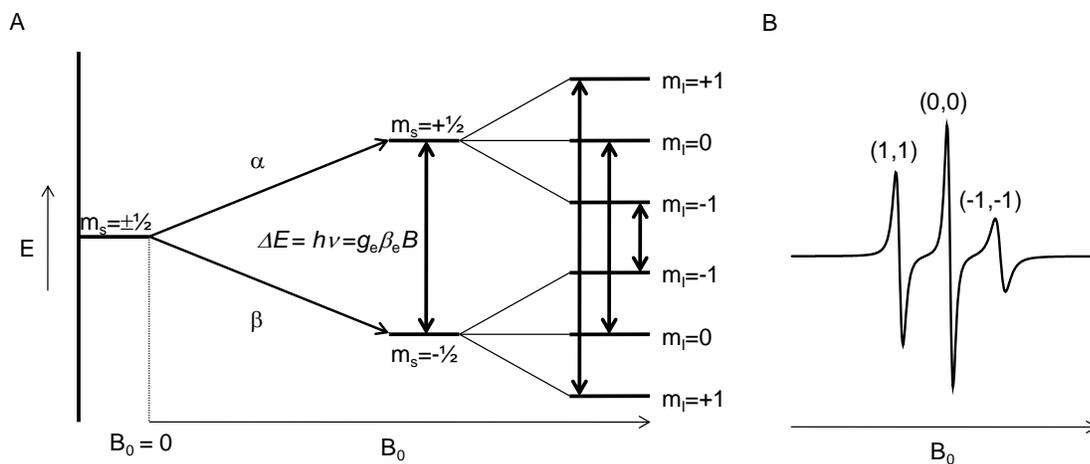


Figure 2-10. A) Energy diagram and B) example of EPR spectrum of a system with an electron spin state of $m_l = 0, \pm 1$.

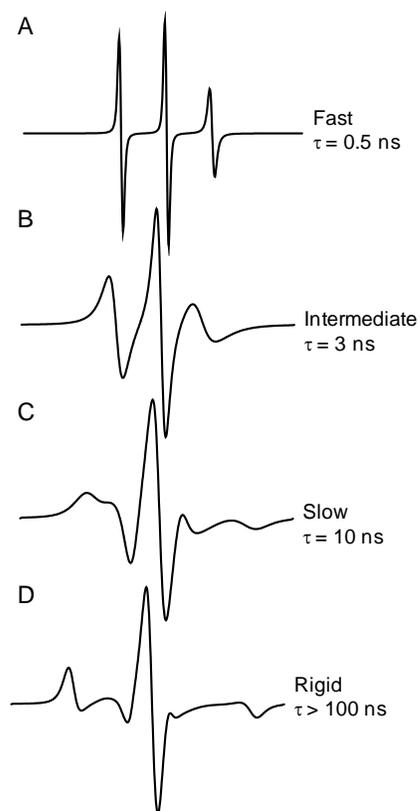


Figure 2-11. EPR line shapes representing various modes of spin motion.

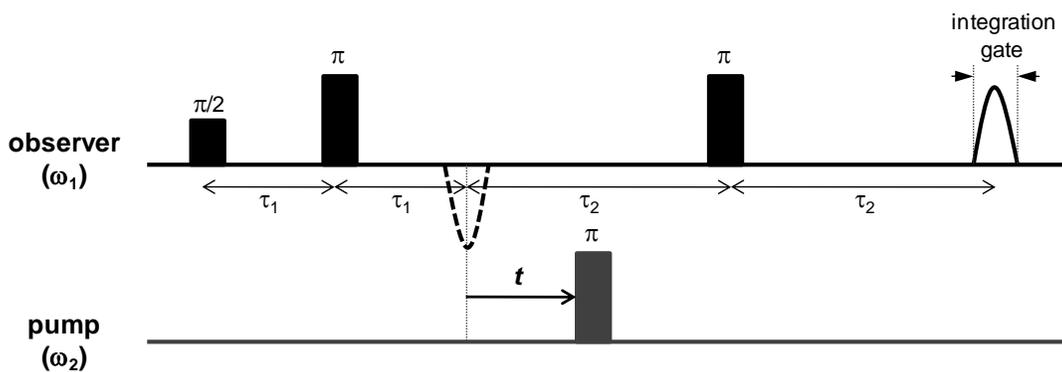


Figure 2-12. Pulse sequence of the four-pulse DEER experiment. τ_1 and τ_2 represents the fixed delay times and the time t between the inverted electron spin echo (dotted line) and the pump π pulse is varied. The intensity of the refocused echo is integrated as shown. Figure adapted from Jeschke (142).

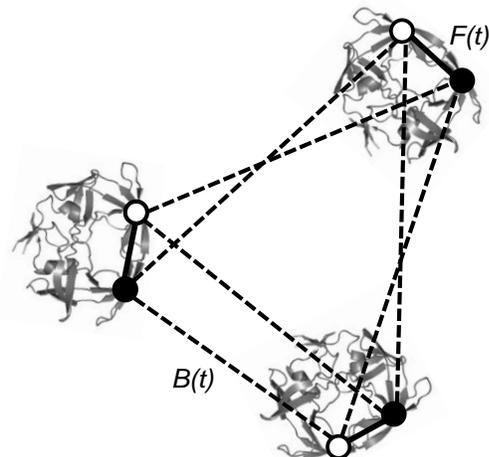


Figure 2-13. Illustration of doubly labeled HIV-1 protease showing the intramolecular $F(t)$ (solid lines) and intermolecular $B(t)$ (dotted line) interactions. Figure adapted from Jeschke (142).

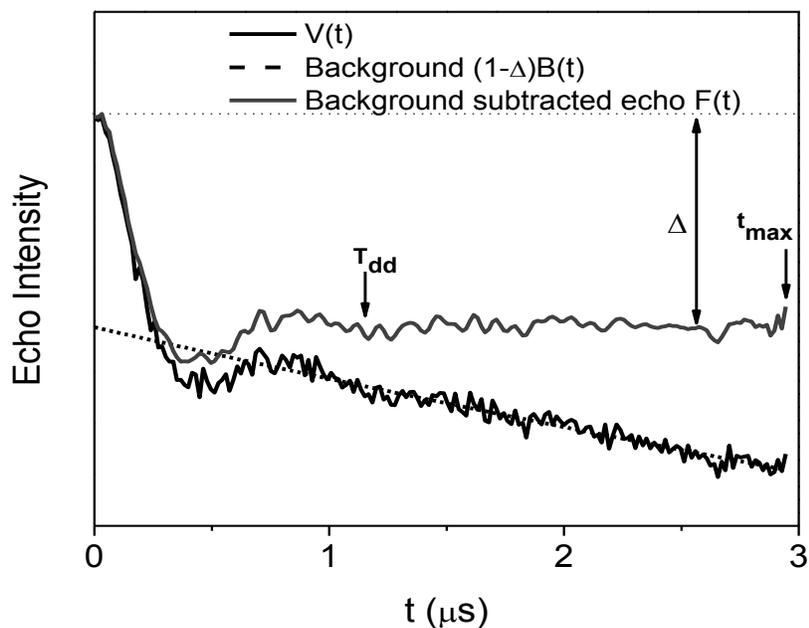


Figure 2-14. Plot of the experimental dipolar evolution function $V(t)$ (black), background factor $B(t)$ (black dotted line), and form factor $F(t)$ (gray). T_{dd} represents the echo decay time, t_{max} refers to the maximum time for collection of echo data, and Δ represents the modulation depth.

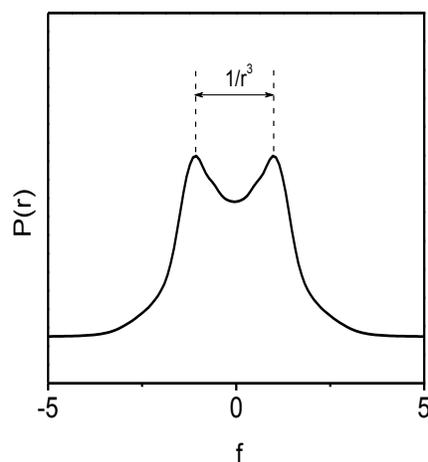


Figure 2-15. Pake pattern obtained by Fourier transformation of the dipolar evolution curve.

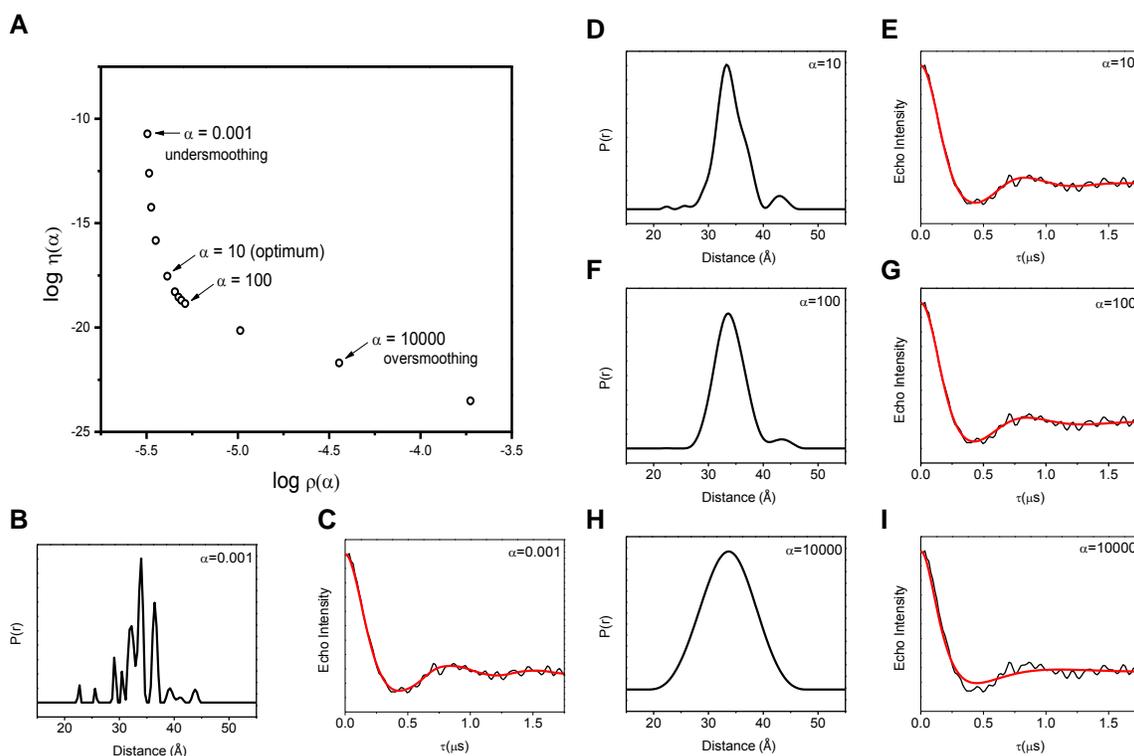


Figure 2-16. Example of an A) L-curve and corresponding distance distribution profiles and background subtracted echo (black) overlaid with the TKR-generated dipolar modulation curves (red) of B) and C) undersmoothed ($\alpha=0.001$), D) and E) optimal ($\alpha=10$), F) and G) $\alpha=100$, and H) and I) oversmoothed ($\alpha=10000$) α values.

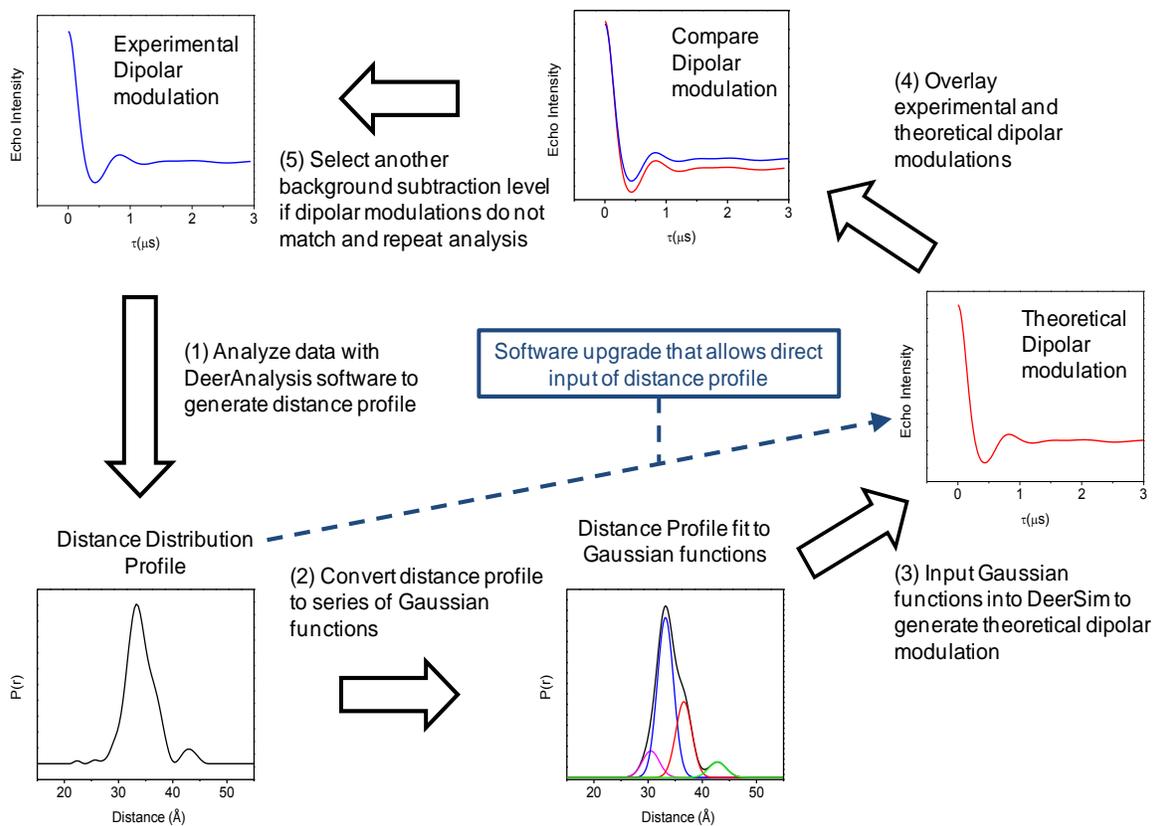


Figure 2-17. Flow-chart of the Self-Consistent Analysis (SCA) method for optimizing background subtraction of the dipolar modulation curve that generates the distance distribution profile.

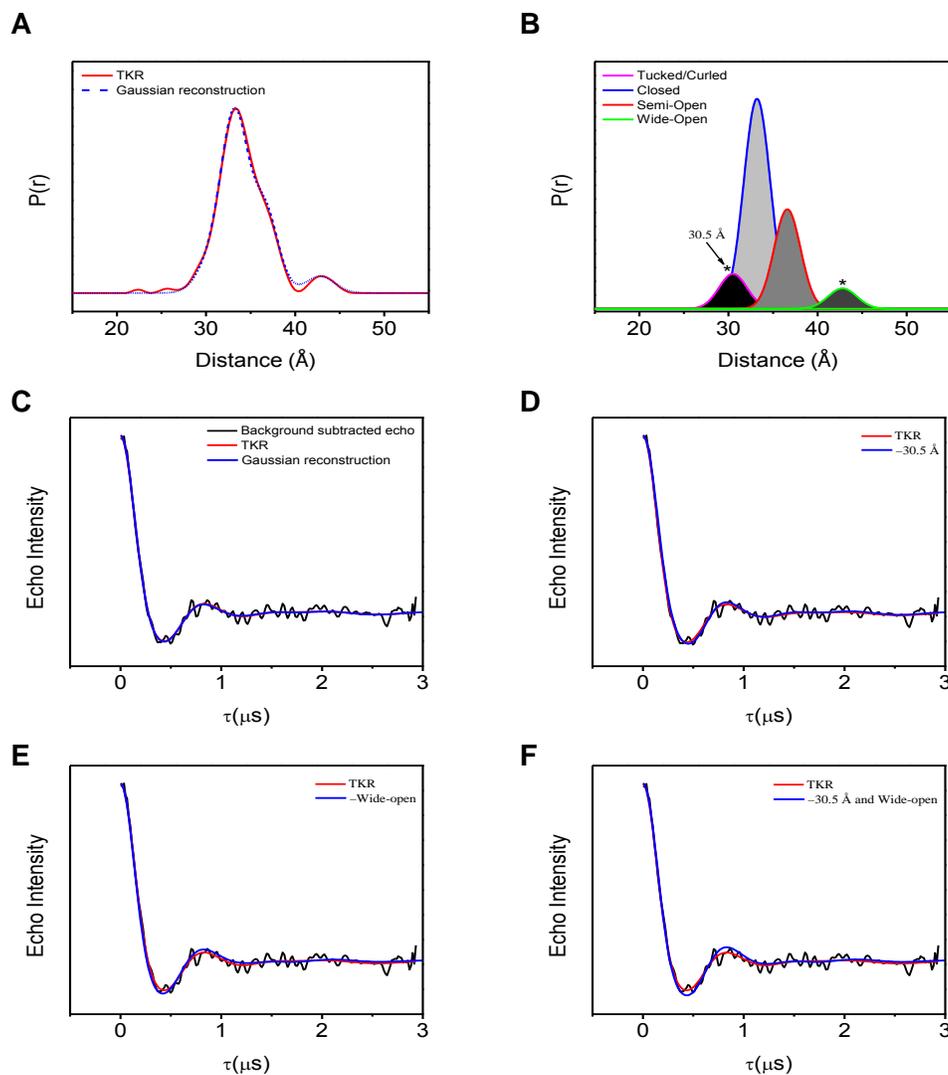


Figure 2-18. Example of population validation process for verifying minor populations. A) Distance distribution profile from TKR analysis (red) overlaid with the sum of the populations obtained by Gaussian reconstruction (blue dashed), B) individual Gaussian populations (suppressed populations are indicated by an asterisk), and background subtracted dipolar echo curve (black) overlaid with the TKR-generated dipolar modulation (red) and Gaussian reconstructed dipolar modulation (blue) generated C) without subtracting any population and D-F) with suppression of populations.

CHAPTER 3 CLONING AND MUTAGENESIS

Materials and Methods

Chemicals, reagents, and supplies were obtained from Fisher Scientific (Pittsburgh, Pennsylvania) unless otherwise indicated. pET23a DNA was purchased from Novagen (Gibbstown, New Jersey). *E. coli* codon-optimized HIV-1 protease DNA was purchased from DNA 2.0 (Menlo Park, California). Restriction enzymes (*Nde*I and *Bam*HI), T4 DNA ligase, and 1 kb standard DNA ladder were purchased from New England Biolabs (Ipswich, Massachusetts). The QuikChange™ site-directed mutagenesis kit was purchased from Stratagene (Santa Clara, California). XL1-Blue *E. coli* cells were purchased from Invitrogen (Carlsbad, California). The QIAquick® gel extraction kit and QIAprep® spin miniprep kit were purchased from Qiagen (Valencia, California).

Cloning of HIV-1 PR PRE and POST Constructs

The *Escherichia coli* (*E. coli*) codon- and expression-optimized genes for HIV-1 protease Pre_i-K55C and Post_i-K55C constructs were received as pJ201:29193 (Figure 3-1A) and pJ201:29194 (Figure 3-1B) vectors, respectively. The corresponding DNA and amino acid sequences of the HIV-1 PR constructs are given in Tables 3-1 and 3-2. Plasmids containing the Pre_i-K55C and Post_i-K55C and pET23a vector were digested by incubating the vectors with *Nde*I and *Bam*HI in the appropriate buffer at 37°C for approximately 2 hours. The digests were run through a 1% agarose gel and visualized using a UV transilluminator. The desired bands were excised from the gel and DNA was extracted using a QIAquick® gel extraction kit using the provided protocol. The Pre_i-K55C and Post_i-K55C gene inserts were ligated into the pET23a vector by

incubating DNA in appropriate proportions with T4 DNA ligase at room temperature for approximately 1 h. The ligation reaction mixture was transformed into XL1-blue *E. coli* cells via standard heat-shock method. The ligated vector was isolated and purified using the QIAprep® spin miniprep kit. The plasmid DNA obtained was submitted for DNA sequencing for confirmation. In this work, pET23_Pre_i-K55C (pET23_PRE) and pET23_Post_i-K55C (pET23_POST) refers to the DNA while Pre_i-K55C (PRE) and Post_i-K55C (POST) refers to the protease. This naming convention also applies to the other constructs (Table 3-3).

Site-Directed Mutagenesis of HIV-1 PR Constructs

To obtain each of the constructs containing drug-pressure selected mutations (Table 1-5 and Table 3-3), several rounds of site-directed mutagenesis were performed on pET23_POST. The DNA and amino acid sequences of the mutant constructs are given in Tables 3-4 to 3-7, where the sites of mutation are highlighted. The pET23_Post_i-A82V-K55C construct (pET23_PMUT1) was obtained by mutation of alanine at position 82 to valine using pET23_Post_i-K55C as the template. pET23_Post_i-A82V-I10L-V15I-K55C (pET23_PMUT3) was obtained by performing successive mutations on the pET23_PMUT1. First, isoleucine at position 10 was converted to leucine and then the product pET23_Post_i-A82V-I10L-K55C intermediate mutant was used as a template to form the pET23_PMUT3 construct. The same mutagenesis strategy was performed for the pET23_Post_i-A54I-E58Q-K55C (pET23_PMUT2) and pET23_Post_i-A54I-E58Q-Q34E-I36M-N37T-K55C (pET23_PMUT5) constructs. The sequences, melting temperatures (T_m), and molecular weights of the primers used for each mutant are provided in Appendix A (Tables A-1 to A-4).

An additional mutant construct, B_{si}-I63P, was obtained by substituting isoleucine at position 63 of B_{si} with proline. This corresponds to the DNA mutation, **ATT**→**CCT**, where the mutation is highlighted (Table 3-9). As explained in Chapter 1, the amino acid residue at position 63 of B_{si}, the stabilized, inactive form of HIV-1 PR subtype B LAI, was converted from leucine to isoleucine as it is a putative cleavage site for HIV-1 PR autoproteolysis. The DNA and amino acid sequences of B_{si} and B_{si}-I63P are given in Tables 3-8 and 3-9, respectively. The primer sequence and parameters used for mutagenesis of pET23_B_{si}-I63P are provided in Table A-5 of Appendix A.

Site-directed mutagenesis was carried out using the QuikChange™ site-directed mutagenesis kit. The primers were designed using the PrimerX software (available online at <http://www.bioinformatics.org/primerx/>). The template DNA was mixed with the primers, deoxynucleotide triphosphate (dNTP) mixture, *PfuUltra* polymerase, polymerase buffer and sterile, nuclease-free water in the specific order and proportions optimized for HIV-1 PR as listed in Table A-6 of Appendix A. A separate mixture containing 5% (v/v) dimethylsulfoxide (DMSO) was also prepared. The mixtures were subjected to polymerase chain reaction (PCR) in an Eppendorf (Hauppauge, New York) thermocycler for 3 hours. The thermal cycling parameters used are shown in Table A-7 of Appendix A. After PCR, *DpnI* was added to the mixture and incubated for another 1 to 2 h at 37 °C. The PCR product was transformed into XL1-blue *E. coli* cells via standard heat-shock method. The mutant DNA was isolated and purified using the QIAprep® spin miniprep kit and submitted for DNA sequencing for confirmation.

Results and Discussion

Cloning

The cloning process involves the insertion of target DNA into a plasmid that is suitable for a chosen bacterial expression strain. The major requirement is for the DNA insert and plasmid to have restriction enzyme sites that are complementary for the ligation process to work. After ligation, DNA is transformed into *E. coli* cells and selectively grown in ampicillin-containing media. Cells that survive should have plasmid containing the target DNA. The success of ligation can also be verified by cleavage using an appropriate restriction enzyme followed by agarose gel electrophoresis, which provides an estimate of the size of DNA by comparison with a DNA standard ladder. The ligated DNA can also be compared with a reference plasmid DNA containing an insert of the same size.

The theoretical size of the pET23 vector is 3666 base pairs (bp) or 3.6 kilobase (kb) while the DNA inserts, which include the restriction enzyme sites, have 315 bp each. Thus, the DNA gel for the successfully ligated plasmid should contain a band that corresponds to about 4 kb. Figure 3-2 shows the DNA gels of the pET23_PRE and pET23_POST. Samples labeled 1A, 1B, and 1C correspond to DNA isolated from different colonies in one LB agar plate whereas samples 2A, 2B, and 2C represent DNA from different colonies in another plate. The gel for both plasmids possesses 2 bands, an upper band that corresponds to roughly 5.5 kb and a lower band, which is estimated to be about 2.5 kb. This is expected as the reference pET23a plasmid containing a HIV-1 PR DNA insert (Figure 3-2, lane 1) also has 2 bands at the same position. This result is not surprising because gel electrophoresis does not separate molecules only according to size but also based on shape. DNA molecules that are exactly alike can

travel through the gel at different rates depending on the topological parameters of the molecule, e.g. degree of supercoiling. Since the plasmid is a closed circular DNA, there is a large possibility for the DNA to twist into a supercoiled conformation. The rate of migration across the gel of circular, linear, and supercoiled DNA increases, respectively. Supercoiled DNA, which is a more compact molecule, can penetrate through the mesh of the gel unlike circular DNA. Overall, gel electrophoresis is useful to verify the presence of DNA and comparison with a reference HIV-1 PR DNA provides evidence of the success of ligation.

DNA sequencing was necessary to confirm whether the DNA insert having the correct sequence was incorporated into the pET23a vector. Results for DNA sequencing revealed that for pET23_PRE, only bands on lanes 4, 6 and 7 (Figure 3-2A) had the correct sequence. The DNA sequence of bands on lanes 2 and 3 correspond to the sequence of the original plasmid. This indicates that the initial step involving the excision of the pET23 plasmid containing the original DNA insert was unsuccessful. The original plasmids were most likely not removed in solution after the succeeding steps in the cloning process and were then transformed into the *E. coli* cells. These formed as separate colonies in the LB agar medium. DNA sequencing analysis of the DNA corresponding to lane 5 did not produce a signal, indicating that the plasmid was not successfully ligated. For pET23_POST, the DNA sequence of bands that correspond to lanes 3, 4, and 6 (Figure 3-2B) were correct. Bands corresponding to lanes 2 and 5 were found to have the DNA sequence of the original plasmid whereas DNA from lane 7 did not generate a signal. Cloning of PRE and POST DNA into the pET23a vector was partially successful in that not all of the DNA were cloned and ligated. Nevertheless, the

successfully cloned pET23_PRE and pET23_POST were used for expression of HIV-1 PR in the bacterial strain, BL21(DE3)pLysS *E. coli* cells.

Mutagenesis

In this work, site-directed mutagenesis was utilized to modify codons in HIV-1 PR DNA and study the effect of specific mutations on protein structure and function. The QuikChange site-directed mutagenesis kit was used because it is a simple, fast, and efficient method. However, it is imperative to select the correct primers and proportions of components in the PCR mixture to avoid failure in obtaining a PCR product.

To verify if DNA amplification is successful, gel electrophoresis of the PCR product was performed. Figure 3-3 shows a gel of the PCR products of two separate runs (A and B) of the same HIV-1 PR DNA sample. The bands that appear at about 4 kb in lanes 2 and 3 represent plasmid DNA containing nicks, which is the expected position for a linear pET23a plasmid with the HIV-1 PR DNA insert. It is noticeable that these bands are not present in lanes 4 and 5 indicating that PCR failed to amplify DNA. As expected, the control PCR mixture on lane 6 does not contain the ~4 kb band. The bands at the bottom of the gel correspond to the primers.

After PCR, the ligation mixture is transformed into XL-1 Blue *E. coli* cells, which is chosen because it is capable of sealing nicks in the plasmid DNA and also enhances the stability of the DNA insert (146). Subsequent growth on ampicillin-containing media allows the selection of cells that have taken up plasmid possessing the HIV-1 PR DNA insert. However, cells that survive do not necessarily contain the plasmid DNA with the correct sequence. Thus, it is imperative to run DNA sequencing on the isolated DNA.

In this work, the following mutants were constructed: 1) pET23_PMUT1, 2) pET23_PMUT3, 3) pET23_PMUT2, and 4) pET23_PMUT5 DNA using pET23_POST as

the template DNA. A single point DNA mutation was performed for each round of mutagenesis with the exception of the pET23_Post_i-A54I-K55C intermediate mutant. DNA gels were obtained for each mutant, wherein samples 1A and 1B represent DNA isolated from different colonies in one LB agar plate and samples 2A and 2B correspond to DNA from different colonies in another plate.

For pET23_PMUT1, the codon for alanine (A82) in pET23_POST was replaced with one that codes for valine (G**CA**→GTA, where the mutation is highlighted, Table 3-4). The DNA gel of pET23_PMUT1 is shown in Figure 3-4. The bands corresponding to lanes 3 and 4, which were subjected to DNA sequencing, have the correct sequence.

Two rounds of mutagenesis were performed on pET23_PMUT1 to construct pET23_PMUT3. The intermediate mutant, pET23_Post_i-A82V-I10L-K55C, was obtained by converting isoleucine to leucine (A**TT**→C**TT**, where the mutation is highlighted, Table 3-5) at position 10 in the protein sequence. This mutant was then used as the template for converting valine (V15) to isoleucine (G**TT**→A**TT**, where the mutation is highlighted, Table 3-5). The DNA gel of pET23_PMUT3 is shown in Figure 3-5. DNA corresponding to bands on lane 2 and 5 were sequenced. Results revealed that the band at lane 5 did not have the correct sequence.

pET23_PMUT2 was obtained by performing two rounds of mutagenesis. Instead of the usual single point DNA mutation, pET23_Post_i-A54I-K55C was initially constructed by replacing two DNA base pairs in the codon for alanine (A54) to convert into isoleucine (G**CC**→A**TC**, where the mutation is highlighted, Table 3-6). Once produced, the resultant DNA was used as the template to construct pET23_PMUT2. This was carried out by changing glutamic acid at position 58 to glutamine (G**AG**→C**AG**,

where the mutation is highlighted, Table 3-6). The DNA gel of pET23_PMUT2 is shown in Figure 3-6. DNA corresponding to bands on lanes 2-4 were sequenced and shown to have the correct sequence.

pET23_PMUT2 was used as the template for constructing pET23_PMUT5. Three rounds of mutagenesis were carried out. The first round involves switching the glutamine residue at position 34 with glutamic acid (**CAG**→**GAG**, where the mutation is highlighted, Table 3-7) to form the pET23_Post_i-A54I-E58Q-Q34E-K55C intermediate mutant. This was followed by the conversion of isoleucine (I36) to methionine (**ATC**→**ATG**, where the mutation is highlighted, Table 3-7) to produce pET23_Post_i-A54I-E58Q-Q34E-I36M-K55C. Finally, pET23_PMUT5 was obtained by substituting asparagine (N37) with threonine (**AAC**→**ACC**, where the mutation is highlighted, Table 3-7). The DNA gel of pET23_PMUT5 is shown in Figure 3-7. The bands on lanes 2 and 4 have the correct DNA sequence.

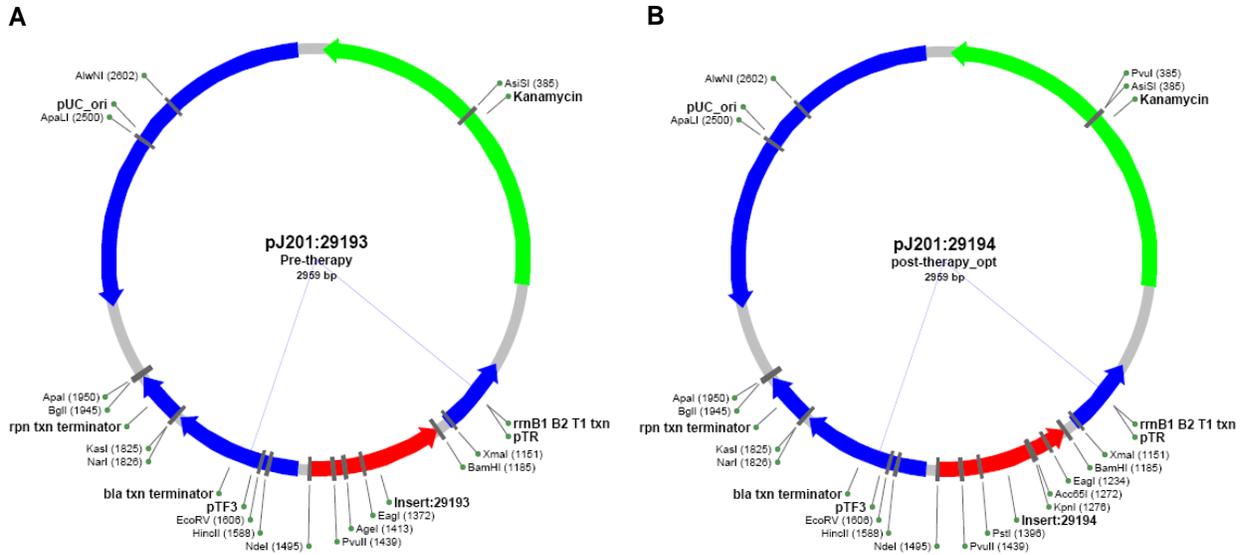


Figure 3-1. Map of original vector containing the A) Pre_i-K55C and B) Post_i-K55C HIV-1 PR DNA inserts (red).

Table 3-1. *E. coli* codon-optimized DNA and amino acid sequence of PRE.

ccg cag att acc ttg tgg caa cgt ccg ctg gtt acg att aag att ggt ggt cag ctg aaa	10	20
P Q I T L W Q R P L V T I K I G G Q L K		
gaa gct ctg ctg aac acc ggt gcc gac gac acg gtg ctg gaa gag atg acc ctg acc ggc	30	40
E A L L N T G A D D T V L E E M T L T G		
cgt tgg aag ccg aaa atg atc ggc ggc att ggt ggc ttt atc tgc gtt cgt caa tac gat	50	60
R W K P K M I G G I G G F I C V R Q Y D		
caa gtc ccg atc gag att gca ggt cac aaa gcg att ggc act gtt ctg gtg ggt cca acg	70	80
Q V P I E I A G H K A I G T V L V G P T		
ccg gtc aat atc atc ggt cgc aat ctg ctg acc cag atc ggt gcg acc ttg aac ttc	90	
P V N I I G R N L L T Q I G A T L N F		

Table 3-2. *E. coli* codon-optimized DNA and amino acid sequence of POST.

ccg caa atc acg ctg tgg caa cgt ccg att gtc acc atc aaa gtt ggc ggt cag ctg aaa	10	20
P Q I T L W Q R P I V T I K V G G Q L K		
gag gca ctg ctg aat acg ggt gcg gac gat acc gtt ctg cag gaa atc aac ttg acg ggt	30	40
E A L L N T G A D D T V L Q E I N L T G		
cgc tgg aaa ccg aag atg att ggc ggc atc ggc ggt ttt gcc tgc gtc cgt gag tac gac	50	60
R W K P K M I G G I G G F A C V R E Y D		
cag gtg ccg att gaa att gct ggt cac aag gcg att ggt acc gtt ttg gtg ggt cca act	70	80
Q V P I E I A G H K A I G T V L V G P T		
ccg gca aac att atc ggc cgt aat ctg ctg acc caa atc ggt gcg acc ctg aac ttc	90	
P A N I I G R N L L T Q I G A T L N F		

Table 3-6. *E. coli* codon-optimized DNA and amino acid sequence of PMUT2. Sites of mutation are in bold.

ccg	caa	atc	acg	ctg	tgg	caa	cgt	ccg	att	gtc	acc	atc	aaa	gtt	ggc	ggt	cag	ctg	aaa	10	20
P	Q	I	T	L	W	Q	R	P	I	V	T	I	K	V	G	G	Q	L	K		
30																				40	
gag	gca	ctg	ctg	aat	acg	ggt	gcg	gac	gat	acc	gtt	ctg	cag	gaa	atc	aac	ttg	acg	ggt		
E	A	L	L	N	T	G	A	D	D	T	V	L	Q	E	I	N	L	T	G		
50																				60	
cgc	tgg	aaa	ccg	aag	atg	att	ggc	ggc	atc	ggc	ggt	ttt	atc	tgc	gtc	cgt	cag	tac	gac		
R	W	K	P	K	M	I	G	G	I	G	G	F	I	C	V	R	Q	Y	D		
70																				80	
cag	gtg	ccg	att	gaa	att	gct	ggt	cac	aag	gcg	att	ggt	acc	gtt	ttg	gtg	ggt	cca	act		
Q	V	P	I	E	I	A	G	H	K	A	I	G	T	V	L	V	G	P	T		
90																					
ccg	gca	aac	att	atc	ggc	cgt	aat	ctg	ctg	acc	caa	atc	ggt	gcg	acc	ctg	aac	ttc			
P	V	N	I	I	G	R	N	L	L	T	Q	I	G	A	T	L	N	F			

Table 3-7. *E. coli* codon-optimized DNA and amino acid sequence of PMUT5. Sites of mutation are in bold.

ccg	caa	atc	acg	ctg	tgg	caa	cgt	ccg	att	gtc	acc	atc	aaa	gtt	ggc	ggt	cag	ctg	aaa	10	20
P	Q	I	T	L	W	Q	R	P	I	V	T	I	K	V	G	G	Q	L	K		
30																				40	
gag	gca	ctg	ctg	aat	acg	ggt	gcg	gac	gat	acc	gtt	ctg	gag	gaa	atg	acc	ttg	acg	ggt		
E	A	L	L	N	T	G	A	D	D	T	V	L	E	E	M	T	L	T	G		
50																				60	
cgc	tgg	aaa	ccg	aag	atg	att	ggc	ggc	atc	ggc	ggt	ttt	atc	tgc	gtc	cgt	cag	tac	gac		
R	W	K	P	K	M	I	G	G	I	G	G	F	I	C	V	R	Q	Y	D		
70																				80	
cag	gtg	ccg	att	gaa	att	gct	ggt	cac	aag	gcg	att	ggt	acc	gtt	ttg	gtg	ggt	cca	act		
Q	V	P	I	E	I	A	G	H	K	A	I	G	T	V	L	V	G	P	T		
90																					
ccg	gca	aac	att	atc	ggc	cgt	aat	ctg	ctg	acc	caa	atc	ggt	gcg	acc	ctg	aac	ttc			
P	V	N	I	I	G	R	N	L	L	T	Q	I	G	A	T	L	N	F			

Table 3-8. *E. coli* codon-optimized DNA and amino acid sequence of B_{si}.

cca	caa	atc	act	ctg	tgg	aaa	cgt	ccg	ctg	gtc	acc	att	aaa	att	ggc	ggt	caa	ctg	aaa	10	20
P	Q	I	T	L	W	K	R	P	L	V	T	I	K	I	G	G	Q	L	K		
30																				40	
gaa	gcg	ctg	ctg	aac	acc	ggt	gca	gat	gat	acc	gtt	atc	gag	gaa	atg	agc	ctg	ccg	ggt		
E	A	L	L	N	T	G	A	D	D	T	V	I	E	E	M	S	L	P	G		
50																				60	
cgt	tgg	aaa	cct	aaa	atg	att	ggc	ggt	att	ggt	ggt	ttc	att	tgt	gtg	cgc	cag	tac	gac		
R	W	K	P	K	M	I	G	G	I	G	G	F	I	C	V	R	Q	Y	D		
70																				80	
cag	atc	att	atc	gaa	atc	gcc	ggc	cac	aag	gca	att	ggt	acc	gtg	ctg	ggt	ggc	ccg	acc		
Q	I	I	I	E	I	A	G	H	K	A	I	G	T	V	L	V	G	P	T		
90																					
ccg	gtt	aac	atc	atc	ggc	cgc	aac	ctg	ctg	act	cag	att	ggc	gcc	acg	ctg	aac	ttc			
P	V	N	I	I	G	R	N	L	L	T	Q	I	G	A	T	L	N	F			

Table 3-9. *E. coli* codon-optimized DNA and amino acid sequence of B_{si}-I63P. Site of mutation is in bold.

10										20									
cca	caa	atc	act	ctg	tgg	aaa	cgt	ccg	ctg	gtc	acc	att	aaa	att	ggc	ggt	caa	ctg	aaa
P	Q	I	T	L	W	K	R	P	L	V	T	I	K	I	G	G	Q	L	K
30										40									
gaa	gcg	ctg	ctg	aac	acc	ggt	gca	gat	gat	acc	ggt	atc	gag	gaa	atg	agc	ctg	ccg	ggt
E	A	L	L	N	T	G	A	D	D	T	V	I	E	E	M	S	L	P	G
50										60									
cgt	tgg	aaa	cct	aaa	atg	att	ggc	ggt	att	ggt	ggt	ttc	att	tgt	gtg	cgc	cag	tac	gac
R	W	K	P	K	M	I	G	G	I	G	G	F	I	C	V	R	Q	Y	D
70										80									
cag	atc	cc t	atc	gaa	atc	gcc	ggc	cac	aag	gca	att	ggt	acc	gtg	ctg	gtt	ggc	ccg	acc
Q	I	I	I	E	I	A	G	H	K	A	I	G	T	V	L	V	G	P	T
90																			
ccg	gtt	aac	atc	atc	ggc	cgc	aac	ctg	ctg	act	cag	att	ggc	gcc	acg	ctg	aac	ttc	
P	V	N	I	I	G	R	N	L	L	T	Q	I	G	A	T	L	N	F	

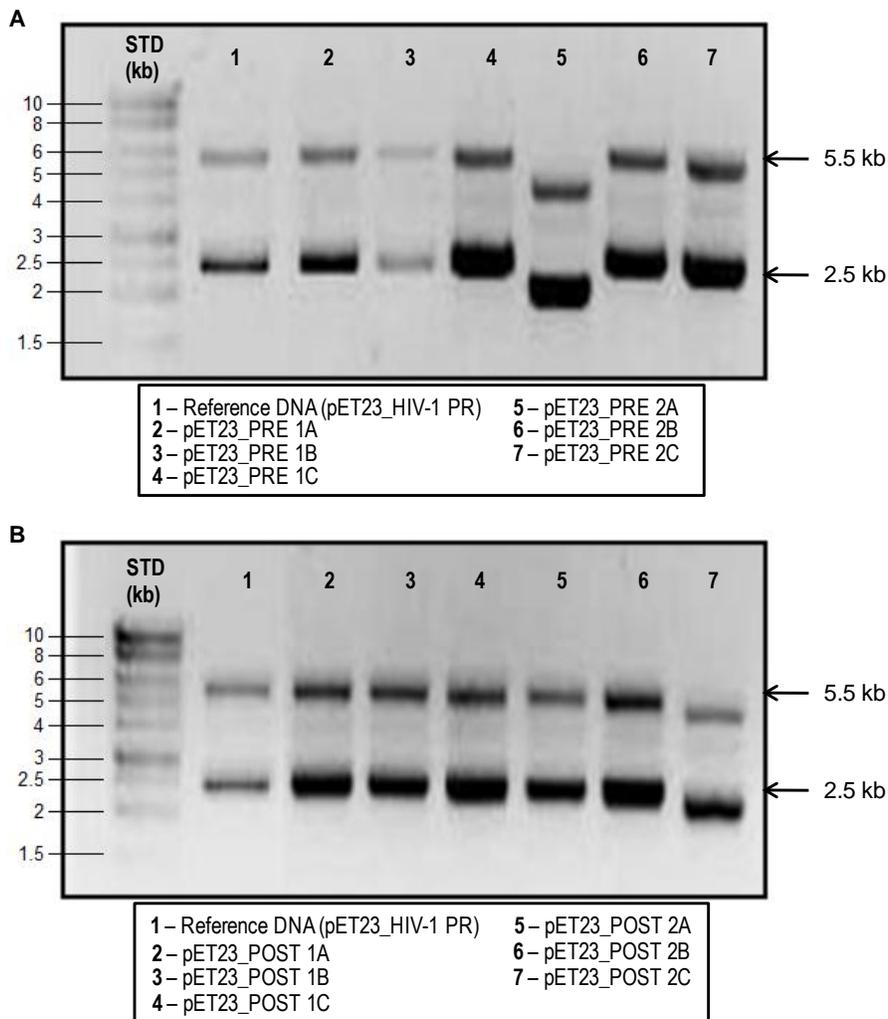


Figure 3-2. DNA gel of pET23a plasmid containing A) PRE and B) POST HIV-1 PR DNA. The STD lane corresponds to NEB 1 kb standard DNA ladder.

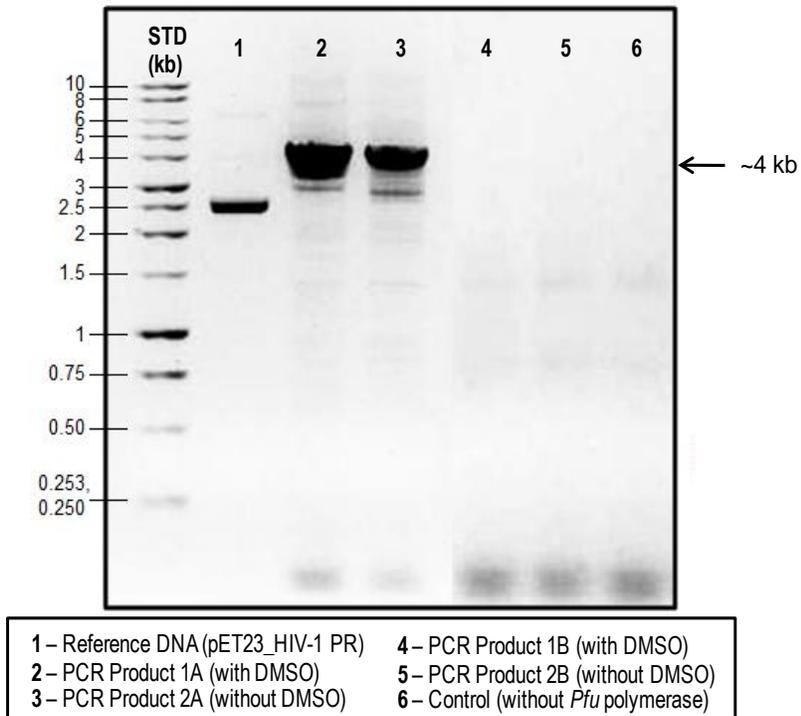


Figure 3-3. Sample DNA gel after *DpnI* digestion. The STD lane corresponds to NEB 1 kb standard DNA ladder.

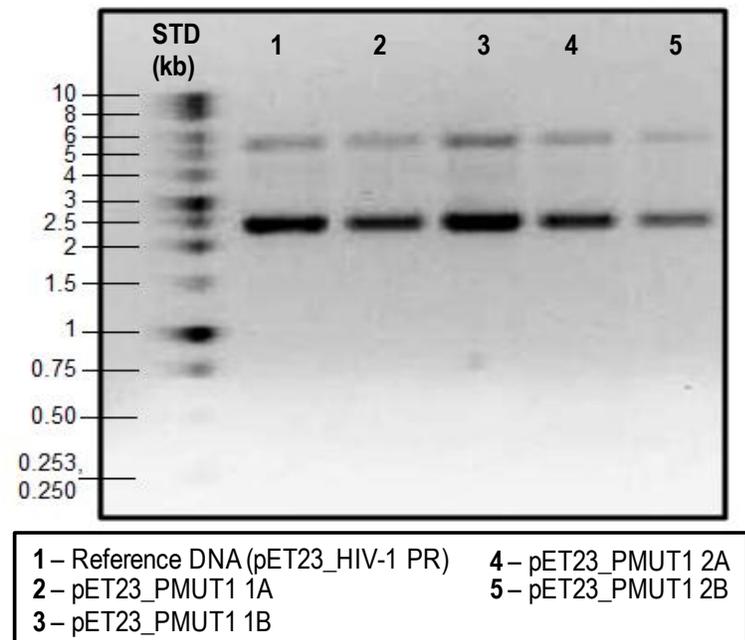


Figure 3-4. DNA gel of pET23_PMUT1. The STD lane corresponds to NEB 1 kb standard DNA ladder.

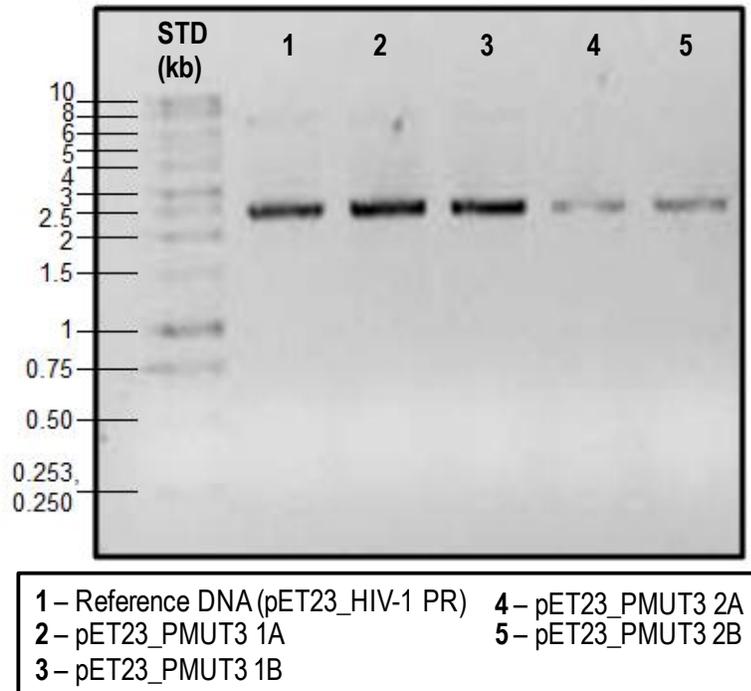


Figure 3-5. DNA gel of pET23_PMUT3. The STD lane corresponds to NEB 1 kb standard DNA ladder.

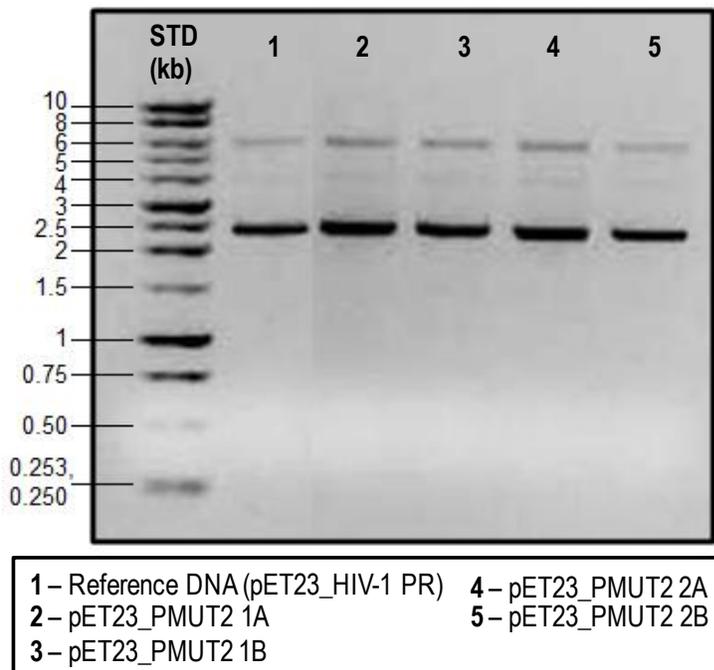


Figure 3-6. DNA gel of pET23_PMUT2. The STD lane corresponds to NEB 1 kb standard DNA ladder.

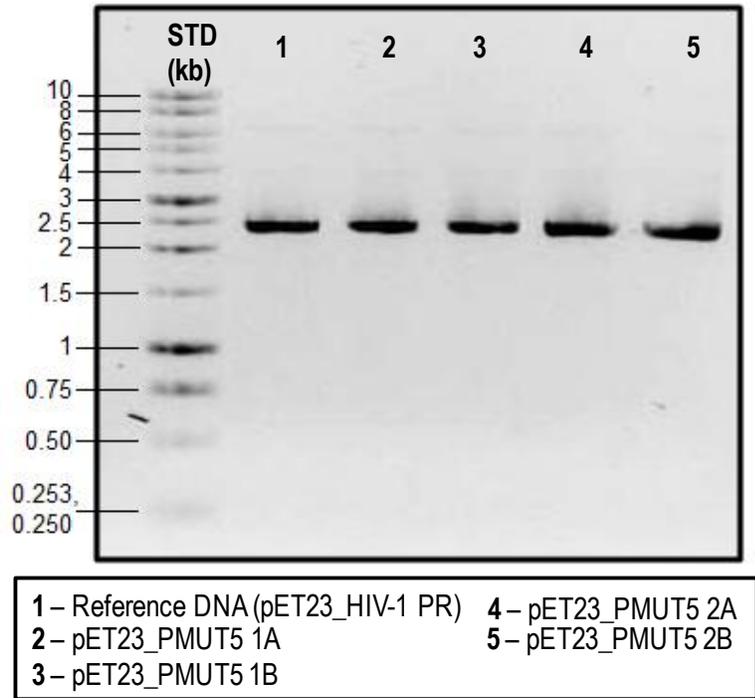


Figure 3-7. DNA gel of pET23_PMUT5. The STD lane corresponds to NEB 1 kb standard DNA ladder.

CHAPTER 4 PURIFICATION AND CHARACTERIZATION OF HIV-1 PROTEASE CONSTRUCTS

Materials and Methods

Chemicals, reagents, and supplies were obtained from Fisher Scientific (Pittsburgh, Pennsylvania) unless otherwise indicated. BL21(DE3) pLysS *E. coli* cells were purchased from Invitrogen (Carlsbad, California). AG 501-X8 (D) resin, 20-50 mesh, Laemmli sample buffer (LSB), Biorad Criterion pre-cast 16.5% Tris-Tricine peptide SDS-PAGE gel and other SDS-PAGE buffers and supplies were purchased from BioRad (Hercules, California). Molecular weight (MW) protein markers were purchased from Promega (Madison, Wisconsin). HiTrap™ Q HP Anion Exchange column and HiPrep™ 26/10 desalting column were purchased from GE Healthcare (formerly Amersham Biosciences, Pittsburgh, Pennsylvania). 0.60 I.D. x 0.84 O.D. capillary tubes were purchased from Fiber Optic Center (New Bedford, Massachusetts).

Expression of HIV-1 PR

The following plasmid DNA constructs were transformed separately into BL21(DE3)pLysS *E. coli* cells *via* standard heat-shock method: 1) pET23_ PRE, 2) pET23_ POST, 3) pET23_PMUT1, 4) pET23_PMUT2, 5) pET23_PMUT3, 6) pET23_PMUT5, 7) pET23_ B_{si}-K55C and 8) pET23_ B_{si} -I63P-K55C (Refer to Table 3-3 for the naming convention). The transformed cells were inoculated in 5 mL sterile Luria-Bertani (LB) media and grown at 37°C with shaking at 250 rpm until the optical density at 600 nm (OD₆₀₀) reached approximately 0.60. This was transferred to 1 L sterile LB media and grown at 37°C with shaking at 300 rpm to approximately OD₆₀₀ of 1.0. The cells were induced by addition of 0.1% (v/v) isopropyl-β-D-thiogalactopyranoside (IPTG) and further incubated for 4 to 5 h at 37°C. Cells were

harvested by centrifugation for 15 min at 8500 g and 4°C using a Sorvall RC6 floor-model centrifuge with SLA-3000 rotor.

Purification of HIV-1 PR

All buffers used for protein purification are listed in Table B-1 of Appendix B. Resuspension and wash buffers were prepared prior to protein purification. The resuspension buffers containing 1 M and 9 M urea were prepared fresh. All buffers were filtered through 0.22 µm membrane. Approximately 1 to 2 g of AG 501-X8 (D) resin (20-50 mesh) was added to the urea buffers, allowed to stir for 1 to 2 h, and the resin was removed by filtration. The pH of the inclusion body (IB) resuspension buffer containing 9 M urea was typically set to pH 8.85 or adjusted according to the specific isoelectric point (pI) of the protein being purified. The theoretical pIs of the HIV-1 PR constructs, which were generated by the ExPASy ProtParam tool (<http://expasy.org>), are shown in Table 4-1.

The cell pellet obtained from expression was resuspended in 30 mL resuspension buffer. The cells were lysed by tip sonication for 2 min (5 sec on, 5 sec off) at approximately 25 W output power using a Fisher Scientific sonic dismembrator (Pittsburgh, Pennsylvania) followed by 3 passes through a 35-mL French pressure cell (Thermo Scientific, Waltham, Massachusetts) operating at approximately 1200 lb/psi. The lysed cells were centrifuged for 30 min at 18500 g and 4°C using the Eppendorf 5810R centrifuge with F34-6-38 rotor. The same rotor was used for all the succeeding washing steps and all other steps in the purification. After centrifugation, the supernatant was discarded and the pellet, which contains the cell debris and inclusion bodies, was collected and resuspended with 40 mL wash buffer 1. The suspension was

homogenized using a 50 mL Dounce Tissue Homogenizer, tip sonicated for 2 min, then centrifuged for 30 min at 18500 g and 4°C. The inclusion bodies were washed again using the same process with 40 mL each of wash buffers 2 and 3. The washing steps served to separate the inclusion bodies from non-target proteins and other cellular components. The inclusion bodies were solubilized in 30 mL inclusion body (IB) resuspension buffer and homogenized followed by tip sonication and centrifugation at 18500 g and 4°C for 30 min. The supernatant was collected and subjected to anion exchange chromatography.

A 5-mL HiTrap™ Q HP anion exchange column (GE Healthcare, Pittsburgh, Pennsylvania) was equilibrated with approximately 60 to 80 mL of IB resuspension buffer on an Akta Prime liquid chromatography system (GE Healthcare, Pittsburgh, Pennsylvania). The supernatant (approximately 30 mL) was applied to the column at a rate of 5 mL/min and the flow through, which contains the HIV-1 PR, was collected in 4-mL fractions. The fractions (approximately 32 mL) were pooled and acidified to pH 5 by addition of formic acid to a final concentration of 25 mM. This was allowed to stand overnight at 4°C allowing precipitation of contaminants. Any precipitate formed was removed by centrifugation at 8500 g for 30 min at 6°C. The solution containing soluble HIV-1 PR was refolded by 10-fold stepwise dilution with 300 mL of 10 mM formic acid solution on ice using a peristaltic pump for approximately 2 h. The pH of the solution was adjusted to 3.8 by adding approximately 1 mL of 2.5 M sodium acetate. The solution was allowed to equilibrate for 1 h at 30°C followed by adjustment of pH to 5 with 2.5 M sodium acetate. The solution was centrifuged for 20 min at 18500 g and 23°C to remove contaminants that precipitated during refolding. HIV-1 PR was

concentrated to $OD_{280} = 0.5$ using an Amicon® 100-mL concentrator equipped with a Millipore 10,000 Da MW cut-off polyethersulfone (PES) membrane and buffer exchanged with 10 mM Tris-HCl buffer pH 6.9 using a 53-mL HiPrep™ 26/10 desalting column for spin labeling. Prior to buffer exchange, the HiPrep™ 26/10 desalting column was equilibrated with 3 to 4 column volumes (~60 to 80 mL) of the Tris-HCl buffer.

Spin Labeling of HIV-1 PR

The concentration of the HIV-1 PR in 10 mM Tris-HCl buffer was adjusted to approximately $OD_{280} = 0.2$. A very small amount of spin label [(1-oxy-2,2,5,5-tetramethylpyrrolinyl-3-methyl) methanethiosulfonate (MTSL)], approximately 20-fold molar excess relative to HIV-1 PR, was dissolved in 100 μ L of pure ethanol and added to about 40 mL sample. The spin-labeling reaction was carried out in the dark at 4°C for 12 to 16 h. The spin labeled protein was buffer exchanged into the storage buffer, 2 mM NaOAc of pH 5.0, using a desalting column. Column equilibration was done as previously using 2 mM NaOAc buffer. The concentration of the HIV-1 PR spin labeled samples was adjusted according to the method of characterization: $OD_{280} = 0.5$ for SDS-PAGE; $OD_{280} = 0.2$ to 0.3 for CD spectroscopy; $OD_{280} = 0.1$ for mass spectrometry; $OD_{280} = 1.0$ for CW-EPR. Samples were stored at -20°C until analysis.

SDS-PAGE

Approximately 15 μ L of samples collected from the purification process and purified, spin labeled HIV-1 PR were mixed in a 1:1 ratio with Laemmli sample buffer (LSB) containing β -mercaptoethanol (BME) and subsequently boiled for 5 min. The samples were then loaded into a pre-cast 16.5% Tris-Tricine peptide SDS-PAGE gel. The gel was run at 150 V until the visible dye reached the bottom edge of the gel.

Circular Dichroism (CD) Spectroscopy

Samples with a concentration of approximately 0.2 to 0.3 mg/mL were prepared in 2 mM NaOAc buffer at pH 5.0. The protein concentration of each construct (Appendix B; Table B-2) was calculated using Beer's Law (Equation 2-1) from the collected absorption at 280 nm and extinction coefficient (ϵ) of 12490 M⁻¹ cm⁻¹ provided by the ExPASy website (<http://expasy.org/>). All measurements were collected on an Aviv 400 CD spectrometer using Hellma cuvettes with a path length of 1 cm. The parameters used for the CD experiments are listed in Table 4-2.

The measured values refer to ΔA and results were calculated by averaging data obtained from 4 scans and subtracting the background signal from the buffer. The mean residue ellipticity, $[\theta]$, was calculated based on the following equation:

$$[\theta] = \theta_d \frac{1}{10 \cdot c \cdot l \cdot n_R} \text{ [deg}\cdot\text{cm}^2\cdot\text{dmol}^{-1}] \quad (4-1)$$

where c is the concentration in mol/L (M), l is the path length in cm, and n_R is the number of residues. θ_d is the observed ellipticity in degrees calculated from the average corrected ΔA :

$$\theta_d = \frac{2.303}{4} \cdot \frac{180}{\pi} \cdot \Delta A \quad (4-2)$$

The CD spectrum is then obtained by plotting $[\theta]$ vs. wavelength in nm.

Mass Spectrometry (MS)

Spin labeled HIV-1 PR samples with concentration of approximately 0.1 mg/mL in 2 mM NaOAc buffer at pH 5.0 were submitted for analysis at the UF Chemistry Department Mass Spectrometry Facility. Approximately 0.5% (v/v) acetic acid in 50/50 (v/v) methanol/water was added to 100 μ L sample and injected directly through

an autosampler, ionized by electrospray ionization (ESI) in positive mode followed by analysis using an Agilent 6210 Time-of-Flight Mass Spectrometer (TOF-MS). All data were processed using the MassHunter™ software.

CW-EPR

Samples were placed in 0.60 I.D. x 0.84 O.D. capillary tubes. The tube was inserted into the loop gap resonator and data were collected at room temperature (approximately 20 to 25°C). The CW-EPR instrument used was a modified Bruker ER200 spectrometer with an ER023 M signal channel and an ER032 M field control unit (Molecular Specialties, Milwaukee, Wisconsin). CW EPR spectra were collected with 1 G modulation amplitude and 100 G sweep width at X-band frequency of 9.6 to 9.7 GHz. Table 4-3 provides a complete list of the typical parameters used for CW-EPR data collection.

Results and Discussion

The expression and purification protocol used in this work was based on Freire et al. (83, 105, 147) and optimized by a fellow group member, Angelo Veloro. In the purification process, buffers containing urea were prepared fresh because urea degrades spontaneously forming products that can carbamylate free cysteines and thereby affect the spin labeling efficiency of HIV-1 PR (148). Any urea decomposition products were reacted by ionization with glycyglycine (diGly), which is present in the wash buffers, followed by removal by ion-exchange with the AG501-X8 (D) mixed-ion bed resin. The purified and spin labeled HIV-1 PR samples were characterized by SDS-PAGE, CD, MS, and CW-EPR.

SDS-PAGE

The purity of HIV-1 PR was monitored by running SDS-PAGE on samples obtained from specified steps in the purification process. The technique can be also be used to estimate the molecular weight (MW) of protein by comparison with known standards. SDS-PAGE gels of HIV-1 PR constructs purified in this work are shown in Figures 4-1 to 4-8. Bands that appeared near the 10 kDa MW marker correspond to the HIV-1 PR monomer, which is expected as its estimated MW is 11 kDa. All constructs possess this band but with varying amounts of protein. The size of the band is roughly proportional to the amount of protein present in the sample. The POST, PMUT3, and PMUT5 constructs have the most abundant HIV-1 protease. Degradation of HIV-1 PR can be monitored by examining the presence of any bands lower than the 10 kDa marker, which corresponds to possible cleavage products. Though hardly noticeable because of the very large band size for POST, PMUT3, and PMUT5 HIV-1 PR, some faint bands can be seen below the HIV-1 PR band.

The SDS-PAGE gel profiles of all constructs were similar with some minor differences. For all constructs, the supernatant after French press produced bands of various MW, which correspond to cellular debris and other components that dissolved in the resuspension buffer and were separated from the inclusion bodies. As seen in Figures 4-1 to 4-8A, comparison of lanes 2-4 and lanes 6-9 reveals that several bands appeared in the wash pellets while only a few faint bands appeared for the wash supernatant solutions. This implies that the wash steps were not substantially effective in removing non-target proteins. It is worth noting that lanes 6-8 of the SDS-PAGE gel of B_{si} (Figure 4-7) showed only a faint band for HIV-1 PR. However, lanes 5 and 9, which correspond to wash 4 supernatant and pellet, respectively, contain the band for HIV-1

PR. This anomaly may be attributed to the SDS-PAGE sample preparation. After dissolving the sample in LSB followed by boiling, the sample was centrifuged. Most of the HIV-1 PR may be insoluble and deposited at the bottom of the microcentrifuge tube. Consequently, only little amount of the protein was loaded into the gel.

For all constructs, distinct bands appeared at approximately 35 kDa during the purification process (Figures 4-1 to 4-8A). This unidentified protein contaminant was not removed after passing through the Q column since bands of the same MW were still present as seen in Figures 4-1 to 4-8B. For the POST construct (Figure 4-2), a light but distinct band was present between 15 and 25 kDa in both gels, which could correspond to the HIV-1 PR dimer of approximately 21 kDa. This also appeared as very faint bands in the gels of the PMUT3 and B_{si} constructs. Based on the gel profiles of the column fractions, fractions 4-9 contained the largest amount of protein and were pooled for the subsequent acidification and refolding steps.

Acidification of the HIV-1 PR facilitated the precipitation of protein contaminants that appeared as several bands in Figure 4-1 to 4-8B. After the acidification and refolding process, a huge amount of solid was typically removed by centrifugation. At this point, the HIV-1 PR obtained is estimated to have >95% purity and suitable for spin labeling. Spin labeled HIV-1 PR samples of OD ~0.5 were run through SDS-PAGE to verify the purity. The SDS-PAGE gel in Figure 4-9 shows that the spin labeled HIV-1 PR samples have >99% purity.

CD Spectroscopy

Isolation of protein from inclusion bodies entails refolding the protein using an appropriate buffer with the correct pH. After the refolding process, the purified protein may or may not be properly folded. In this work, CD experiments were performed to

verify if the purified and spin labeled HIV-1 PR constructs have the proper secondary structure. Figure 4-10 shows the CD spectra of the HIV-1 PR constructs. All constructs have a similar profile, which is typical of a protein with primarily β -sheet conformation. This is in agreement with the structure of HIV-1 PR, which consists of predominantly β strands. The experimentally-determined CD spectra also match published CD results for HIV-1 PR, thus confirming the proper folding of the protein samples. However, slight differences can be seen in the minima, particularly with PMUT5, which may be attributed to the mutations present in the construct. Further CD analysis needs to be performed to calculate the % β -sheet structure of each of the spin labeled samples.

MS

The purified and spin labeled HIV-1 PR samples were analyzed by ESI-MS to confirm the molecular weight (MW) of HIV-1 PR. In addition, this method was used to check the spin labeling efficiency, that is, whether MTSL was attached successfully to the protein. Table 4-4 shows the list of MW of the HIV-1 PR constructs with and without spin label and the observed MW obtained from the MS spectra. The theoretical MW with MTSL corresponds to $MW(\text{HIV-1 PR}) + MW(\text{MTSL}) - MW(\text{Hydrogen})$, where the MW (MTSL) is 185.38 Da.

Figures 4-11 to 4-18 show mass spectra of the HIV-1 PR constructs. A typical ESI-MS mass spectrum contains a series of peaks that represent multiple-charge variants. The m/z peaks may be protonated or adducts of Na^+ , NH_4^+ or any ion present in solution. The mass spectra of HIV-1 PR constructs possess m/z peaks that correspond to $[M/n]^{n+}$ ions, where n represents the charge state. Because the MW of the protein is known, identifying the characteristic m/z peaks is relatively straightforward. The simplest

strategy for peak identification is to pick the +10 charge state, divide the MW by 10, and search for the m/z peak. For example, the theoretical MW of the PRE construct is 10874.98 Da, thus the m/z peak having a +10 charge state corresponds to an approximate m/z of 1087.498. This peak corresponds to m/z 1087.494 in Figure 4-11. All the other peaks can easily be identified because they are adjacent to each other, differing by a consecutively decreasing number of charges as m/z increases. The ESI-MS results provide confirmation that all spin labeled HIV-1 PR constructs have the correct MW.

The spin labeling efficiency was determined by verifying the presence of any unlabeled protein. The MS spectra of PMUT3 (Figure 4-14), PMUT5 (Figure 4-16), and B_{si} (Figure 4-17) show the presence of unlabeled HIV-1 PR monomer and dimer. By comparison of intensities of the m/z peaks, the spin labeling efficiency can be roughly estimated. PMUT3 contains approximately 40% unlabeled protein whereas PMUT5 and B_{si} have about 20 to 30% unlabeled protein. All the other HIV-1 PR constructs contain little to no unlabeled protein and are thus efficiently labeled.

CW-EPR

Site-directed spin labeling in conjunction with EPR has been used as a tool to probe the structure and conformational dynamics of proteins (132-134, 140, 141). The motional dynamics of the nitroxide side chain, as reflected in the EPR line shape, can be correlated to motion arising from the secondary structure through local backbone fluctuations (LBF) and the overall protein structure due to conformational changes in the protein fold. In this work, SDSL-EPR was used to observe any changes in the nitroxide line shape of different variants of HIV-1 PR. Figure 4-19 shows an overlay of the CW-EPR spectra of HIV-1 PR constructs. By visual inspection, only a minor difference

in the EPR line shape, particularly on the high-field portion (Figure 4-19 inset), can be observed. This is in agreement with previous studies in our group, which revealed only slight changes in the EPR spectra of different HIV-1 PR subtypes (51). In fact, it has also been shown that addition of inhibitor to HIV-1 PR has a minor effect on the spectral line shape. NMR and ITC studies have suggested that inhibitor binding induces a conformational change and that the major source of protein motion arises from minor backbone fluctuations (28, 29, 33-37). These types of motion, however, were not detected using CW-EPR at X-band frequency. The spectral line shape rather reflects only the internal motion of the nitroxide spin label. Because CW-EPR did not reveal any information on protein conformational changes, the technique was used instead to confirm whether the HIV-1 PR constructs are efficiently spin labeled for DEER studies. Pulsed-EPR, described in further detail in Chapter 5, was employed to monitor the conformational changes in HIV-1 PR.

Qualitative inspection of the signal-to-noise ratio (SNR) of the CW-EPR spectra (Figure 4-19) reveals that all HIV-1 PR samples are substantially spin labeled. Spin labeled HIV-1 PR samples for DEER analysis were typically concentrated to OD₂₈₀ of 2.5 to 2.7. However, it should not be assumed that spin labeled proteins with the same optical density have the same concentration of spin label. A standard curve using 4-oxo-TEMPO was obtained (Appendix B; Figure B-1) to verify the concentration of spin labeled protein that produces CW-EPR spectra with good SNR. Table B-3 of Appendix B provides a list of values for optical density, protein concentration, and spin label concentration of the HIV-1 PR constructs. It has been determined that samples with OD₂₈₀ of ~1.0 and spin label concentration of greater than 15 μ M for the HIV-1 PR

monomer are efficiently labeled and suitable for analysis by DEER. Based on Table B-3, it is noticeable that PMUT3, PMUT5, and B_{si} have high OD₂₈₀ values but correspond to relatively low spin label concentrations. This suggests that these constructs may not be efficiently labeled.

Table 4-1. Theoretical isoelectric points (pI) of HIV-1 PR constructs.

Construct #	Name of Construct ^a	pI
1	PRE	9.06
2	POST	9.06
3	PMUT1	9.06
4	PMUT2	9.06
5	PMUT3	9.45
6	PMUT5	9.06
7	B _{si}	9.39
8	B _{si} -I63P	9.39

^aSee Table 3-3 for the naming convention of HIV-1 PR constructs.

Table 4-2. Standard parameters used for circular dichroism (CD) spectroscopy data collection.

Parameter	Value
Experiment Type	Wavelength
Temperature	25 °C
Bandwidth	1.0 nm
Wavelength Start	250 nm
Wavelength End	200 nm
Wavelength Step	1.0 nm
Averaging Time	1.000 s
Settling Time	0.333 s
Multi-Scan Wait	1.000 s
Number of Scans	4

Table 4-3. Standard parameters used for CW-EPR data collection.

Parameter	Value
Frequency	~9.6 GHz
Center Field	~3230-3270 G
Sweep Width	100 G
Time Constant	0.16384 s
Acquisition Time	40.6323 s
Modulation Amplitude	~1 G
Power	20 dB
Receiver Gain	1 x 10 ⁵
Receiver Phase	100
Number of Scans	1
Number of Points	1024

Table 4-4. Theoretical MW of HIV-1 PR constructs with and without MTSL and MW determined from ESI-MS analysis.

HIV-1 PR Construct	Theoretical MW (Da) (without MTSL)	Theoretical MW (Da) (with MTSL)	Observed MW ^a (Da) (with MTSL)	Standard Deviation
PRE	10690.6	10874.98	10874.94	0.20
POST	10601.4	10785.78	10785.60	0.15
PMUT1	10629.4	10813.78	10813.62	0.18
PMUT3	10643.4	10827.78	10827.68	0.11
PMUT2	10642.5	10826.88	10826.74	0.11
PMUT5	10648.5	10832.88	10832.73	0.11
B _{si}	10702.7	10887.08	10886.93	0.07
B _{si} -I63P	10686.6	10870.98	10870.86	0.08

^aValue corresponds to the average MW calculated from the m/z peaks of varying charge states.

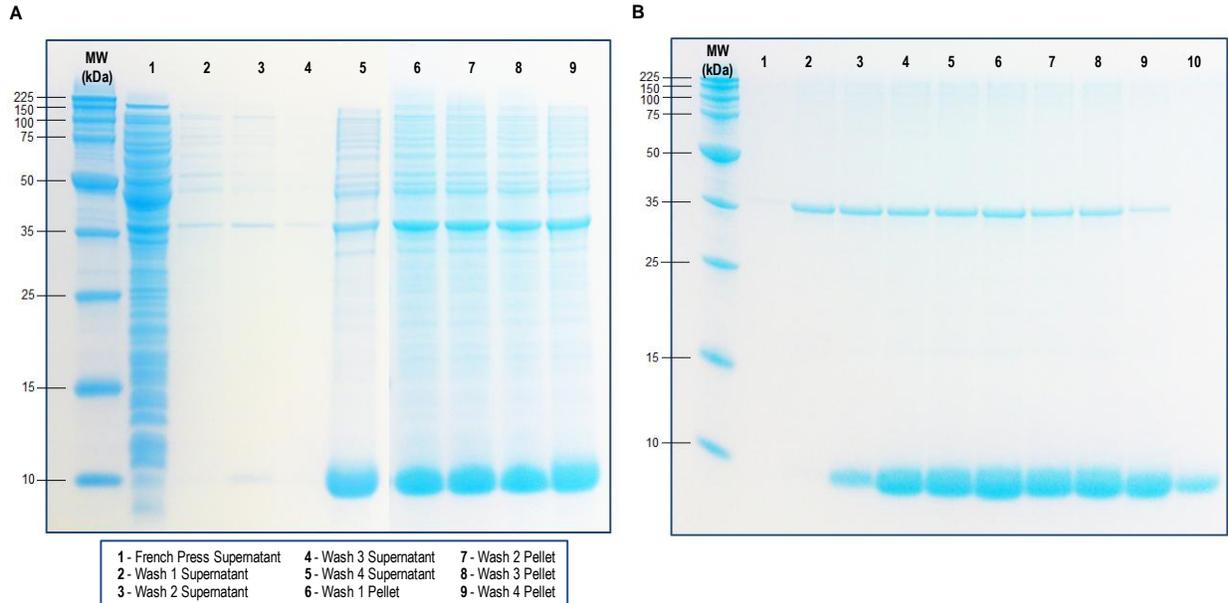


Figure 4-1. SDS-PAGE gel of PRE. A) Samples obtained from steps in the purification as indicated. B) Lanes correspond to fractions 1-10 collected from Q column chromatography. MW lane corresponds to Promega broad range protein markers.

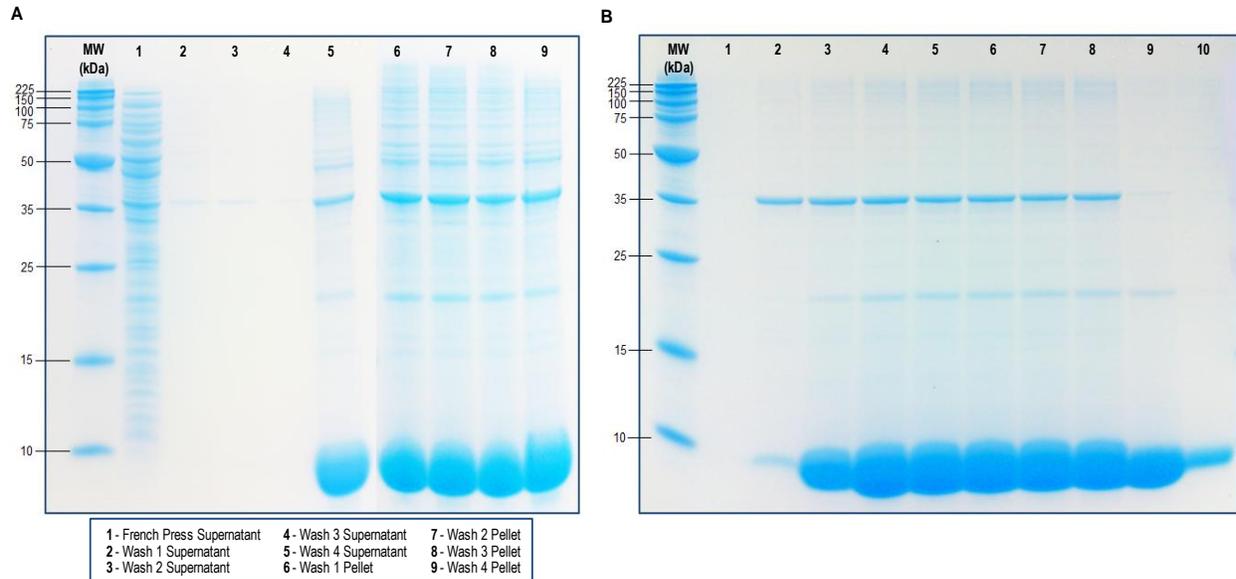


Figure 4-2. SDS-PAGE gel of POST. A) Samples obtained from steps in the purification as indicated. B) Lanes correspond to fractions 1-10 collected from Q column chromatography. MW lane corresponds to Promega broad range protein markers.

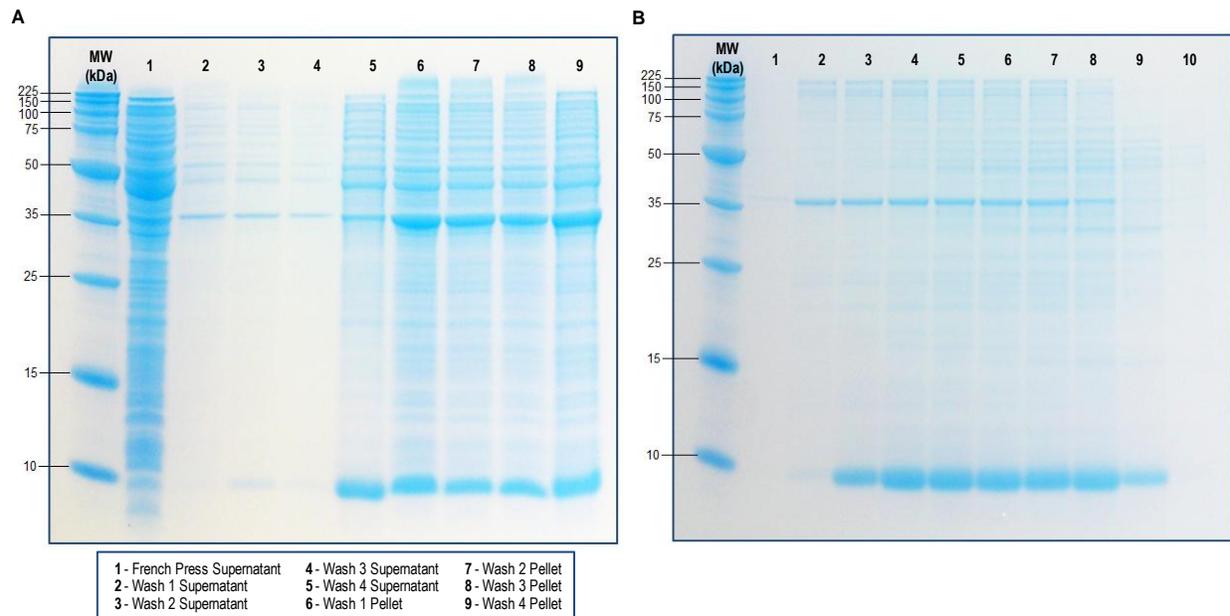


Figure 4-3. SDS-PAGE gel of PMUT1. A) Samples obtained from steps in the purification as indicated. B) Lanes correspond to fractions 1-10 collected from Q column chromatography. MW lane corresponds to Promega broad range protein markers.

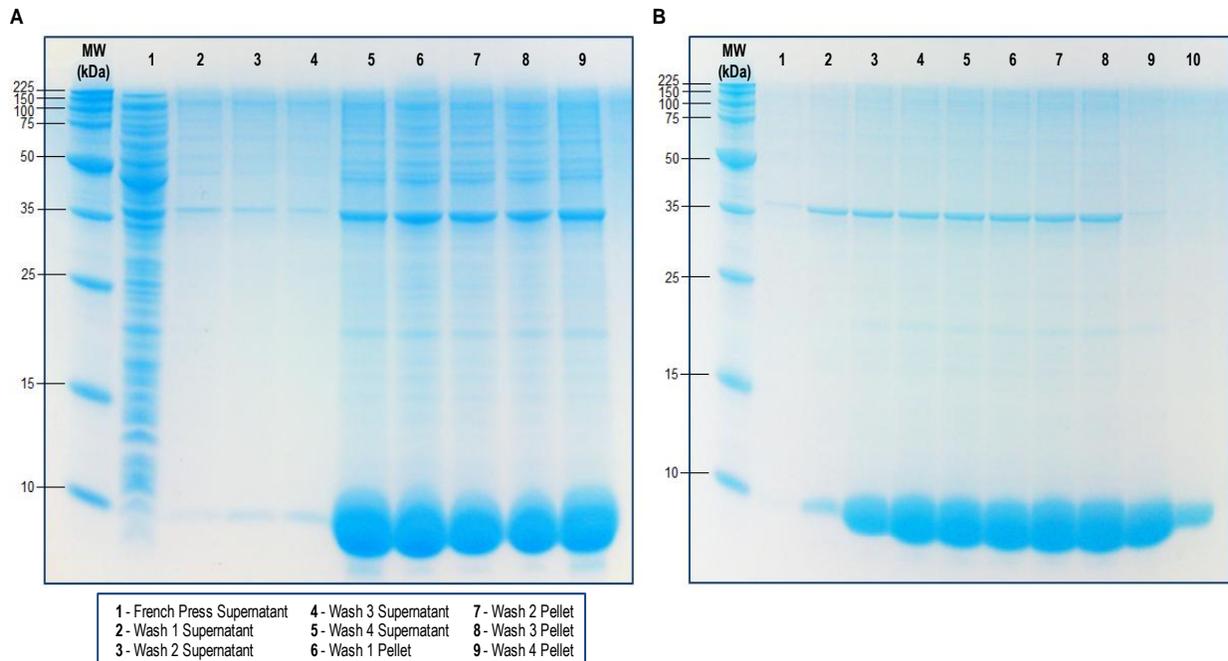


Figure 4-4. SDS-PAGE gel of PMUT3. A) Samples obtained from steps in the purification as indicated. B) Lanes correspond to fractions 1-10 collected from Q column chromatography. MW lane corresponds to Promega broad range protein markers.

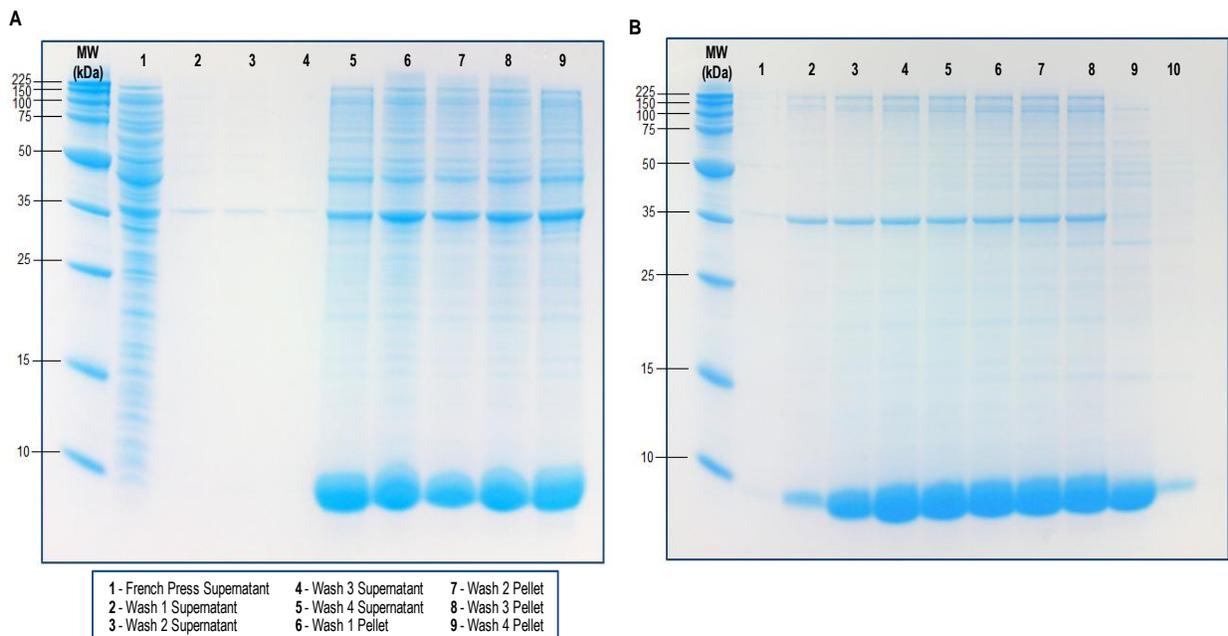


Figure 4-5. SDS-PAGE gel of PMUT2. A) Samples obtained from steps in the purification as indicated. B) Lanes correspond to fractions 1-10 collected from Q column chromatography. MW lane corresponds to Promega broad range protein markers.

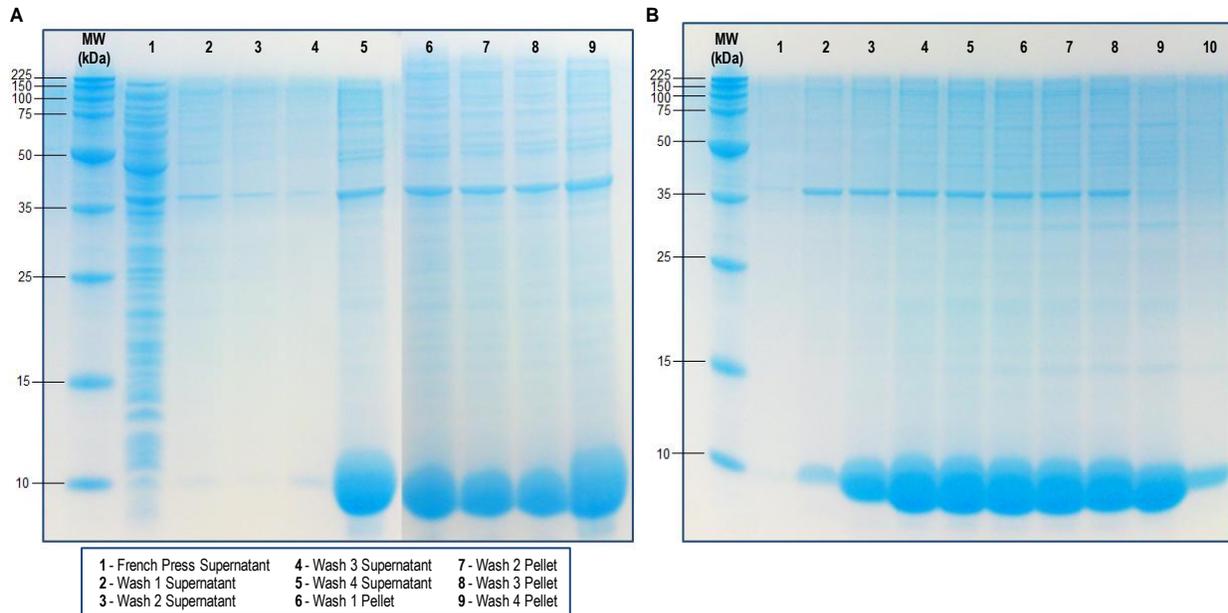


Figure 4-6. SDS-PAGE gel of PMUT5. A) Samples obtained from steps in the purification as indicated. B) Lanes correspond to fractions 1-10 collected from Q column chromatography. MW lane corresponds to Promega broad range protein markers.

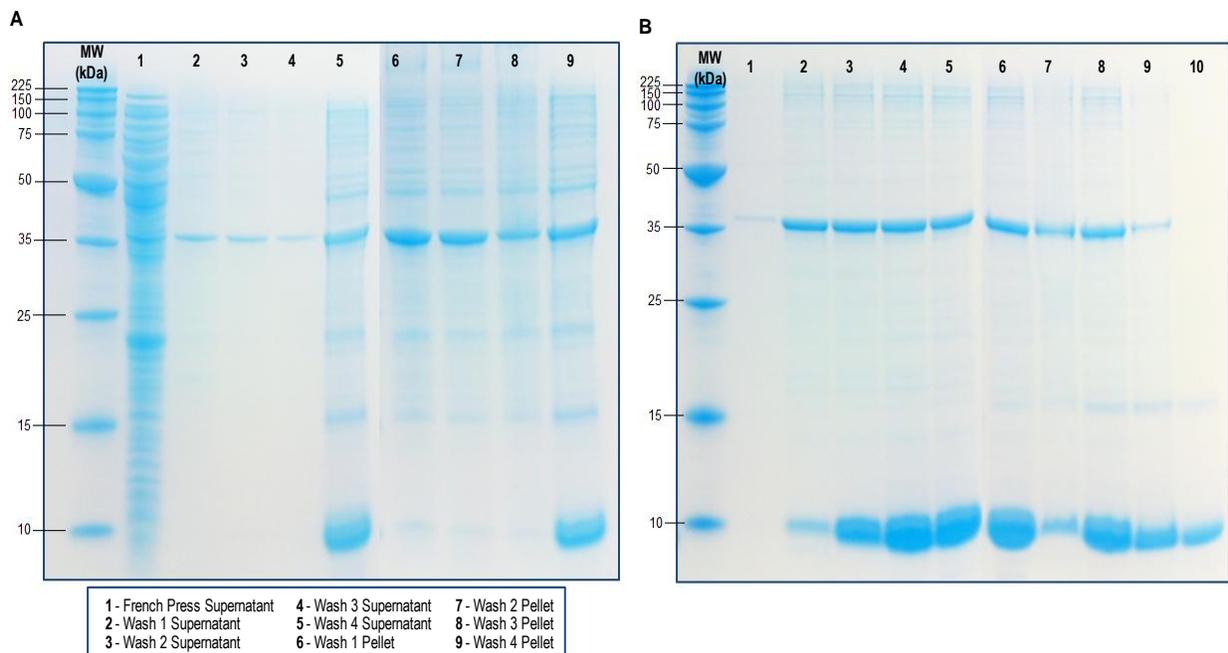


Figure 4-7. SDS-PAGE gel of B_{Si}. A) Samples obtained from steps in the purification as indicated. B) Lanes correspond to fractions 1-10 collected from Q column chromatography. MW lane corresponds to Promega broad range protein markers.

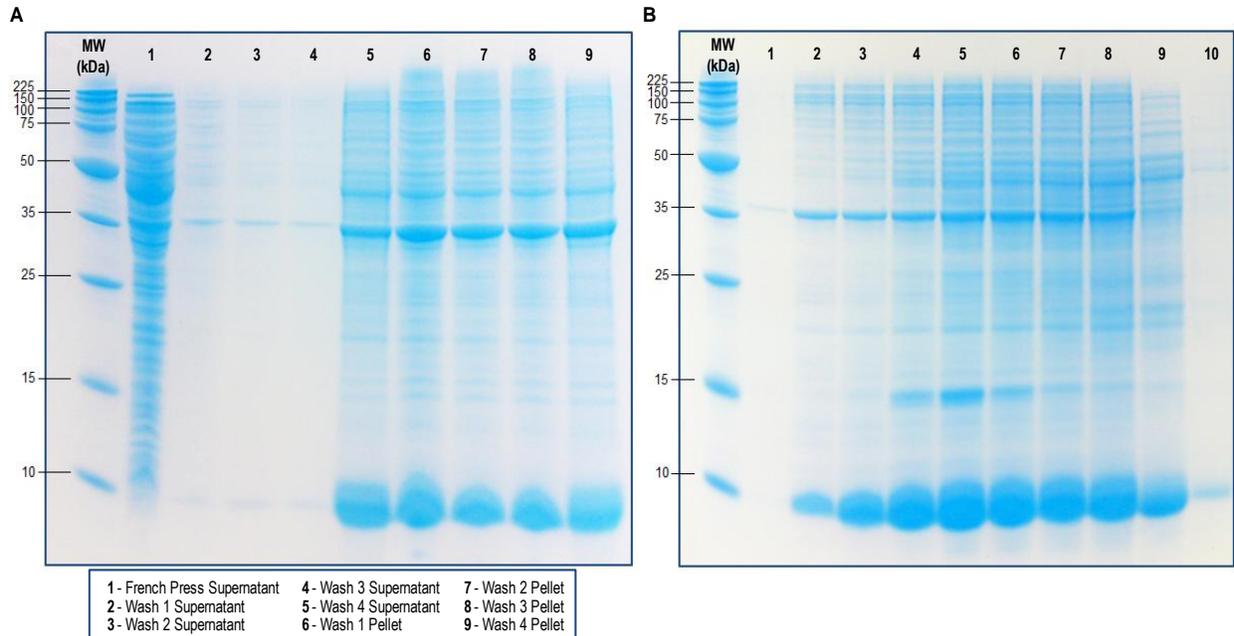


Figure 4-8. SDS-PAGE gel of B_{si}-I63P. A) Samples obtained from steps in the purification as indicated. B) Lanes correspond to fractions 1-10 collected from Q column chromatography. MW lane corresponds to Promega broad range protein markers.

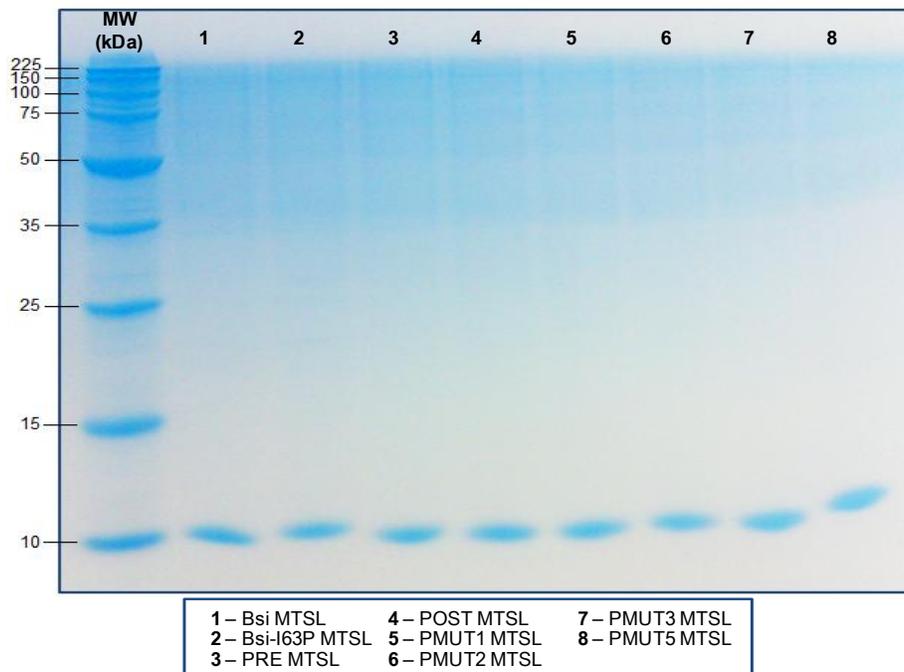


Figure 4-9. SDS-PAGE gel of spin labeled HIV-1 PR constructs. MW lane corresponds to Promega broad range protein markers.

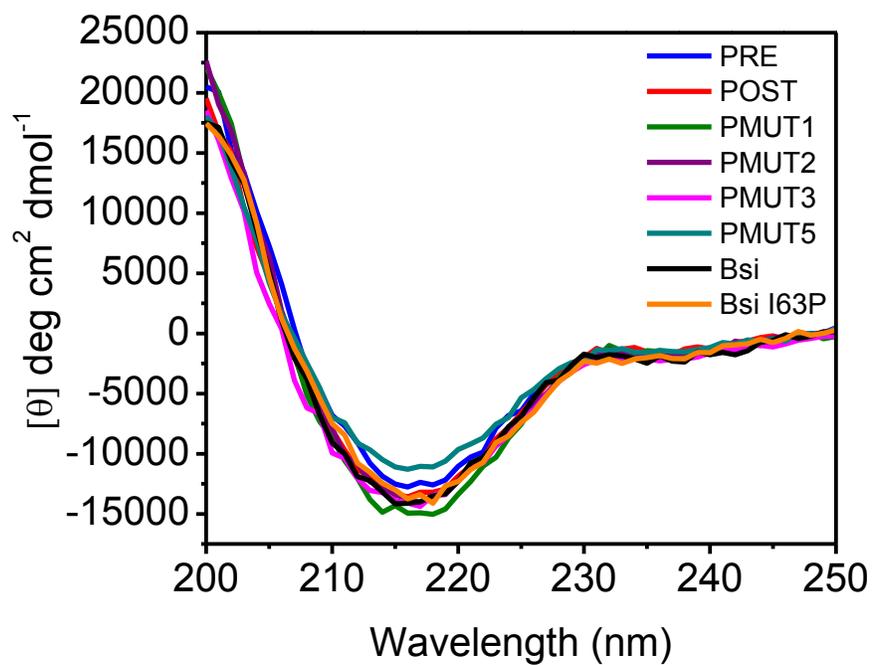


Figure 4-10. Circular dichroism spectra for spin labeled HIV-1 PR PRE (blue), POST (red), PMUT1 (green), PMUT2 (purple), PMUT3 (magenta), PMUT5 (cyan), B_{si} (black), and B_{si}-I63P (orange).

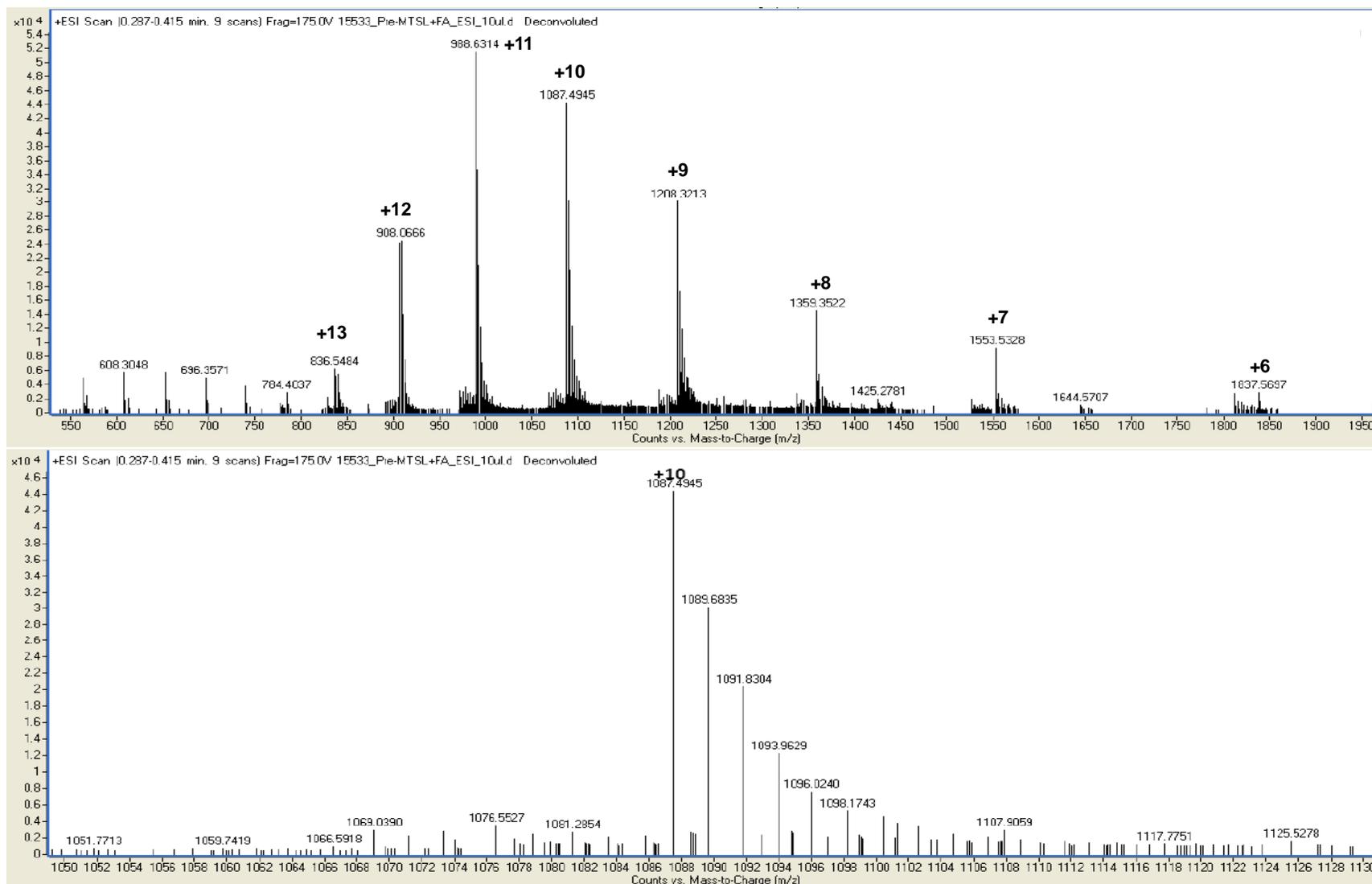


Figure 4-11. Mass spectra of spin labeled PRE. The upper spectrum shows the major peaks with the charge states indicated. The lower spectrum is an enlarged view of the m/z peak with +10 charge state.

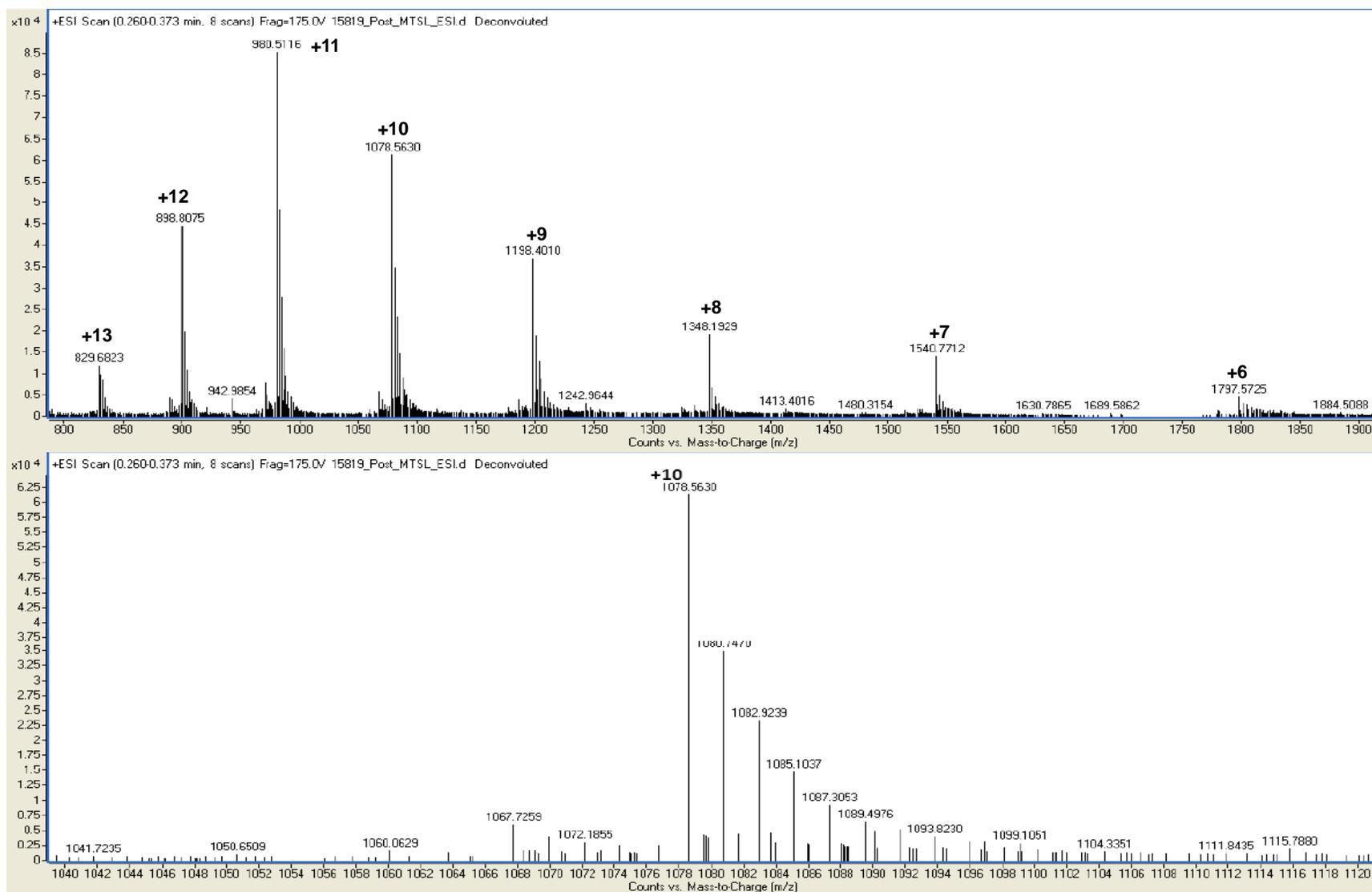


Figure 4-12. Mass spectra of spin labeled POST. The upper spectrum shows the major peaks with the charge states indicated. The lower spectrum is an enlarged view of the m/z peak with +10 charge state.

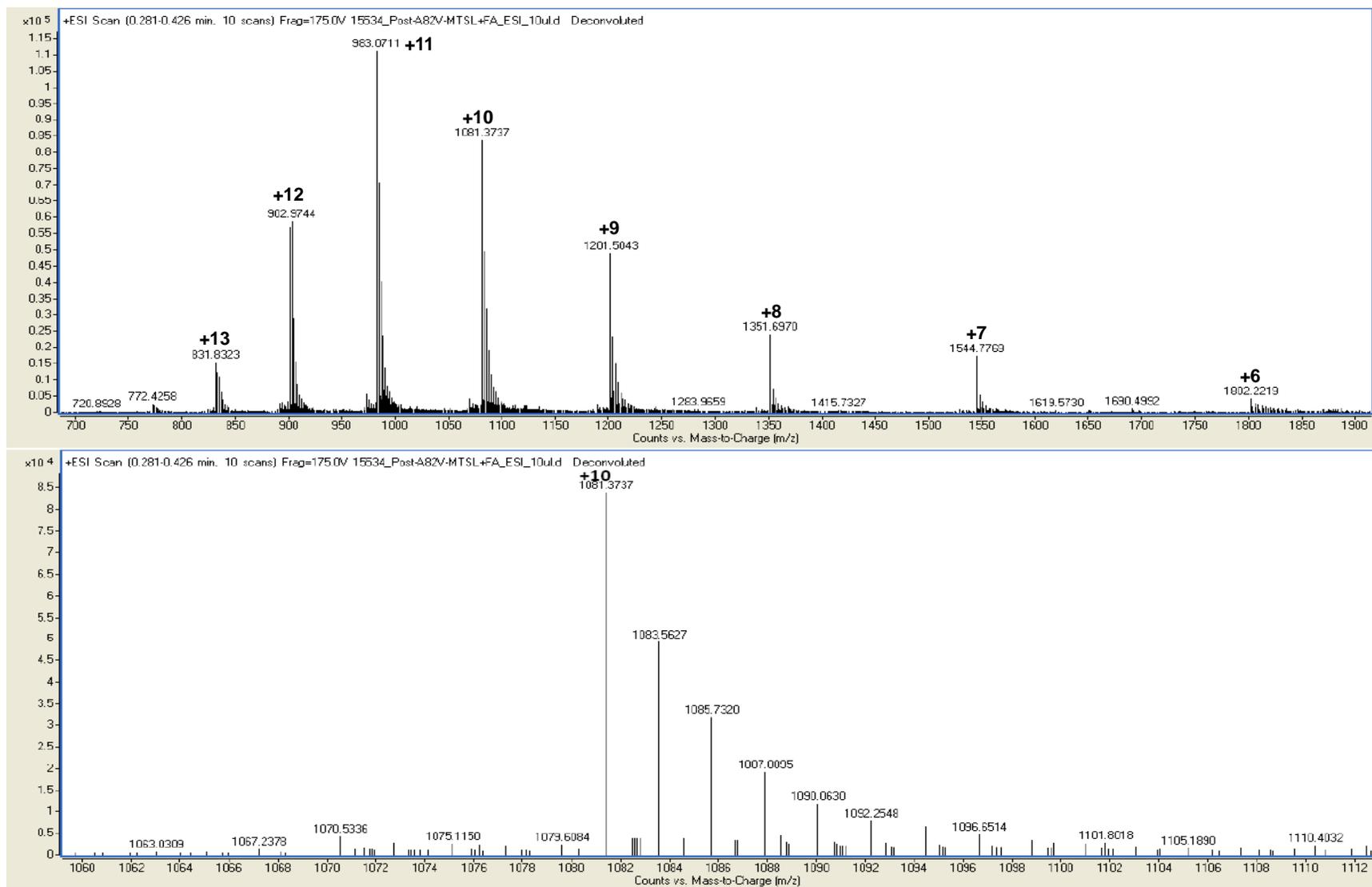


Figure 4-13. Mass spectra of spin labeled PMUT1. The upper spectrum shows the major peaks with the charge states indicated. The lower spectrum is an enlarged view of the m/z peak with +10 charge state.

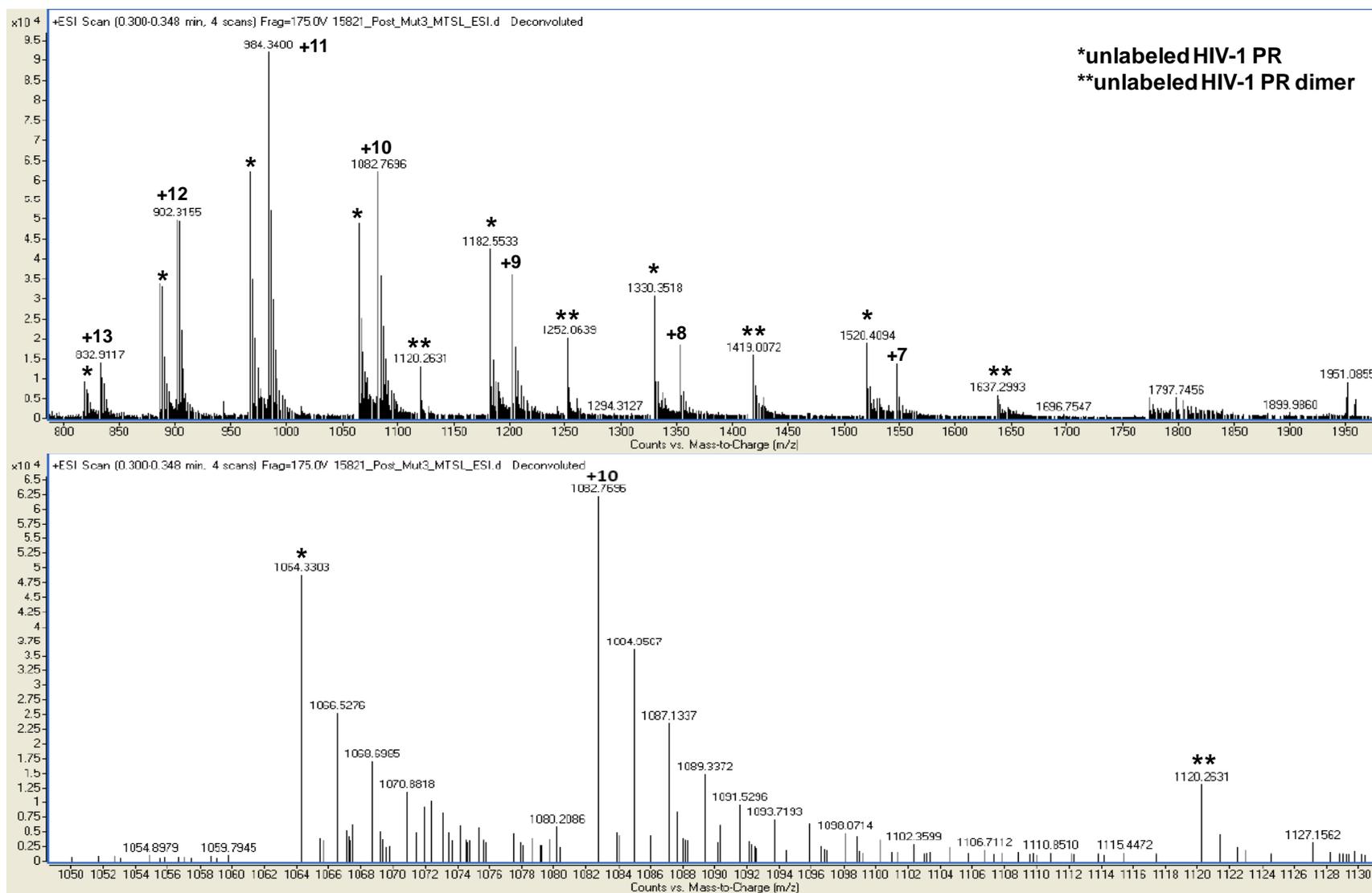


Figure 4-14. Mass spectra of spin labeled PMUT3. The upper spectrum shows the major peaks with the charge states indicated. The lower spectrum is an enlarged view of the m/z peak with +10 charge state.

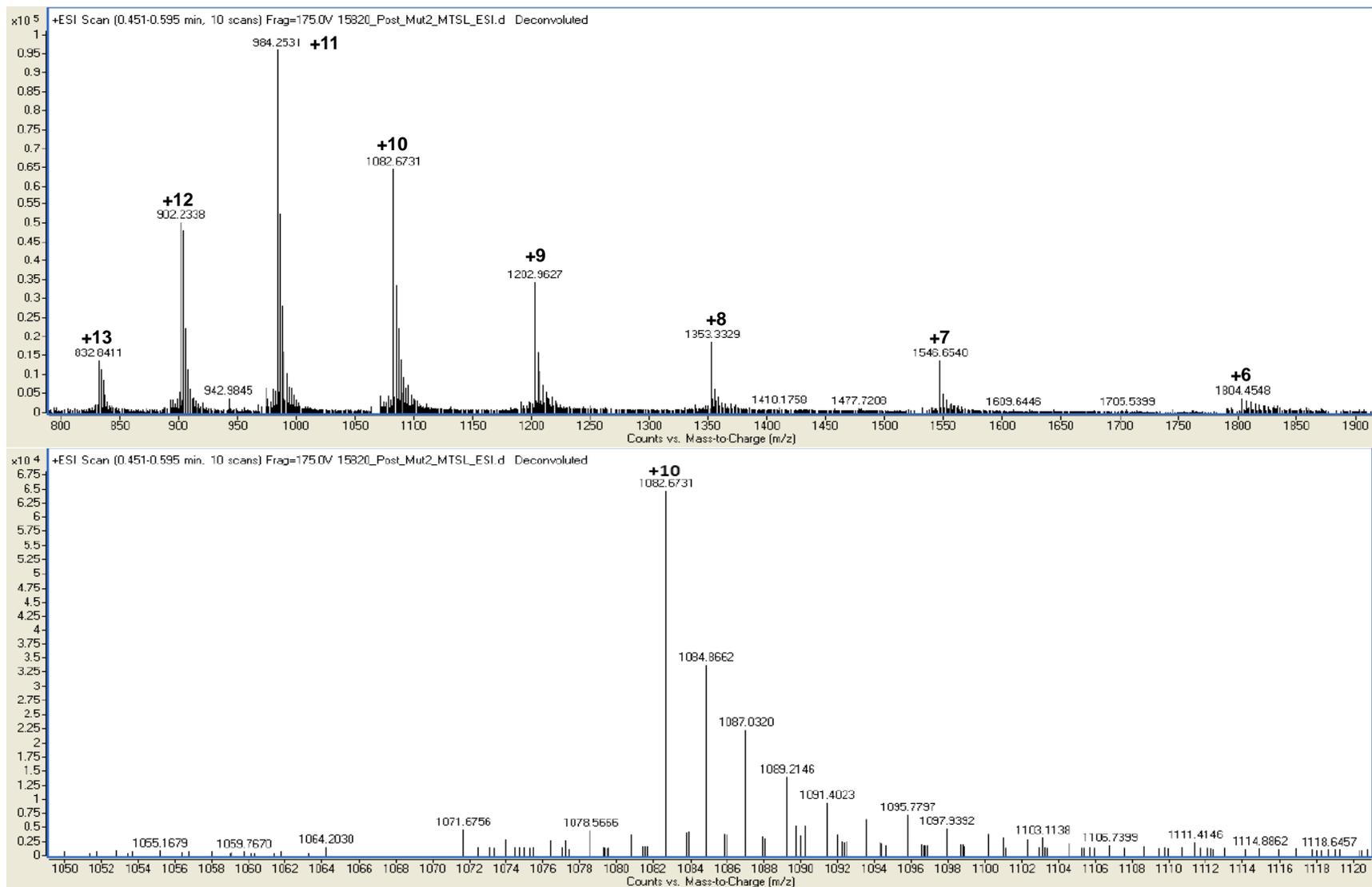


Figure 4-15. Mass spectra of spin labeled PMUT2. The upper spectrum shows the major peaks with the charge states indicated. The lower spectrum is an enlarged view of the m/z peak with +10 charge state.

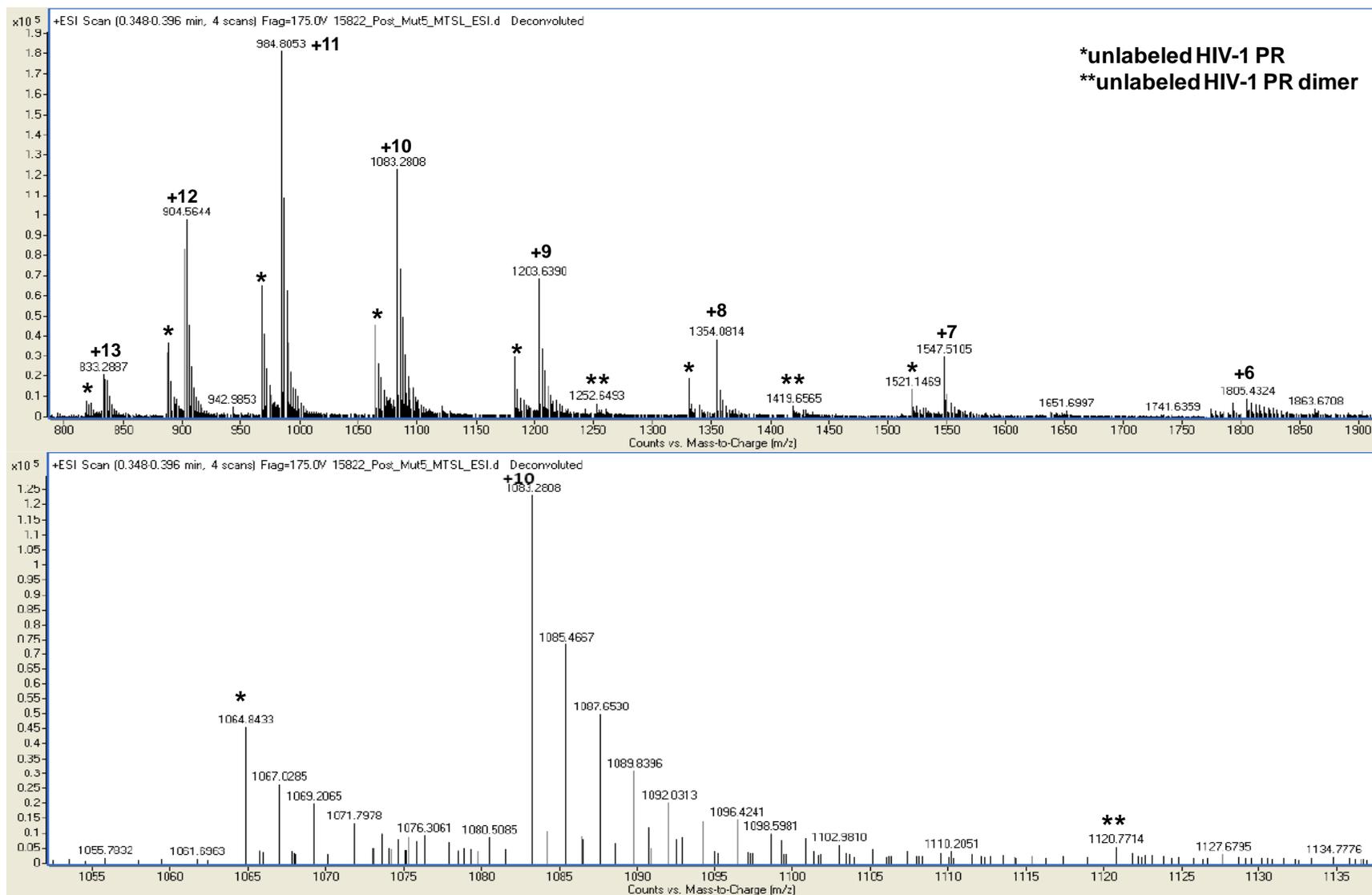


Figure 4-16. Mass spectra of spin labeled PMUT5. The upper spectrum shows the major peaks with the charge states indicated. The lower spectrum is an enlarged view of the m/z peak with +10 charge state.

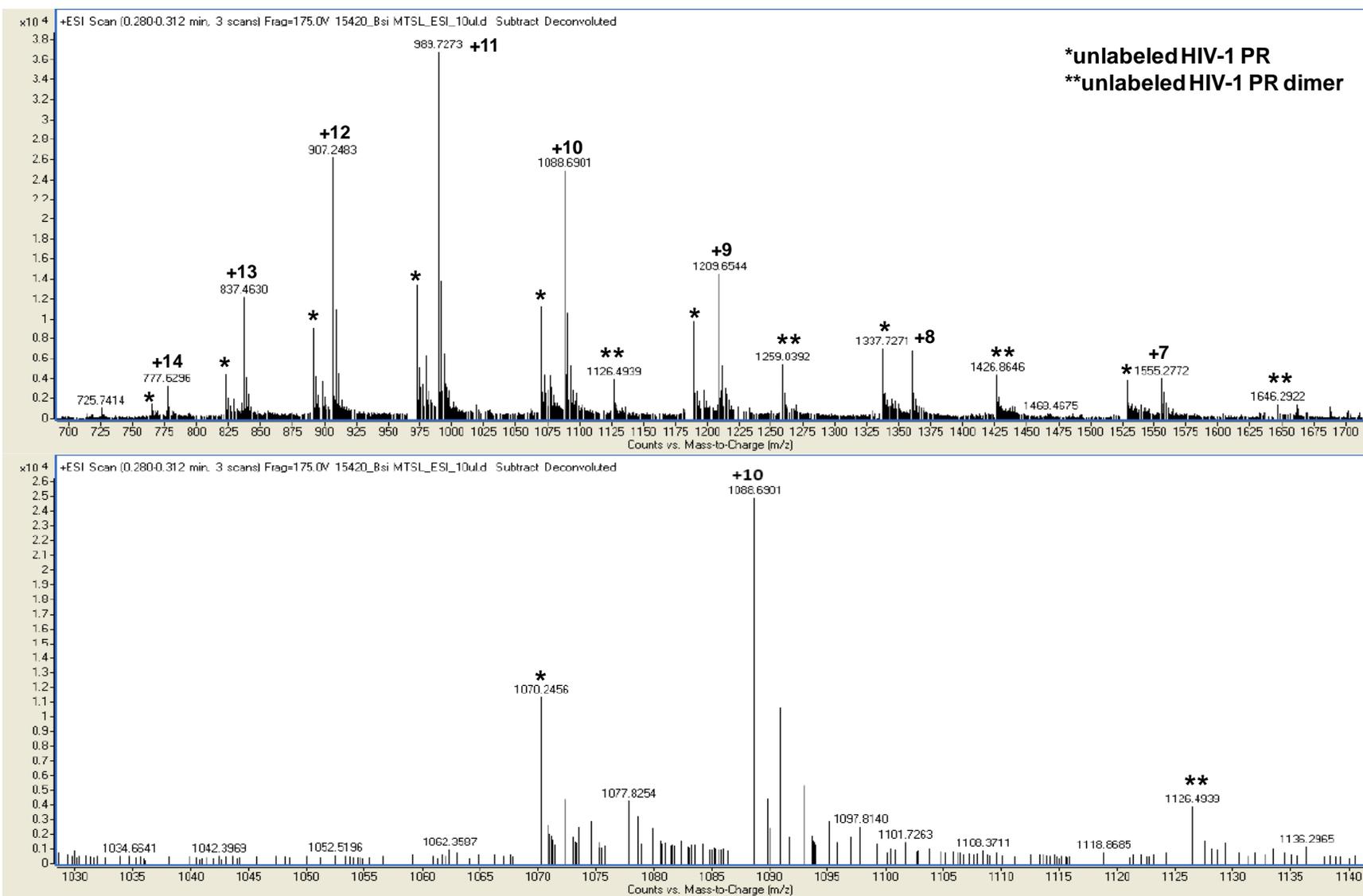


Figure 4-17. Mass spectra of spin labeled B_{si}. The upper spectrum shows the major peaks with the charge states indicated. The lower spectrum is an enlarged view of the m/z peak with +10 charge state.

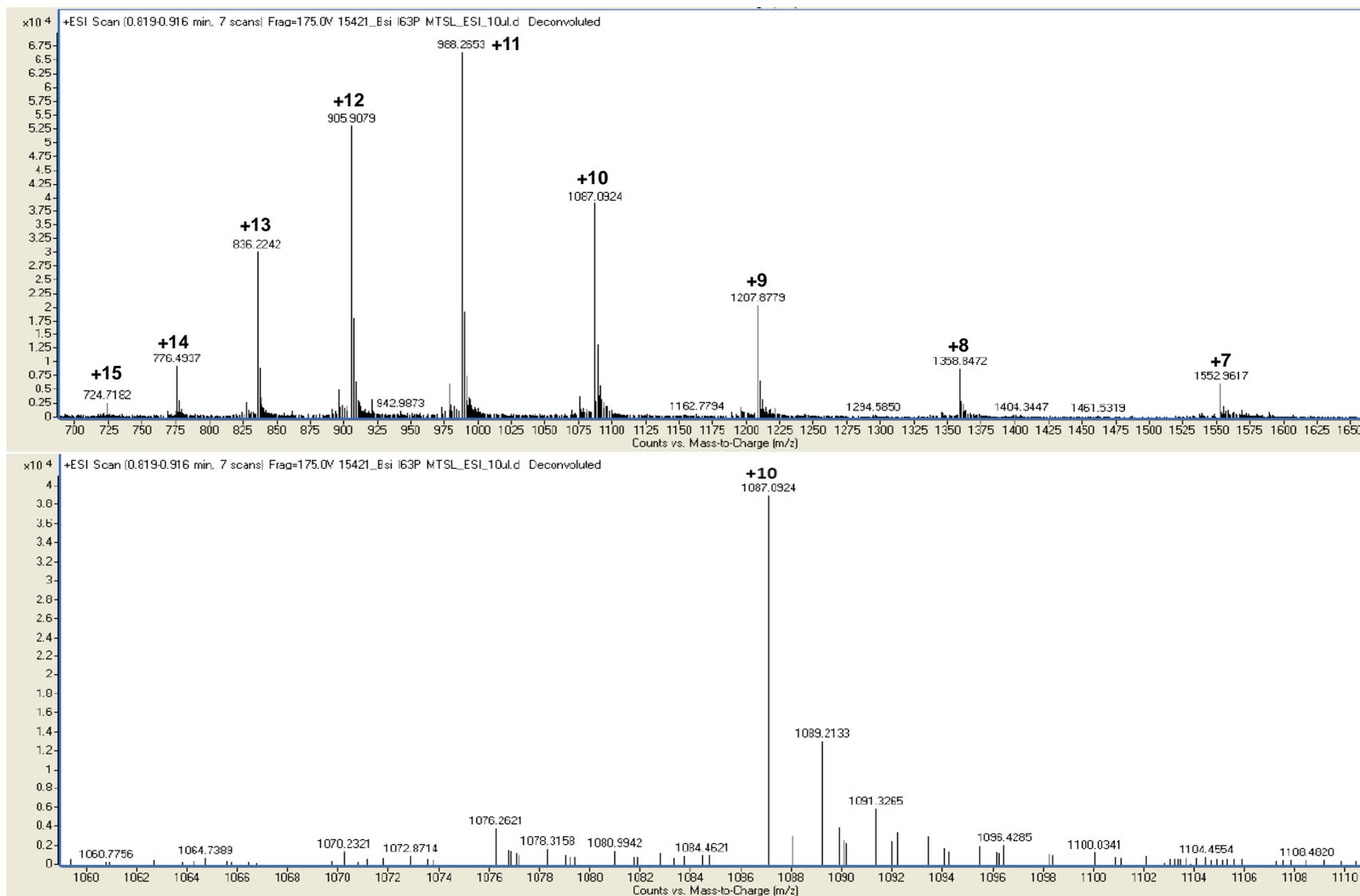


Figure 4-18. Mass spectra of spin labeled B_{si}-I63P. The upper spectrum shows the major peaks with the charge states indicated. The lower spectrum is an enlarged view of the m/z peak with +10 charge state.

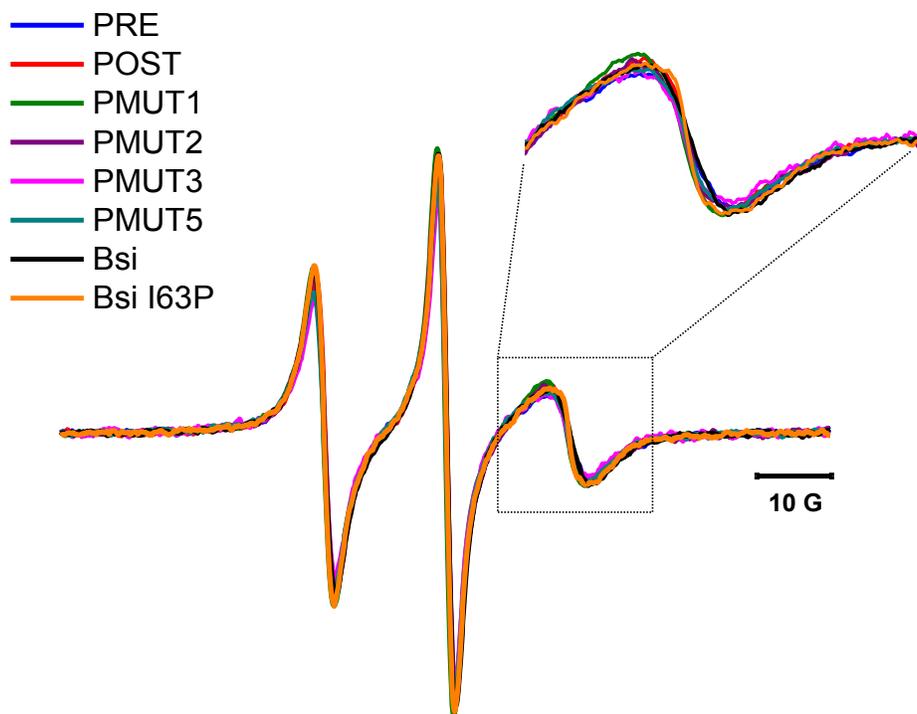


Figure 4-19. CW-EPR spectra for spin labeled HIV-1 PR PRE (blue), POST (red), PMUT1 (green), PMUT2 (purple), PMUT3 (magenta), PMUT5 (cyan), B_{si} (black), and B_{si}-I63P (orange). Inset shows an enlarged view of the high-field peak.

CHAPTER 5 DISTANCE MEASUREMENTS OF HIV-1 PROTEASE CONSTRUCTS

As described in Chapter 2, Double Electron-Electron Resonance (DEER) spectroscopy is a technique for determining the distance between specific sites in a protein molecule. In HIV-1 protease, the distance between spin-labeled sites on the flaps of each monomer can be measured. From the distance profiles obtained, the conformational ensemble, which corresponds to the sum of the protein conformational states, of various HIV-1 PR variants and subtypes can be compared. This chapter describes previous DEER studies on HIV-1 PR and results of DEER experiments on HIV-1 PR pre-therapy (PRE), post-therapy (POST), PMUT1 (POST-A82V), B_{si}, and B_{si}-I63P apo constructs (See Table 3-3 for the naming convention). DEER analysis was also performed on the PRE and POST constructs with the protease inhibitor, ritonavir (RTV).

The K55C mutation was introduced into the protease constructs to serve as a site for attaching the methanethiosulfonate spin label (MTSL). Previous studies have shown that spin-labeling the protein at this position, which is a solvent-exposed site, does not significantly alter the structure and activity of the HIV-1 protease (149). Based on X-ray crystal structures and MD simulation studies of HIV-1 PR, the distance between the terminal amine groups of K55 and K55' varies from 25 to 36 Å, which is suitable for DEER measurement (39, 45). Figure 5-1 shows X-ray crystal structures of HIV-1 PR with the flaps in the closed and semi-open conformations. The distance between the spin labels at positions K55 and K55' is predicted to decrease by approximately 3 Å, from about 36 Å to 33 Å, as the flaps shift from semi-open to a closed conformation.

Previous DEER Studies on HIV-1 PR

As previously mentioned, our group has utilized DEER spectroscopy to characterize the flap conformations of HIV-1 PR. Luis Galiano, a former member of our group, demonstrated that the DEER technique can be used to measure the distance between flaps of HIV-1 PR and thereby determine the flap conformations (45, 47). Results of the study revealed significant variation in the distance distribution profile of B_{sj} in the absence and presence of a protease inhibitor, ritonavir (RTV). Figure 5-2 shows the dipolar evolution curve and distance profile of apo and inhibitor-bound HIV-1 PR. A noticeable difference is seen in the dipolar evolution curves such that the inhibitor-bound protease consists of more oscillations compared to an almost dampened oscillation for the apo protease. This reflects the motion and flexibility of the flaps and the possible changes in conformation the protease undergoes upon binding of inhibitor. From the distance profile, it has been determined that the most probable distance of the apo HIV-1 PR is 35.4 Å, which corresponds to a semi-open conformation. In addition, the distance distribution is broad, which is indicative of flaps sampling a wide range of conformations. Upon addition of RTV, the most probable distance decreased by about 3 Å (from ~36 Å to ~33 Å) and the breadth of the distribution narrowed. This clearly indicates that the flaps closed-in to interact with the inhibitor and as such restricted the motion of the flaps, which is characteristic of a closed conformation.

Mandy E. Blackburn, also a previous member of our group, optimized the process of obtaining DEER data (49). By improving the signal-to-noise ratio (SNR), the sub-populations of HIV-1 PR can be determined from the distance profiles. Using the improvised technique, the distance distribution profiles of B_{sj} in the absence and presence of nine FDA-approved protease inhibitors and a substrate mimic, CA/p2, were

compared (Figure 5-3) (48). These results mainly showed the effect of inhibitor binding on the distance profiles. The inhibitors and CA/p2 substrate were classified based on how the inhibitor/substrate affected the closing of the flaps. CA/p2, RTV, LPV, TPV, SQV, APV, and DRV were classified as “strong/moderate”, where 70% of the conformational ensemble is in the closed conformation. The lower percentage of closed conformation for IDV, NFV, and ATV reveals a “moderate/weak” effect on flap closing. HIV-1 PR bound with these inhibitors possess a predominant semi-open conformation except for ATV, which has approximately 40% closed conformation.

Galiano et al. (46) studied the effect of mutations in drug-resistant HIV-1 PR variants, V6 (90) and MDR769 (103). V6 is a clinical isolate from a pediatric patient treated with RTV while MDR769 was isolated from a patient previously treated with IDV, NFV, SQV, and APV. Figure 5-4 shows DEER results for V6 and MDR769. Compared to B_{si} , the distance profile reveals that MDR769 has a longer interflap distance, corresponding to a more open conformation. In contrast, the most probable distance for V6 is shorter than B_{si} , which corresponds to a more closed conformation. This is in agreement with the MD simulation studies done by Simmerling et al. (39, 46). These findings provide evidence that drug-pressure selected mutations, particularly those that are relatively far from the active site, can influence HIV-1 PR function by altering flap motion and flexibility. Thus, distance distribution profiles, which provide information on flap conformations, can be used to compare the different HIV-1 PR variants containing drug-pressure selected mutations.

Blackburn extended Galiano's work on the V6 drug-resistant construct by studying the effect of addition of inhibitors on the distance distribution profile (47, 49). The apo

form of V6 was in agreement with Galiano's results in that the most probable distance is shorter than B_{si} apo. However, the breadth of the profile for V6 is broader whereas Galiano reported a broader profile for B_{si} , which may be attributed to the low signal-to-noise ratio (SNR) as seen in Figure 5-4. The effects of inhibitors on the distance profile of V6 were relatively similar to B_{si} with some notable exceptions. Similar to B_{si} , IDV, NFV, and ATV only slightly changed the most probable distance and the breadth of distribution. However, RTV, which had a strong effect on flap closing of B_{si} , has a minimal effect on the distance profile of V6 and almost no change in the breadth of the distance profile. All the other inhibitors had moderate to strong effect on the V6 distance profile, which was also observed in B_{si} . Overall, the effects of inhibitors on V6 compared to B_{si} were almost the same; however, the breadth of the distance profiles of V6 were relatively broader.

Jamie L. Kear, a previous member of our group, determined the distance distribution profiles of non-B subtype HIV-1 PR variants as well as previously studied drug resistant constructs, V6 and MDR769 (50, 51). Distance measurements on the non-B subtypes and drug-resistant constructs were carried out to examine the effects of naturally occurring polymorphisms and drug-pressure selected mutations on the flap conformation and flexibility of HIV-1 PR. Results showed that polymorphisms and drug-pressure selected mutations alter the average flap conformation and flexibility of HIV-1 PR. Figure 5-5 shows the distance distribution profiles of the apo forms of HIV-1 PR subtypes B, C, F, and CRF01_A/E and drug-resistant constructs V6 and MDR769. The most probable distance for almost all subtypes and the drug-resistant constructs are centered between 35 to 36 Å, with the exception of subtype C which has a slightly

longer inter-flap distance of about 37 Å. The differences in the overall shape and breadth of the profiles reveal a variation in flap conformation and flexibility as a consequence of the polymorphisms and mutations. Kear also studied the effects of the various FDA-approved inhibitors on subtypes CRF01_A/E and F (51). The results for both non-B subtypes were similar to that of subtype B_{SI} in that IDV, NFV, and ATV had a “weak/moderate” effect on flap closing whereas all the other inhibitors displayed a “strong” effect.

As previously cited, Galiano and Blackburn reported a more open conformation for MDR769 and a more closed conformation for V6 (47, 49). Kear determined that the most probable distance of V6 and MDR769 is almost the same as B_{SI} and the predominant flap conformation for both constructs is the semi-open conformation (51). Compared to B_{SI}, MDR769 has a larger relative percentage of wide-open conformation whereas V6 shows a greater relative percentage of tucked/curled conformation.

Materials and Methods

Chemicals, reagents, and supplies were obtained from Fisher Scientific (Pittsburgh, Pennsylvania) unless otherwise indicated. Deuterated solvents were purchased from Cambridge Isotope Laboratories, Inc. (Andover, Massachusetts). Ritonavir (RTV) was generously received from the AIDS Research and Reference Reagent Program, Division of AIDS, NIAID, NIH (Bethesda, Maryland).

Sample Preparation for DEER Data Collection

Inactive HIV-1 PR pre-therapy (PRE), post-therapy (POST), PMUT1 (POST-A82V), B_{SI}, and B_{SI}-I63P constructs were purified and spin-labeled as described in Chapter 4. The purified and spin-labeled samples were concentrated to OD₂₈₀ ~2.5 to 2.7 using an Amicon® Ultra-4 Centrifugal Filter Unit with Ultracel-3 membrane. The

sample was buffer exchanged as follows: 1 mL of concentrated HIV-1 PR was diluted with 3 mL deuterated 20 mM NaOAc ($C_2D_3O_2$) in D_2O (pH 5.0) and centrifuged until the volume of sample was approximately 1 mL. This process was repeated 3 times such that by the end of the fourth concentrating step, the sample buffer is approximately 100% deuterated solvent. Also, the OD_{280} was obtained to verify if it is roughly the same as the initial OD_{280} . To avoid multiple freeze/thaw cycles, 70 μ L of the spin-labeled protein in deuterated buffer was aliquoted into 0.2 mL PCR tubes and stored at $-20^\circ C$.

For apo samples, 30 μ L of deuterated glycerol was directly added to 70 μ L of sample so that the final concentration of glycerol is approximately 30%. For RTV-bound samples, a 4:1 molar excess of inhibitor (typically 3 to 6 μ L) was added to 65 μ L of sample. The sample was left standing for 15 to 30 min to allow sufficient time for binding, after which deuterated glycerol was added to the sample. Samples were mixed thoroughly and transferred to a 4-mm quartz EPR tube using a syringe fitted with plastic tubing.

DEER Experiments

DEER data were collected on a Brüker EleXsys E580 EPR spectrometer (Billerica, Massachusetts) equipped with the ER 4118X-MD-5 dielectric ring resonator at a temperature of 65 K. The EPR tube containing the sample was “flash-frozen” using liquid nitrogen and inserted into the resonator. A series of preliminary experiments was performed to determine the following parameters: center field, T_m , d_0 (time in ns at which the echo begins), pulse gate (the breadth of the echo in ns), and appropriate positions for the observer and pump frequencies. The “pump” frequency is applied to the central resonance located at approximately 3460 G in the spectrum (Figure 5-6).

The “observer” frequency is applied at the low-field resonance approximately 26 G lower than the “pump” frequency. The typical parameters are listed in Table 5-1.

The four-pulse DEER sequence (Figure 2-11), described in Chapter 2, is then applied to collect DEER data. The pulse parameters used with the Xepr software package from Bruker Biospin are given in Table 5-2.

DEER Data Analysis

Data analysis was carried out as described in Chapter 2. The raw experimental data was processed *via* Tikhonov Regularization (TKR) using DeerAnalysis2008 software to generate distance distribution profiles and background subtracted dipolar evolution curves. The level of background subtraction was optimized *via* the self-consistent analysis (SCA). Finally, distance profiles were fit with a series of Gaussian functions. Population validation, described in Chapter 2, was done as necessary.

Results and Discussion

As described earlier in this chapter, the flap conformations of different HIV-1 PR subtypes (B, C, F, and CRF01_A/E) and drug-resistant constructs (V6 and MDR769) were previously determined by our group using DEER spectroscopy (45-51). The distance distribution profile represents the conformational ensemble, which corresponds to the sum of the sub-populations that exist in HIV-1 PR (49). The four major conformational states or sub-populations, determined from NMR and ITC studies and modeled by molecular dynamics (MD) simulations, are referred to as semi-open, closed, wide-open, and tucked/curled conformations (28, 29, 33-37). In this work, HIV-1 PR variants pre-therapy (PRE), post-therapy (POST) and PMUT1 (POST-A82V), a construct containing drug-pressure selected mutations, were analyzed by DEER

spectroscopy. DEER results obtained were compared with B_{si} and B_{si}-I63P apo constructs. Also, the PRE and POST constructs were analyzed after addition of the protease inhibitor, ritonavir (RTV). The spin labeling efficiency of the HIV-1 PR samples were verified by mass spectrometry (MS) and CW-EPR as described in Chapter 4. MS and CW-EPR results for the B_{si} and B_{si}-I63P constructs were presented in Chapter 4. However, for the PRE, POST and PMUT1 constructs, DEER analysis was performed on older samples, for which MS and CW-EPR results are given in Appendix B.

Comparison of HIV-1 PR Apo Constructs

The comprehensive results of DEER data analysis for the apo constructs are shown in Figures 5-7 to 5-11. Figures 5-7 to 5-11A show the overlay of the theoretical dipolar evolution curve from Gaussian reconstruction and the dipolar modulation fit to the experimental background subtracted echo. As described in Chapter 2, the TKR-generated dipolar modulation and theoretical curves are overlaid to ensure the curves matched. This procedure ascertains that the most accurate background subtraction is chosen. Figures 5-7 to 5-11B show the overlay of the TKR-generated distance distribution and the distance profile obtained by Gaussian reconstruction. The process of Gaussian reconstruction is necessary for generating the individual populations that sum up to form the distance profile. This method determines the center, full width at half maximum (FWHM), and relative percentage of each peak, which are important parameters that characterize each individual population. Figures 5-7 to 5-11C show the Gaussian populations and Table 5-3 to 5-7 lists the distance, FWHM, and relative percentage of the individual populations. Figures 5-7 to 5-11D show the L-curve used for choosing the appropriate regularization parameter (α). As stated in Chapter 2,

the optimum value of α usually corresponds to the corner of the L-curve. However, this was not applicable here, for reasons that will be explained later in the Experimental Limitations section. Figures 5-7 to 5-11E show the Pake dipolar pattern resulting from Fourier transformation of the background subtracted echo and TKR-generated fit.

Figure 5-12A and B shows the dipolar modulation and distance distribution profile, respectively, for the apo constructs generated by TKR analysis. Long-pass filter, a digital filtering system that selectively attenuates shorter wavelengths, was applied to generate a dipolar evolution curve with an enhanced signal-to-noise ratio (SNR) and consequently improve DEER data analysis. Comparison of the dipolar evolution curves reveals some variations in the oscillations. The oscillations for B_{si}, POST, and PMUT1 constructs are relatively broad and not well-defined whereas oscillations for the PRE and B_{si}-I63P are more pronounced. Dipolar evolution curves that have well-defined oscillations correspond to narrower distance distributions and shorter most probable distance, which is the most intense point of the distance distribution profile (50). The shift in the first minimum of the echo curve (marked by a dashed vertical line on Figure 5-12A) indicates a difference in the frequency of the oscillations and variation in the most probable distance of the distance distribution profile (Figure 5-12B) (49).

Comparison of the most probable distance for HIV-1 PR apo constructs, given in Table 5-8, reveals that PRE and B_{si}-I63P have the same center at 33.2 Å whereas B_{si} and POST have nearly similar values at 36.8 Å and 37.0 Å. The PMUT1 construct has a relatively farther most probable distance of 38.5 Å, where there is a noticeable splitting of the distance distribution into populations. This can be seen in the echo curve of PMUT1, where the first minimum is slightly shifted to the right, indicating a relatively

longer distance compared to the other constructs. Figure 5-13 shows the distance profiles of PRE, POST, PMUT1, and B_{si}-I63P overlaid with the distance profile of B_{si}. The shift in most probable distance relative to B_{si} is clearly evident for the PRE and B_{si}-I63P constructs. Figure 5-14 also shows a comparison of most probable distance between constructs.

The Gaussian reconstruction process reveals the sub-populations that comprise the distance distribution profile. The individual populations correspond to the four major flap conformations: tucked/curled (25-30 Å), closed (~33 Å), semi-open (36-38 Å), and wide-open (40-46 Å). The tucked/curled conformation refers to the state of HIV-1 PR when its flaps tuck or curl into the active site pocket (41, 44). This has been observed in MD simulations and the distances measured were consistent with findings from DEER studies. However, the exact distance assignment of these conformations has not been established. The Gaussian populations reported for each of the construct are based on results of the population validation process. Population validation was done to verify if some of the minor populations need to be suppressed. A comprehensive summary of the population validation process is given in Appendix C.

Each of the peaks assigned in the Gaussian profile should represent real protein conformational states rather than artifacts or peaks arising from background or noise. For example, as seen in the background subtracted dipolar evolution curve of the PRE construct on Figure 5-7A, the echo signal between $\tau = 2$ to 3 μs appears to have a large amount of noise compared to experimental data previously collected by our group. To verify this, distance profiles were obtained at different truncated values of τ at 2.0 μs , 2.5 μs , and 3.0 μs . Figure 5-15 shows the dipolar evolution curves, the corresponding

Gaussian population profiles, and the L-curves. Table 5-9 shows the relative percentages of the individual populations at the indicated τ values. When τ is truncated, it is noticeable that the populations at longer distances corresponding to the wide-open conformation are shifted to a shorter distance. At $\tau = 2.0 \mu\text{s}$, the peak corresponding to the semi-open conformation, disappeared. In effect, truncation of the τ value to $2.0 \mu\text{s}$, altered the relative percentage of the populations for the closed and wide-open conformations. Other than the large shift in the distance of the wide-open conformation, there is only a slight difference in the relative percentage of the populations at $\tau = 2.5 \mu\text{s}$ and $\tau = 3.0 \mu\text{s}$. In addition, only very minor changes were observed in the breadth of the populations for all τ values. Overall, the effect of truncating τ is the reduction of the span of the distance profile because essentially removing some of the data at longer τ changes the level of background subtraction and manifests as a shift in the population at longer distances. However, it is still uncertain whether this results in the removal of noise and whether the population assigned to a wide-open conformation is real or not.

Table 5-10 and Figure 5-16 show the relative percentage of sub-populations for each of the HIV-1 PR apo constructs. The population analysis is also presented as a 3D bar graph in Figure 5-17. The most predominant conformation for PRE and B_{si}-I63P is the closed conformation whereas B_{si}, POST, and PMUT1 have the largest percentage for the semi-open conformation. Results for B_{si} are consistent with previous findings although with a small variation in the center and relative percentages for each sub-population. This may be attributed to the inefficient labeling of the B_{si} construct, which will be discussed in further detail in the Experimental Limitations section.

The population breadth, which is measured by the FWHM, reflects the motion of the spin label in each conformation and the range of conformational sampling in the protein. A broad distribution indicates that the spin label is more flexible and that the flaps sample a wide range of conformations (45). Distributions corresponding to semi-open and wide-open conformations are usually broad whereas the closed conformation typically has a narrower population breadth. Upon addition of inhibitor, HIV-1 protease adopts the closed conformation, where the flexibility and range of HIV-1 PR flap motion is limited (39, 45). DEER analysis of HIV-1 PR in the presence of ritonavir (RTV) showed that the breadth of the distance profile is 3.0 Å, which is narrower than that of the apoenzyme (45). In the inhibited state, it is known that large amplitude backbone motions are restricted and that the flaps undergo only very small and rapid oscillations (28, 150). Thus, the narrow breadth of distance profile for HIV-1 PR with inhibitor is attributed to the motion of the spin label about the flexible linker (49).

The FWHM values of the semi-open population for the B_{si} and POST are 5.6 Å and 5.1 Å, respectively, which correspond to a relatively wide population breadth compared to the narrower breadth for the closed conformation at about 4 Å for both constructs. This is consistent with results of previous DEER studies on apo B_{si} HIV-1 PR (49, 51). For the PRE and B_{si}-I63P constructs, the population breadths of the closed conformation are 3.8 Å and 3.5 Å, respectively. However, the difference between the FWHM of the closed and semi-open populations for the PRE construct is 0.4 Å, which is evidently small, and for B_{si}-I63P the FWHM values are the same. Because PRE and B_{si}-I63P have a large percent closed population of 56% and 57%, respectively, >50% of the population adopt a conformation in which the motion of the flaps is restricted. As the

protease transitions from closed to the open conformations, there is little to no change in the range of motion, as reflected in the values for FWHM. Interestingly, results for PMUT1 are different in that the FWHM values are smaller than expected for a construct having a predominant semi-open conformation. The population breadth for both closed (3.4 Å) and semi-open conformation (3.6 Å) is almost the same and the breadth is smaller for the wide-open conformation at 2.7 Å.

DEER results for the PRE construct are unexpected because an apo construct with a large percentage of closed conformation has not been previously reported. X-ray crystal structures of ligand-free HIV-1 PR primarily exhibit the semi-open conformation (28, 29, 33-37, 40). This has been confirmed by DEER studies for subtypes B, C, CRF_01 A/E, F and multi-drug resistant constructs V6 and MDR769 (48-51). It has been shown that V6 has a shorter most probable distance and is more closed compared to B_{si}, but its predominant conformation is semi-open. Crystal structures of all inhibitor-bound HIV-1 PR are found in the closed conformation because the flaps, particularly the Ile50/50' amide group, close-in and interact with the inhibitor *via* a conserved water molecule (27). This is consistent with DEER results and MD simulations that inhibitor-bound HIV-1 PR adopts the closed conformation (39, 45). However, MD simulation studies of unbound closed HIV-1 PR, which is achieved by removing the inhibitor from the bound state of HIV-1 PR, showed that the average distance and profile breadth are similar to the bound closed simulations (39). In addition, intraflap hydrogen bonding, which may have stabilized the closed flap conformation, was observed in the simulations. These results indicate that despite the loss of flap-inhibitor interactions, ligand-free protease may adopt a closed conformation.

The predominant closed conformation for the PRE construct may be due to some interactions within the protease that stabilize the closed conformation. It is possible that the closed conformation is stabilized by specific amino acid residues, which are located in critical sites of HIV-1 PR. This led to careful examination of the amino acid differences of PRE relative to B_{si}. PRE and B_{si} differ in amino acids at sites 37, 39, 62, and 63, which are located in the hinge area (Figure 5-18). In drug resistant HIV-1 PR variants, the L63P mutation is identified as a secondary mutation that compensates for the decrease in function and activity as a consequence of primary mutations. Interestingly, a number of studies have revealed that the L63P mutation, although associated with progression to drug resistance, is also commonly present in drug-naïve patient isolates. The presence of proline at site 63 in the PRE construct may have some functional advantage and its location in the protein possibly leads to a more rigid protease and less flexible flaps. It is also possible that the L63P mutation is preserved in HIV-1 PR pre-therapy variants to maintain its activity in the presence of inhibitors. Thus, it is of interest to investigate the effect of the L63P mutation on the flap conformations of HIV-1 PR. DEER analysis on the B_{si}-I63P mutant was performed to compare its flap conformation with the B_{si} and PRE constructs. Recall that the mutation is I63P rather than L63P because the B_{si} construct contained the L63I stabilizing mutation. This further supports that position 63 is a critical site that can affect the stability of the protease. DEER results clearly showed a shift in the most probable distance of the B_{si}-I63P construct relative to B_{si} and that it has a large percentage of closed conformation of 57%, similar to the PRE construct. This provides evidence that a single secondary mutation can drastically shift the conformational ensemble of HIV-1 PR. Additionally, the

proline at site 63 in the PRE construct may have contributed largely to its conformational ensemble.

The combined percentage of semi-open and wide-open conformations for the POST construct (74%) is greater than PRE (18%) and the percent population for the wide-open conformation (26%) is greater relative to B_{si} (12%). Because the POST construct is composed of several mutations, which are selected for by RTV-IDV combination inhibitor therapy, these may have a profound effect on the flap conformation and flexibility. Relative to PRE, it contains mutations in the active site (V82A), near-active site (L10I and I15V), flap (I54A and Q58E), and hinge (E34Q, M36I, and T37N) regions (Figure 5-18). The flap and hinge mutations are located in sites that can influence the motion and flexibility of the flaps. Therefore, the combined effect of the mutations is to shift the conformational ensemble, in which the majority of the population is in an open conformation. Additionally, the broader than average breadth for the distance distribution profile compared to previously reported results on other constructs, indicates either an increase in flexibility of the flaps or the flaps are relatively unstable in the closed conformation.

The PMUT1 construct consists of the same mutations as the POST except that the mutation at position 82 was back-mutated to Val, which is present in the PRE construct. The V82A mutation is a primary mutation common in most drug-resistant constructs such as in V6 and MDR769. Because it is an active site mutation, it directly affects inhibitor binding. It is associated with the initial decrease in effectiveness of RTV and also observed in patient isolates receiving IDV and LPV (58, 81). As discussed in Chapter 1, the substitution of Ala with Val was studied by Ho et al. to determine the

effect of reverting a particular site in the POST sequence to a pre-therapy amino acid residue, on viral replicative capacity and PI sensitivity (100). The amino acid substitution resulted in 50% recovery of replicative capacity and about 50% increase in sensitivity to either RTV or IDV relative to PRE. In this work, the effect of the A82V mutation on the flap conformation and flexibility was investigated. DEER results revealed that the predominant conformation of PMUT1 is semi-open. However, the percent closed population of 33% is almost 2 times larger than that for POST (18%) and the percentage of wide-open conformation decreased markedly from 29% for POST to 4% for PMUT1. In addition, the percentage of closed conformation increased to almost 50% of PRE, indicating that there may be some rearrangement in the active site pocket occurring in a certain fraction of the population such that the flaps adopt a more closed conformation. Overall, the changes observed in the populations of PMUT1 relative to PRE and POST provides evidence that a single active site mutation can dramatically alter the protein conformational ensemble.

Comparison of HIV-1 PR Constructs With and Without RTV

As previously stated, the effect of addition of inhibitor to HIV-1 PR is a shift from semi-open to a closed conformation. Upon inhibitor binding, the flaps are expected to close-in on the active site and interact with the inhibitor, allowing HIV-1 PR to adopt a rigid conformation known as the closed conformation. NMR studies on flap dynamics and structure-based analysis on binding energetics have provided evidence that inhibitor-bound HIV-1 PR adopts a closed conformation (28, 29, 33, 34). DEER results of different HIV-1 PR subtypes have shown that the inter-flap distance shifts from about 36 Å to 33 Å upon inhibitor binding (48-51).

Multi-drug resistant constructs contain both primary and secondary mutations that influence flap conformation and flexibility. It is expected that a drug-resistant HIV-1PR, which is previously exposed to an inhibitor that selected for mutations in the protease, will not bind to a protease inhibitor as effectively. The active site mutations directly affect protease-inhibitor interaction whereas non-active site mutations alter the motion of the flaps by either preventing HIV-1 PR to reach the wide-open conformation and thus disallowing access to inhibitor or preventing the flaps from closing so that the inhibitor cannot interact with active-site amino acid residues effectively (23). A multi-drug resistant construct, V6, was analyzed by DEER to determine the effect of addition of inhibitor on the flap conformation (49). Results were mostly similar to B_{si}, with one notable exception. The protease inhibitor, RTV, which had a strong effect on flap closing of B_{si}, has only a minimal effect on the distance distribution profile of V6. These results are expected since the mutations present in V6 were selected under RTV therapy.

In this work, the effect of addition of RTV to the PRE and POST HIV-1 PR variants was investigated. DEER data analysis of the PRE and POST constructs in the presence of ritonavir (RTV) are given in Figures 5-19 and 5-20 and a summary of the Gaussian population profiles is listed on Tables 5-11 and 5-12. A comparison between the dipolar evolution curves and distance distribution profiles of HIV-1 PR in the absence and presence of RTV is shown in Figure 5-21.

The dipolar modulation of PRE in the presence of RTV (Figure 5-21A) clearly reveals a shift in the first minimum and an increase in the frequency of oscillations compared to PRE apo. This indicates a shift in the distance distribution to shorter distances and a narrower breadth of population. There is a considerable decrease in the

breadth of each population; however, Figure 5-21C shows that the most probable distance of both apo and bound PRE is exactly the same at 33.2 Å. The observed changes in oscillations in the dipolar evolution curve manifest as a variation in the percentages of the sub-populations. Table 5-13 presents a summary of the relative percentage for each sub-population and Figure 5-22 shows separate bar graphs for each population. The differences in the relative percent for each population between the constructs with and without RTV are clearly illustrated in Figure 5-23. Compared to PRE apo, values for PRE in the presence of RTV reveal a slight increase in the percent closed population (from 56% to 65%) and small decrease in the semi-open conformation (from 17% to 7%). This indicates that some fraction of the population may be accounted for by RTV-bound PRE. The total percent tucked/curled conformation and the percentage population of wide-open, however, are almost the same.

In contrast to PRE, the dipolar evolution curves of POST in the absence and presence of RTV (Figure 5-21B) do not reveal any changes in the frequency of oscillations. Although the oscillation for the POST with RTV is slightly more well-defined than POST apo, both echo curves are relatively broad indicating wider distribution profiles. The noticeable downward shift in the first minimum of dipolar evolution curve of POST in the presence of RTV may be attributed to the shift in the most probable distance from 37.0 Å to 36.0 Å as seen in the distance distribution profile (Figure 5-21D), but it may also be due to the large change in percentage of populations at longer distances. Figure 5-23 clearly shows a relatively large decrease in the percentage of wide-open conformation of 17%. However, there is almost no change in the percent semi-open population relative to POST apo. A slight increase in the percentages of

closed (from 18% to 24%) and tucked/curled (from 8% to 17%) conformations is also observed. The increase in % tucked/curled and % closed and corresponding decrease in % wide-open possibly indicates the presence of RTV-bound HIV-1 PR.

To some extent, the DEER results for the PRE construct in the presence of RTV are inconsistent with the expected results that there should be a significant change in the percentage of closed population. This is not surprising, however, because the PRE apo construct is predominantly closed, indicating that its flaps are less flexible. Having a restricted motion in the flaps decreases the likelihood of flap opening and allowing access of an inhibitor to the active site cavity. Thus, only a small fraction of the closed population may be accounted for by RTV-bound HIV-1 PR. As mentioned previously, PRE contains some amino acid residues that are different from B_{si} and this may have some effect on the flexibility of the flaps. Although the variant is a pre-therapy isolate, it contains the L63P mutation, which is associated with progression to high-level drug resistance. It is possible that the proline at that particular site has some impact on the flap motion and conformation of HIV-1 PR and a subsequent effect on inhibitor binding.

DEER results for POST in the presence of RTV are consistent with expectations for a drug-resistant construct in that addition of inhibitor has a minimal effect on the conformational ensemble. In addition, these findings, which are similar to V6, are expected because the mutations present in the POST construct were selected under RTV therapy. The combination of mutations may have stabilized the wide-open conformation, which prevents the flaps from closing and binding with the inhibitor. Thus, drug-pressure selected mutations clearly influence the motion and flexibility of HIV-1 PR, which ultimately determines the binding of inhibitor.

Experimental Limitations

A number of experimental factors affect the quality of DEER data obtained. These include quality of sample, signal-to-noise ratio (SNR), and maximum dipolar evolution time (t_{max}). The sample is considered of good quality on the basis of spin labeling efficiency and absence of any protein aggregation. Obtaining a sample of good quality was addressed in Chapter 4. This is essential because it determines the signal-to-noise ratio (SNR), which ultimately affects the quality of DEER data (49). Most of the signal collected from DEER is attributed to background, which means only a small fraction is due to the spin-spin interaction of interest. If a sample is poorly spin-labeled, it will contain mostly unlabeled or singly-labeled proteins, resulting in a decrease in the DEER echo signal. Aggregation of protein is not particularly desirable for DEER analysis because it increases the signal due to background and dampens oscillations in the dipolar evolution curve. In addition, in the presence of aggregation, the intermolecular distance can be smaller and thus overlap with the population of spins at longer distances, which complicates DEER data analysis.

Signal-to-noise ratio (SNR) largely affects the quality and accuracy of the distance distribution profile (49). As mentioned, good SNR can be obtained in a short amount of time if samples are efficiently spin-labeled. Samples with low SNR usually contain a large amount of noise that masks oscillations in the data. As demonstrated earlier, the effect of truncating τ is to shift the farthest populations to shorter distances. However, this method did not determine whether some noise was effectively removed and that assignment of the population at longer distances is still uncertain. Low SNR is also manifested as a distortion in the shape of the L-curve. This complicates DEER data

analysis because the corner of the L-curve, which often represents the optimum value, is obscured. The distortion is primarily observed in the small α -value region, where the curve is dominated by the smoothing function ($\log \eta$). An indication of poor SNR is the large increase in deviation ($\log \rho$) at small α values (49). This can be observed in almost all of the DEER results presented with a few exceptions, particularly B_{si}-I63P, which clearly has good SNR based on its experimental dipolar evolution curve (Figure 5-10A). Samples with a large amount of noise seen in the experimental dipolar modulation usually have distorted L-curves, which is particularly observed for POST-RTV. Other constructs, such as PRE apo, PRE RTV and POST apo, have higher SNR relative to POST RTV but do not have perfectly-shaped L-curves, which led to difficulty in selecting the optimum α value. In contrast, PMUT1 and B_{si}, which have relatively lower SNR compared to B_{si}-I63P, adopt a normal shaped L-curve. However, selection of the optimal α value for PMUT1 is complicated. Another consequence of poor SNR is the dampening of oscillations leading to broader than average distance profiles. This was particularly observed in the POST and B_{si} samples. The discrepancies in the results of the B_{si} construct compared to previous results reported by our group may be attributed to the spin labeling efficiency, which was reported in Chapter 4 to be only about 70-80%, and may have resulted in a lower SNR. The MS data for the PRE construct (Figure B-3) also shows underlabeling of the sample, which may have an effect on the L-curve. The POST and PMUT1 constructs have acceptable spin labeling efficiency based on the MS data shown in Figures B-4 and B-5, respectively. However, DEER analysis of the samples is complicated because of poor SNR of data for both constructs. B_{si}-I63P, which was reported in Chapter 4 to be efficiently labeled, is the only construct

that produced an acceptable SNR and therefore did not complicate DEER analysis. The possible inaccuracies and uncertainty of the DEER results, particularly in assigning conformations corresponding to the minor populations, are certainly attributed to low SNR.

Spin labeling efficiency and SNR are important factors that determine the time for data collection. In addition, the time for collecting data also depends on the maximum dipolar evolution time, t_{max} . It has been determined that increasing the t_{max} results in a decrease in the breadth of the distance profile, which correspond to dipolar evolution curves having more frequent oscillations (49). Therefore, to obtain a distance profile with an accurate breadth of population, the dipolar evolution curves must be collected at sufficiently long t_{max} to capture at least two oscillations.

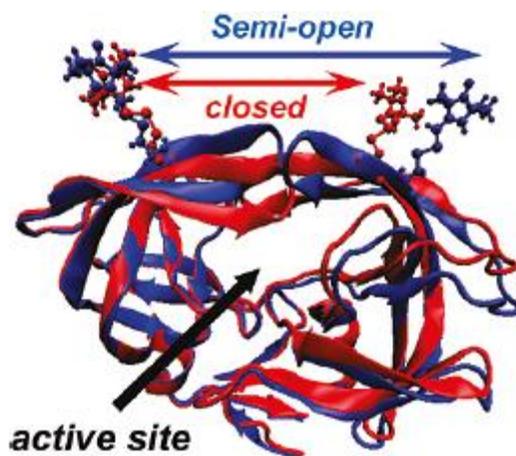


Figure 5-1. HIV-1 PR crystal structures showing the (red) closed (PDB ID 2BPX) and (blue) semi-open (PDB ID 1HHP) flap conformations with spin labels at the K55C and K55C' sites. Reprinted with permission from Blackburn et al. (48). Copyright 2009 American Chemical Society.

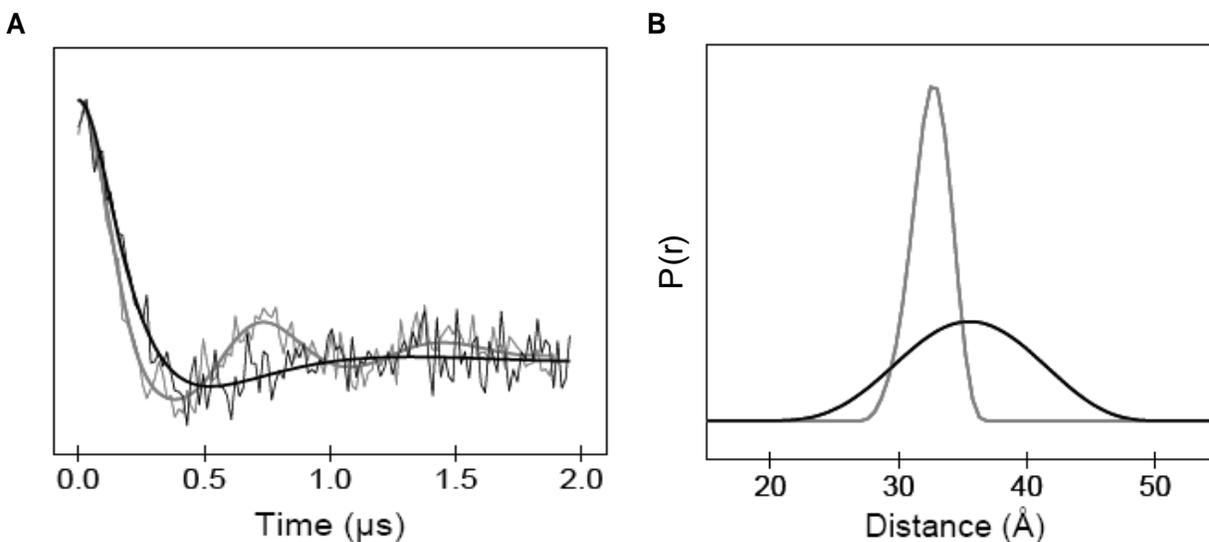


Figure 5-2. A) Dipolar evolution curves for apo (black) and RTV-bound (gray) subtype B HIV-1 PR and B) corresponding distance distribution profiles. Figure adapted from Galiano (47).

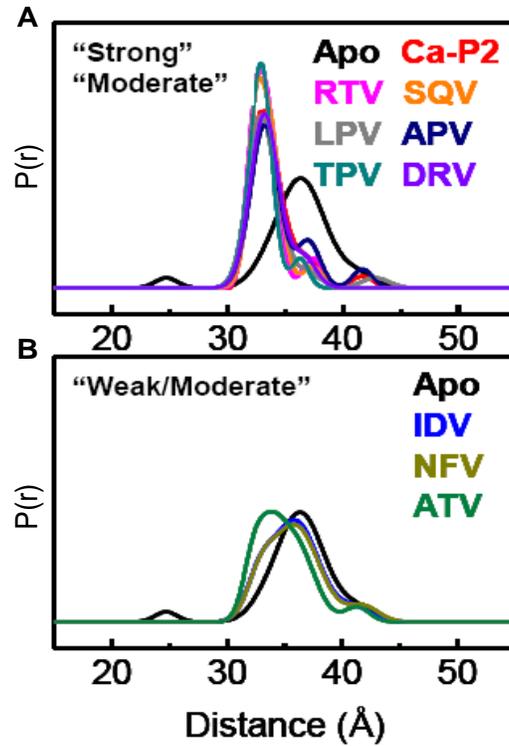


Figure 5-3. Distance distribution profiles of subtype B HIV-1 PR apo and with inhibitors showing (A) "strong/moderate" and (B) "weak/moderate" effect on flap closing. Figure adapted from Blackburn (49).

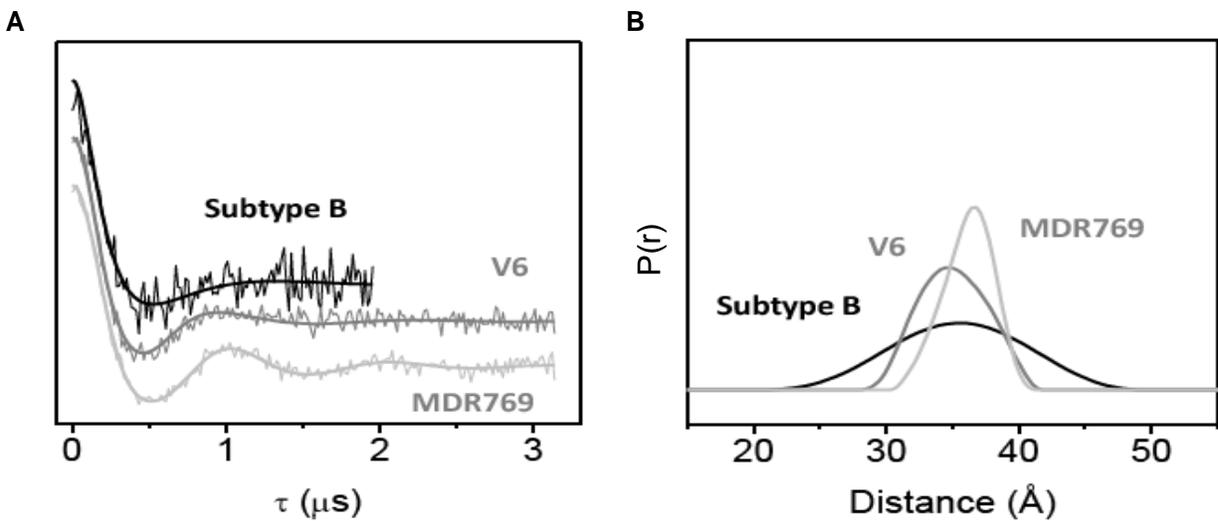


Figure 5-4. A) Dipolar evolution curves for subtype B apo (black), V6 (gray), and MDR769 (light gray) and B) corresponding distance distribution profiles. Figure adapted from Galiano (47).

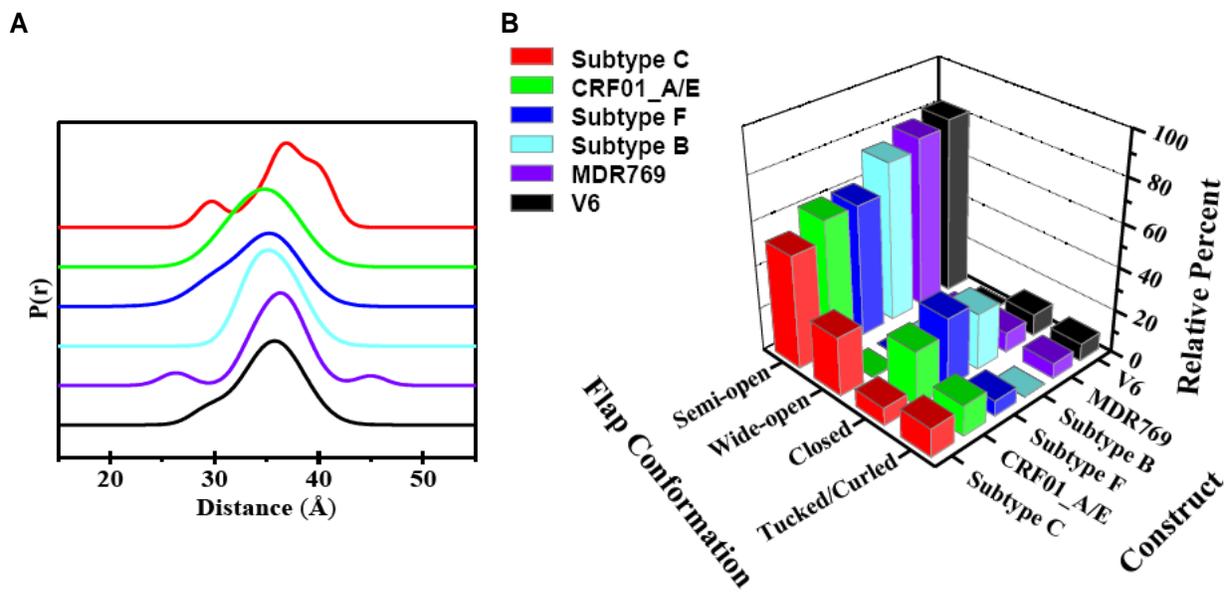


Figure 5-5. A) Distance distribution profiles of apo forms of HIV-1 PR subtype B, C, F, CRF01_A/E and multi-drug resistant constructs, V6 and MDR769, and B) population analysis summarizing the relative percentage of each conformation. Figure adapted from Kear (51).

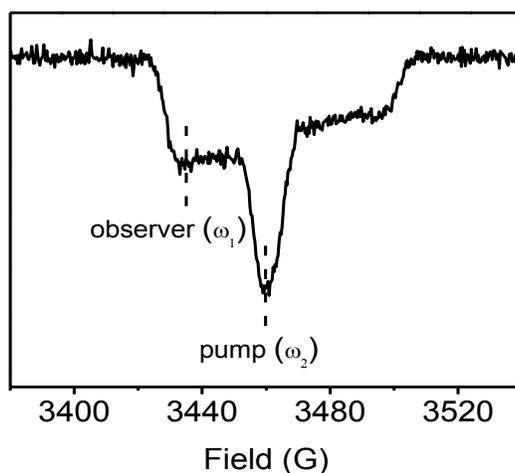


Figure 5-6. Absorption spectra for a nitroxide spin-label. The peak maximum at the center field is referred to as “pump” frequency (ω_1) and the low field transition is referred to as the “observe” frequency (ω_2).

Table 5-1. Standard parameters for DEER data collection.

Parameter	Value
Sweep width	160 G
Shot repetition time	4000
Shots/point	100
Center Field	~3460 G
Low Field	~3432 G
Frequency	~9.6 GHz
Pulsed Attenuation	0.1
Video Bandwidth	25 MHz
Modulation Amplitude	~1 G
Time Constant	0.082-0.164 s
Receiver Phase	100
Number of Scans	Variable

Table 5-2. Parameters used for the pulse sequence.

Parameter	d0	d1	d2	d3	PG	dx
Value	100	200	3000	100	220	12

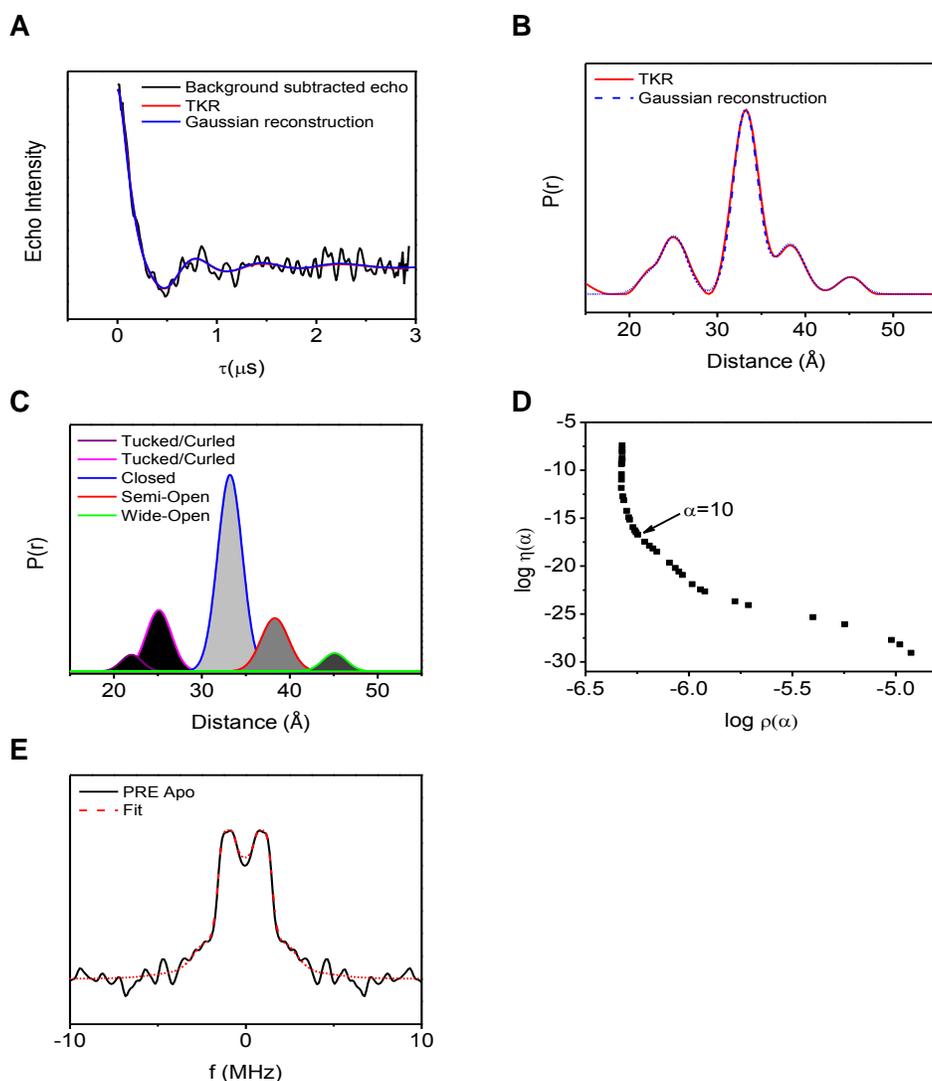


Figure 5-7. DEER data analysis for PRE apo. A) Background subtracted dipolar echo curve (black) overlaid with the TKR-generated dipolar modulation (red) and Gaussian reconstructed dipolar modulation (blue), B) distance distribution profile from TKR analysis (red) overlaid with the sum of the populations obtained by Gaussian reconstruction (blue dashed), C) individual Gaussian populations, D) L-curve ($\alpha = 10$), and E) Pake dipolar pattern.

Table 5-3. Individual population profiles for PRE apo.

	Center (Å) (± 0.3 Å)	FWHM (Å) (± 0.5 Å)	Relative Population (%) ($\pm 5\%$)
Tucked/Curled 1	22.0	2.8	4
Tucked/Curled 2	25.1	3.4	18
Closed	33.2	3.3	56
Semi-Open	38.3	3.7	17
Wide-Open	45.1	3.2	5

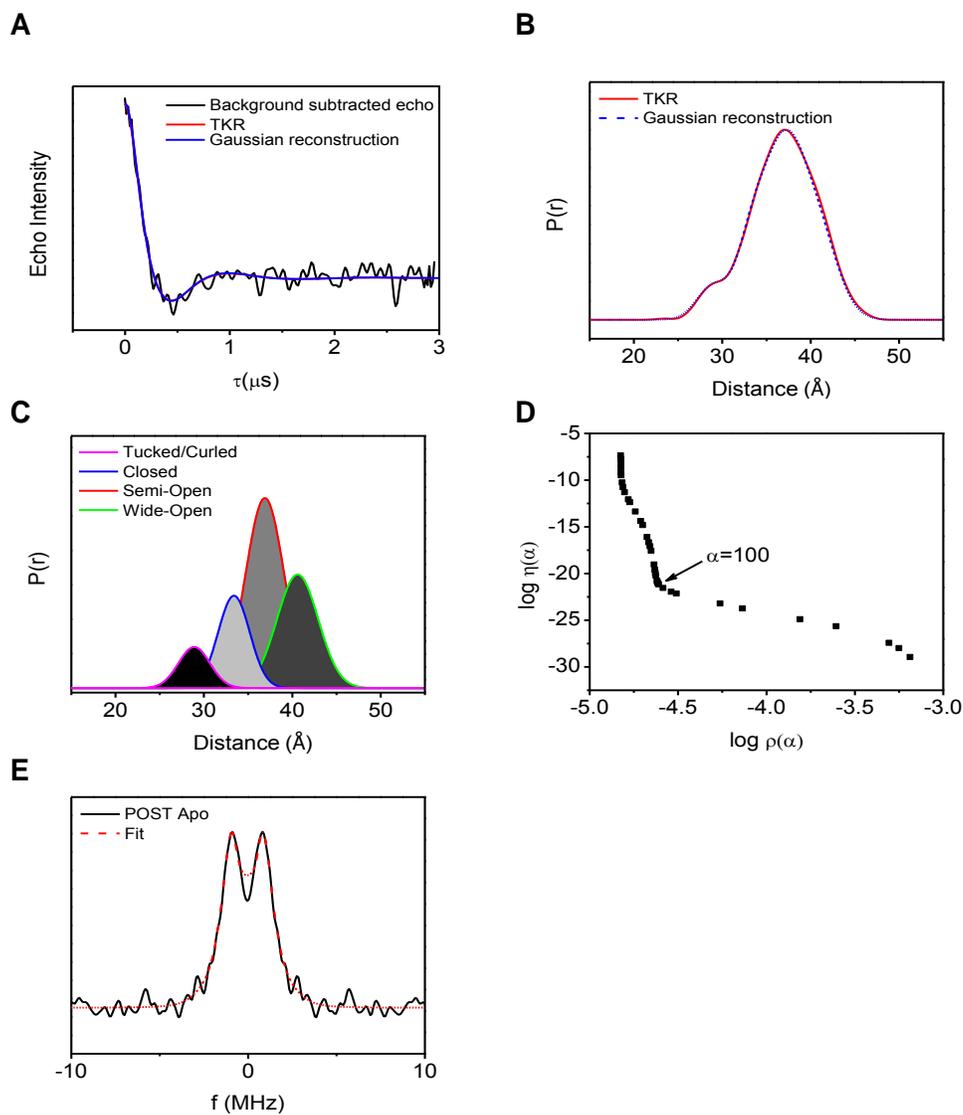


Figure 5-8. DEER data analysis for POST apo. A) Background subtracted dipolar echo curve (black) overlaid with the TKR-generated dipolar modulation (red) and Gaussian reconstructed dipolar modulation (blue), B) distance distribution profile from TKR analysis (red) overlaid with the sum of the populations obtained by Gaussian reconstruction (blue dashed), C) individual Gaussian populations, D) L-curve ($\alpha = 100$), and E) Pake dipolar pattern.

Table 5-4. Individual population profiles for POST apo.

	Center (Å) (± 0.3 Å)	FWHM (Å) (± 0.5 Å)	Relative Population (%) ($\pm 5\%$)
Tucked/Curled	28.9	4.2	8
Closed	33.4	4.2	18
Semi-Open	36.9	5.1	45
Wide-Open	40.6	5.5	29

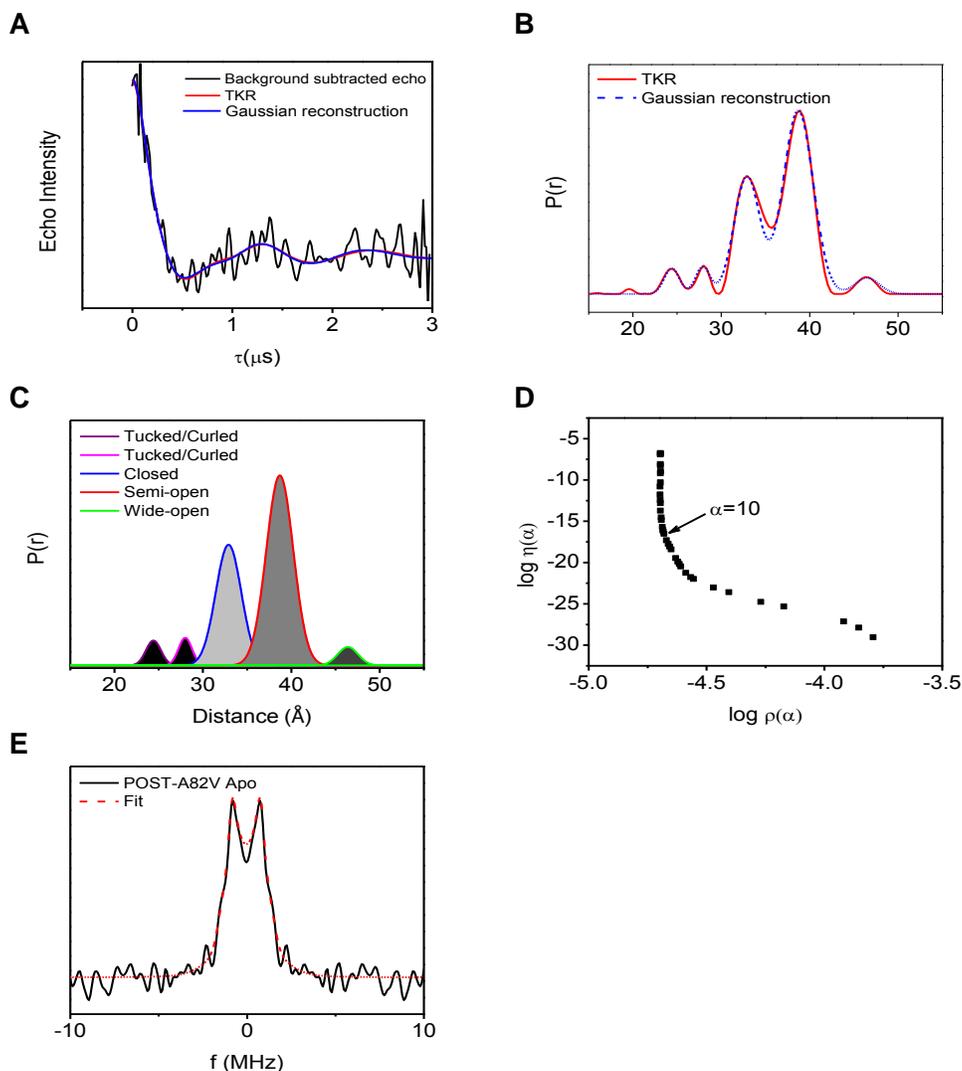


Figure 5-9. DEER data analysis for PMUT1 apo. A) Background subtracted dipolar echo curve (black) overlaid with the TKR-generated dipolar modulation (red) and Gaussian reconstructed dipolar modulation (blue), B) distance distribution profile from TKR analysis (red) overlaid with the sum of the populations obtained by Gaussian reconstruction (blue dashed), C) individual Gaussian populations, D) L-curve ($\alpha = 10$), and E) Pake dipolar pattern.

Table 5-5. Individual population profiles for PMUT1 apo.

	Center (Å) (± 0.3 Å)	FWHM (Å) (± 0.5 Å)	Relative Population (%) ($\pm 5\%$)
Tucked/Curled 1	24.4	2.0	4
Tucked/Curled 2	28.0	1.8	4
Closed	32.9	3.4	33
Semi-Open	38.7	3.6	55
Wide-Open	46.4	2.7	4

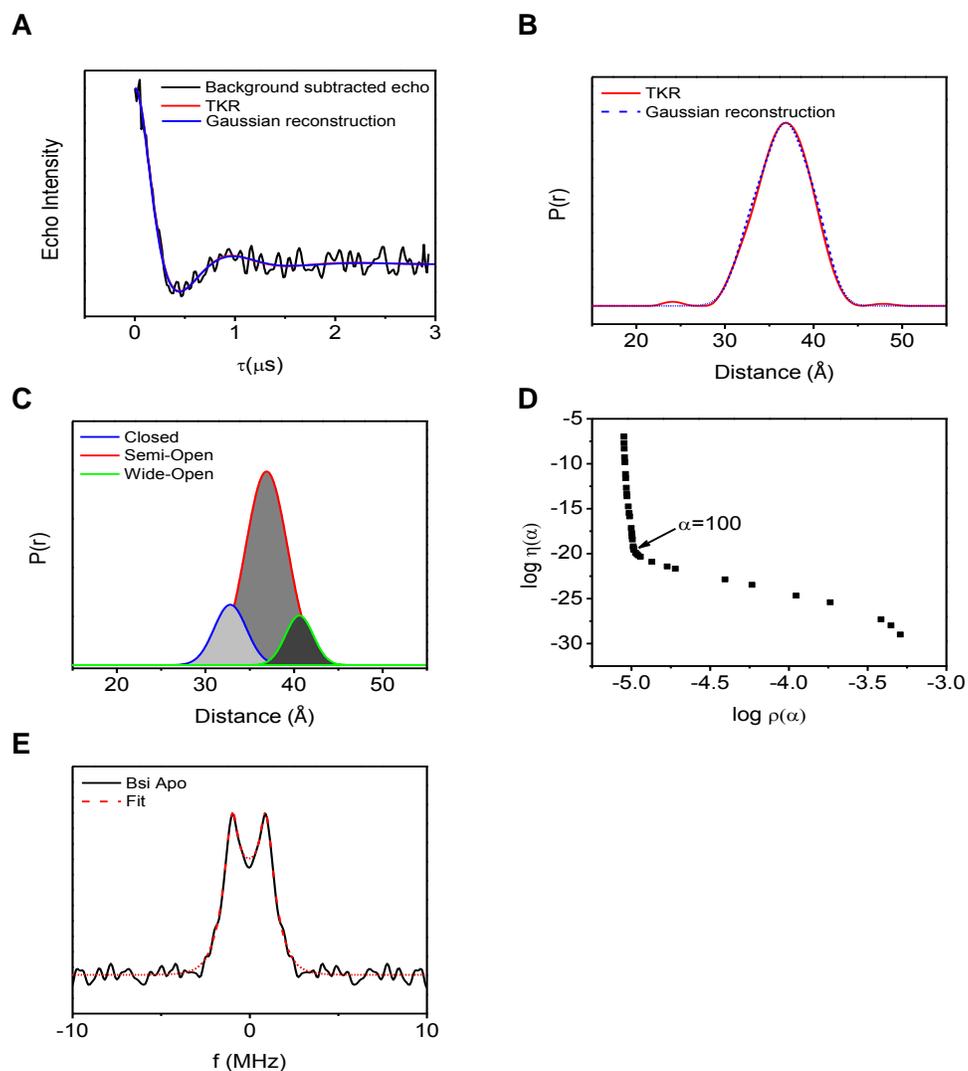


Figure 5-10. DEER data analysis for B_{si} apo. A) Background subtracted dipolar echo curve (black) overlaid with the TKR-generated dipolar modulation (red) and Gaussian reconstructed dipolar modulation (blue), B) distance distribution profile from TKR analysis (red) overlaid with the sum of the populations obtained by Gaussian reconstruction (blue dashed), C) individual Gaussian populations, D) L-curve ($\alpha = 100$), and E) Pake dipolar pattern.

Table 5-6. Individual population profiles for B_{si} apo.

	Center (Å) (± 0.3 Å)	FWHM (Å) (± 0.5 Å)	Relative Population (%) ($\pm 5\%$)
Closed	32.8	4.3	17
Semi-Open	36.9	5.6	71
Wide-Open	41.6	3.7	12

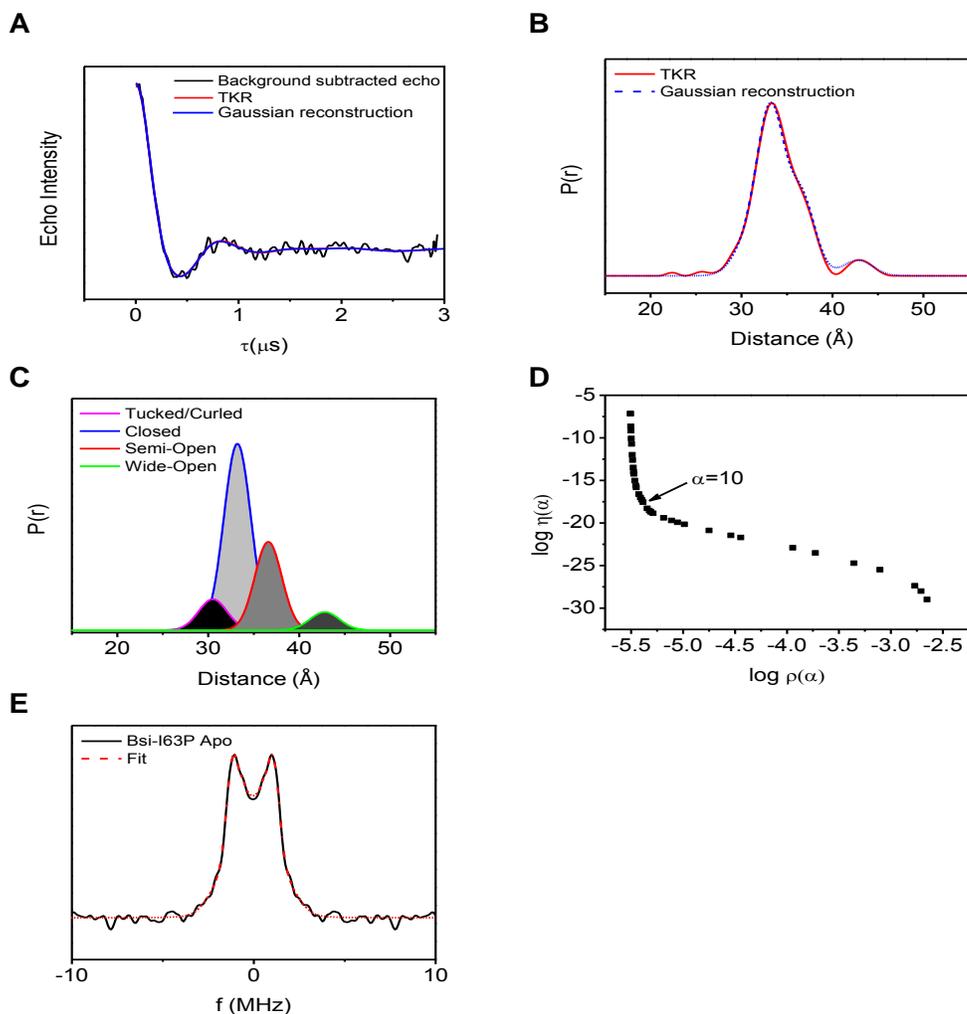


Figure 5-11. DEER data analysis for B_{si}-I63P apo. A) Background subtracted dipolar echo curve (black) overlaid with the TKR-generated dipolar modulation (red) and Gaussian reconstructed dipolar modulation (blue), B) distance distribution profile from TKR analysis (red) overlaid with the sum of the populations obtained by Gaussian reconstruction (blue dashed), C) individual Gaussian populations, D) L-curve ($\alpha = 10$), and E) Pake dipolar pattern.

Table 5-7. Individual population profiles for B_{si}-I63P apo.

	Center (Å) (±0.3 Å)	FWHM (Å) (±0.5 Å)	Relative Population (%) (±5%)
Tucked/Curled	30.5	3.7	10
Closed	33.2	3.5	57
Semi-Open	36.6	3.5	27
Wide-Open	42.8	3.8	6

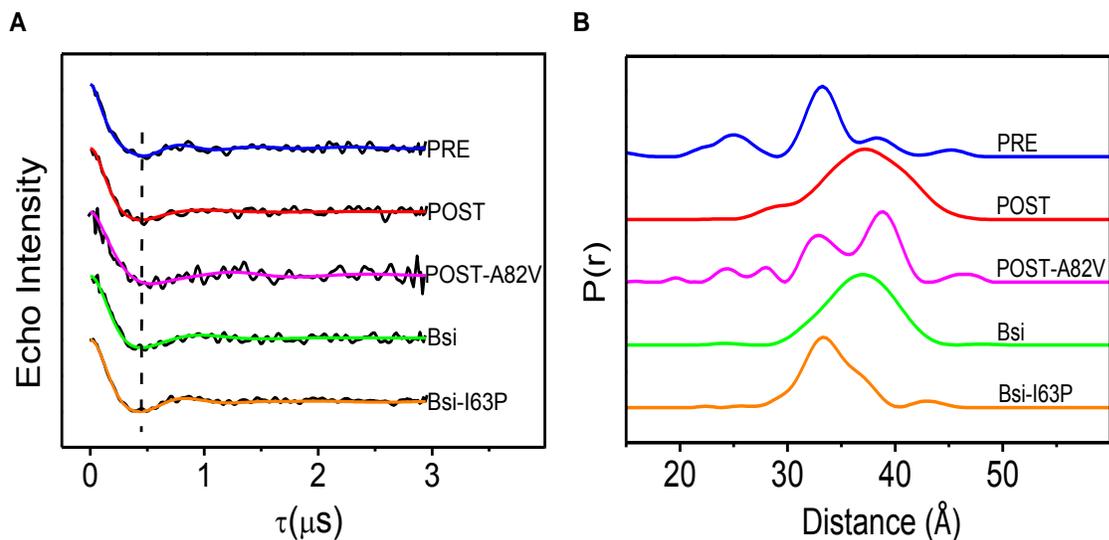


Figure 5-12. (A) Dipolar evolution curves, where background subtracted echo (black) is overlaid with the TKR-generated fit and (B) distance distribution profiles of HIV-1 PR PRE (blue), POST (red), PMUT1 (magenta), B_{si} (green), and B_{si}-I63P (orange) apo constructs.

Table 5-8. Summary of most probable distance of HIV-1 PR apo constructs.

	Most Probable Distance (Å) (± 0.2 Å)
PRE	33.2
POST	37.0
PMUT1	38.5
B _{si}	36.8
B _{si} -I63P	33.2

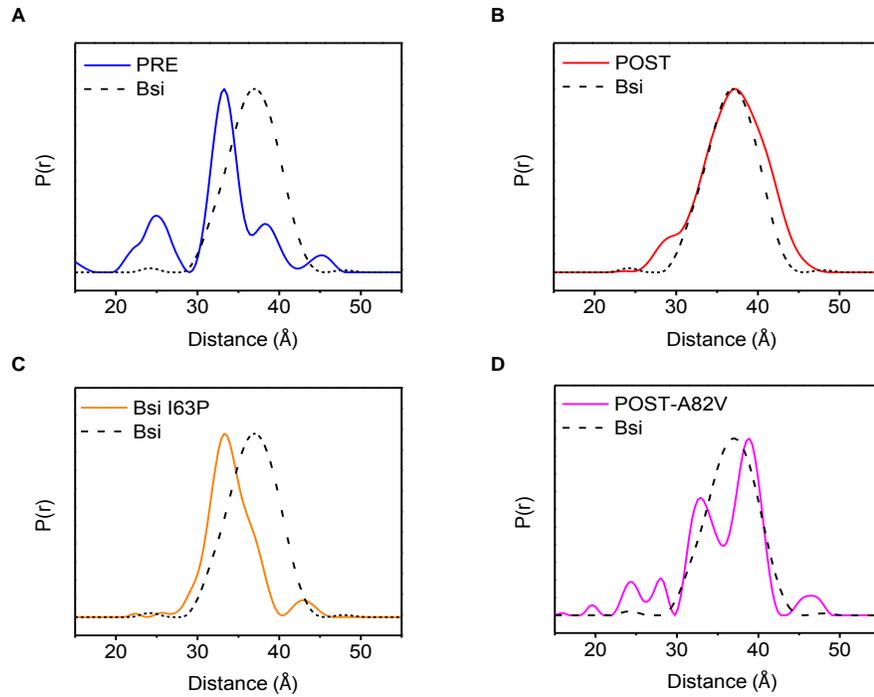


Figure 5-13. Overlay of distance distribution profiles of HIV-1 PR A) PRE (blue), B) POST (red), C) B_{si}-I63P (orange), and D) PMUT1 (magenta) with B_{si} (black dashed) apo constructs.

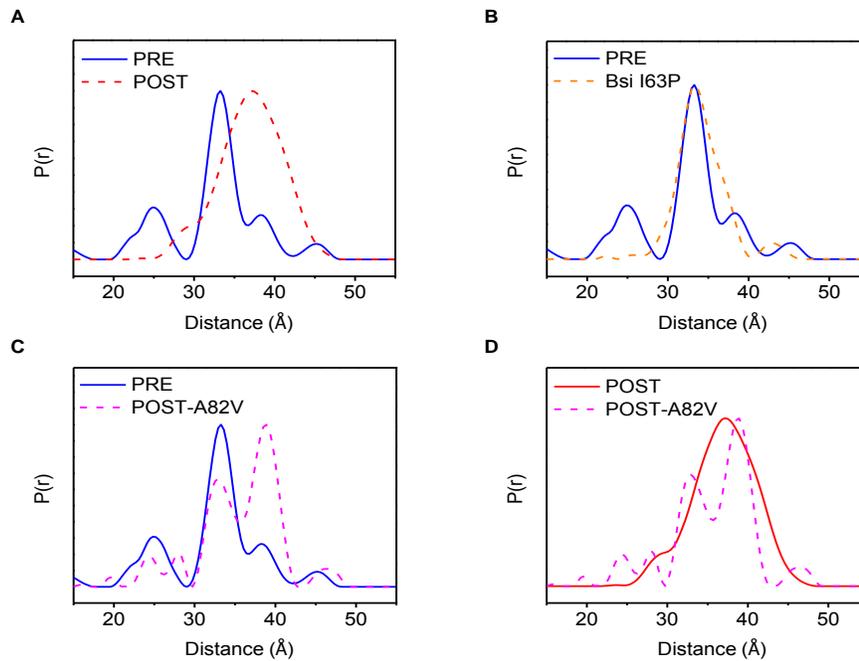


Figure 5-14. Overlay of distance distribution profiles of HIV-1 PR A) PRE (blue) and POST (red), B) PRE (blue) and B_{si}-I63P (orange), C) PRE (blue) and PMUT1 (magenta), and D) POST (red) and PMUT1 (magenta).

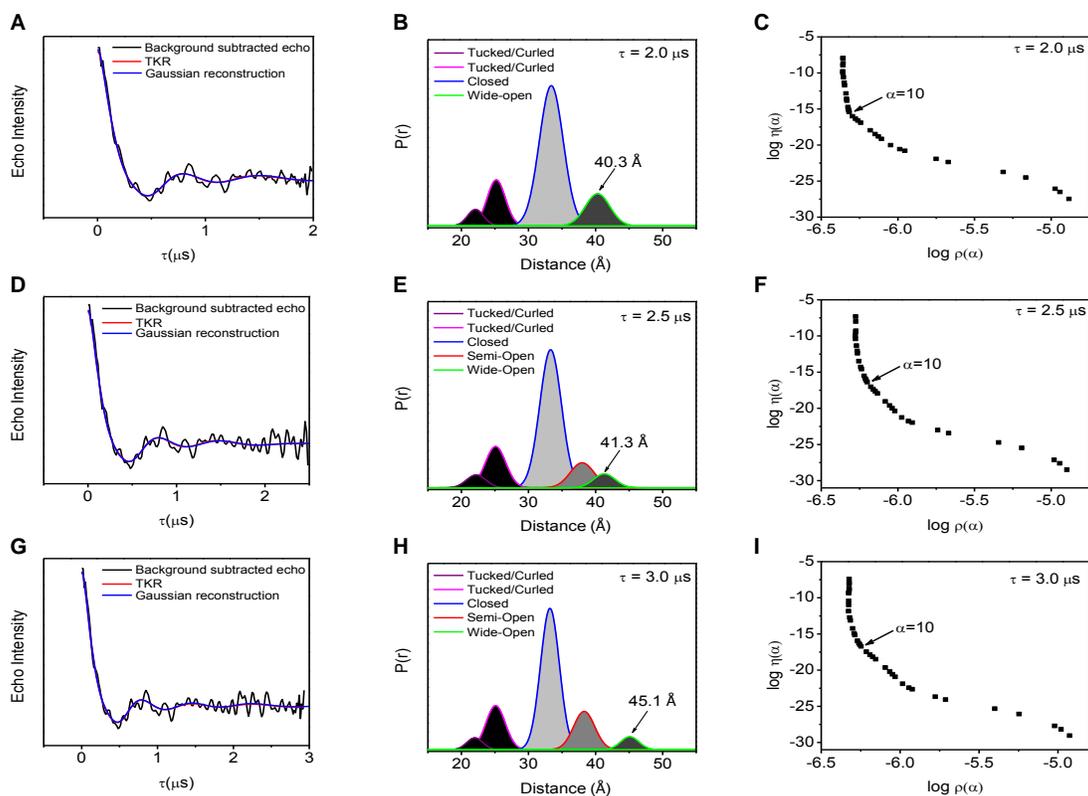


Figure 5-15. Background subtracted dipolar echo curve (black) overlaid with the TKR-generated dipolar modulation (red) and Gaussian reconstructed dipolar modulation (blue), individual Gaussian populations, and L-curves, respectively, for HIV-1 PR PRE apo at A), B), and C) $\tau = 2.0 \mu\text{s}$; D), E), and F) $\tau = 2.5 \mu\text{s}$; and G), H) and I) $\tau = 3.0 \mu\text{s}$.

Table 5-9. Comparison of relative percentages of the individual populations for HIV-1 PR PRE apo at different τ values.

	Distance Assignment (Å)	Relative Population (%) ($\pm 5\%$)		
		$\tau = 2.0 \mu\text{s}$	$\tau = 2.5 \mu\text{s}$	$\tau = 3.0 \mu\text{s}$
Tucked/Curled 1	25-30	5	5	4
Tucked/Curled 2	25-30	16	17	18
Closed	33	64	58	56
Semi-open	36-38	0	15	17
Wide-open	40-46	15	5	5

Table 5-10. Comparison of the relative percentages of the individual populations for HIV-1 PR apo constructs.

	Distance Assignment (Å)	Relative Population (%) ($\pm 5\%$)				
		PRE	POST	PMUT1	B _{si}	B _{si} -I63P
Tucked/Curled	25-30	22	8	8	0	10
Closed	33	56	18	33	17	57
Semi-open	36-38	17	45	55	71	27
Wide-open	40-46	5	29	4	12	6

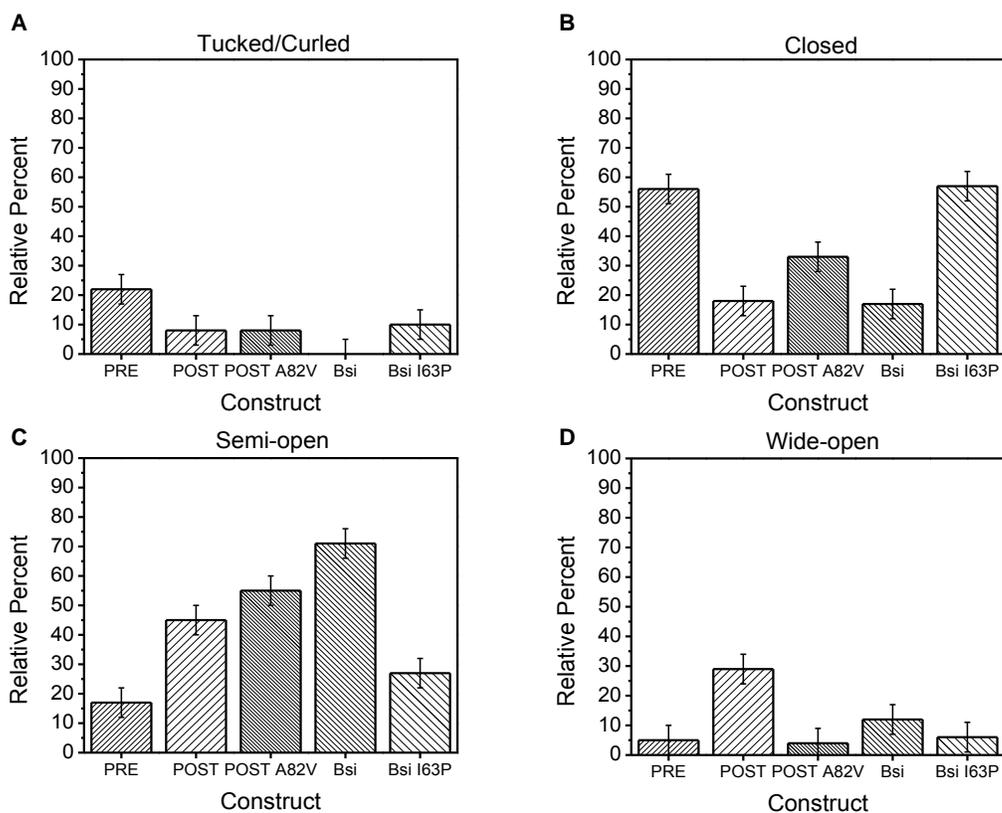


Figure 5-16. Individual plots showing relative percentage of A) tucked/curled, B) closed, C) semi-open and D) wide-open conformations for HIV-1 PR PRE, POST, PMUT1, B_{si}, and B_{si}-I63P apo constructs. Error is $\pm 5\%$.

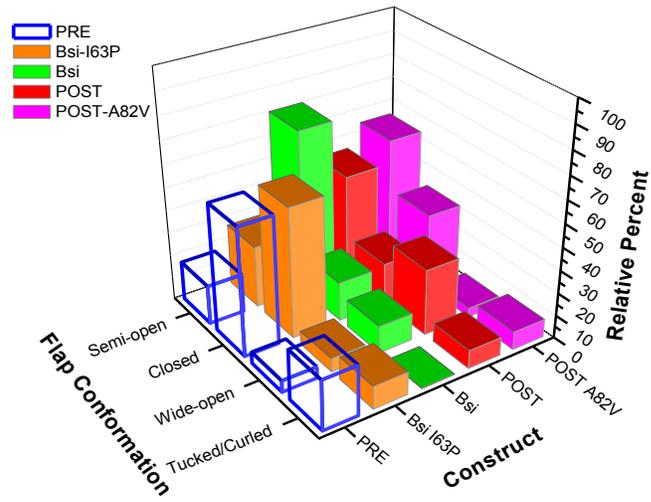


Figure 5-17. Population analysis of HIV-1 PR PRE (blue), POST (red), PMUT1 (magenta), B_{si} (green), and B_{si}-I63P (orange) apo constructs. Error is $\pm 5\%$.

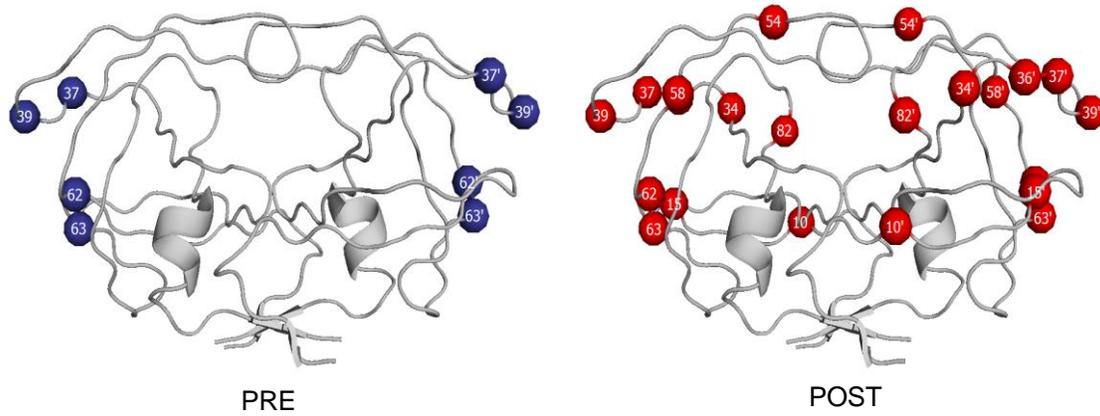


Figure 5-18. Ribbon diagrams of HIV-1 PR PRE and POST in which the amino acid differences relative to subtype B LAI are highlighted and labeled.

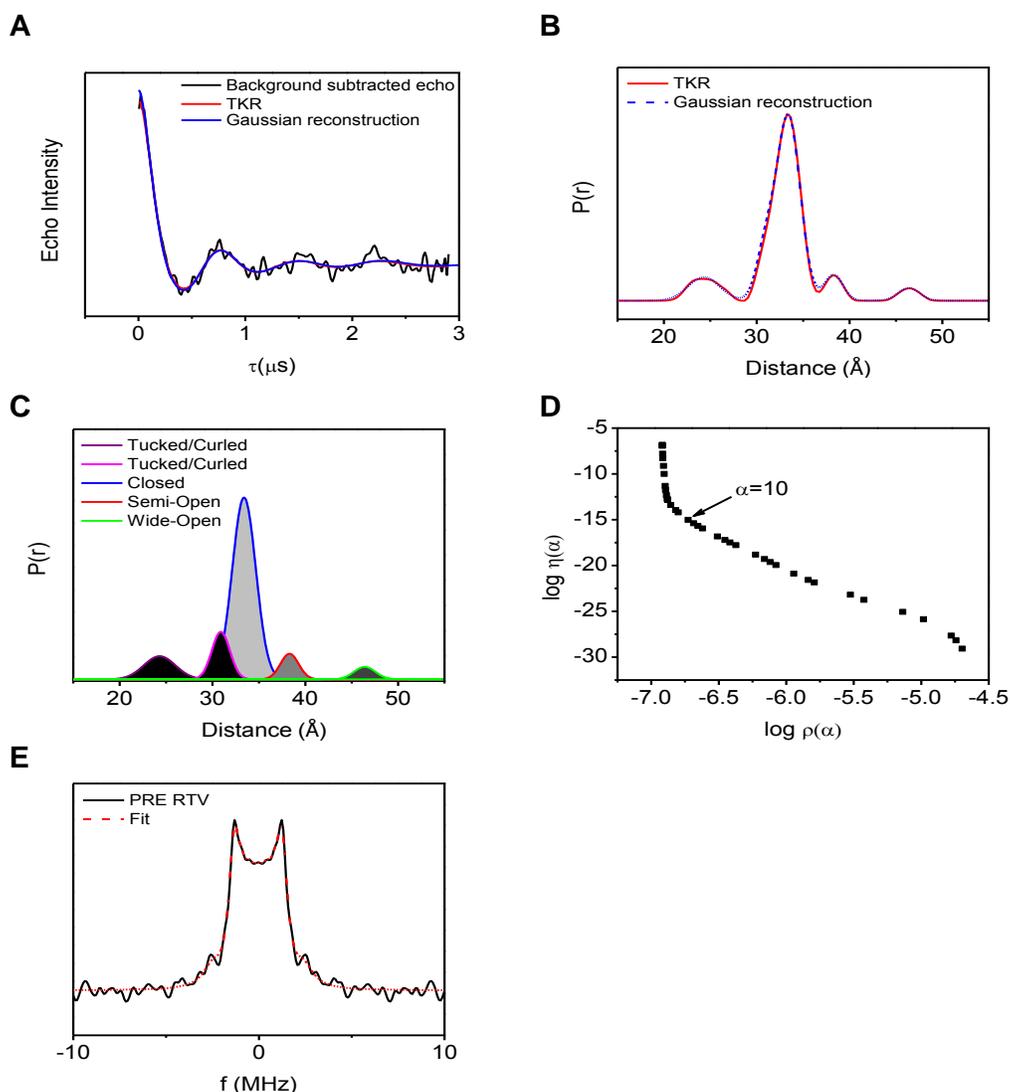


Figure 5-19. DEER data analysis for PRE RTV. A) Background subtracted dipolar echo curve (black) overlaid with the TKR-generated dipolar modulation (red) and Gaussian reconstructed dipolar modulation (blue), B) distance distribution profile from TKR analysis (red) overlaid with the sum of the populations obtained by Gaussian reconstruction (blue dashed), C) individual Gaussian populations, D) L-curve ($\alpha = 10$), and E) Pake dipolar pattern.

Table 5-11. Individual population profiles for PRE RTV.

	Center (Å) (± 0.3 Å)	FWHM (Å) (± 0.5 Å)	Relative Population (%) ($\pm 5\%$)
Tucked/Curled 1	24.3	4.0	11
Tucked/Curled 2	30.9	2.3	13
Closed	33.4	3.0	65
Semi-Open	38.3	2.3	7
Wide-Open	46.4	2.7	4

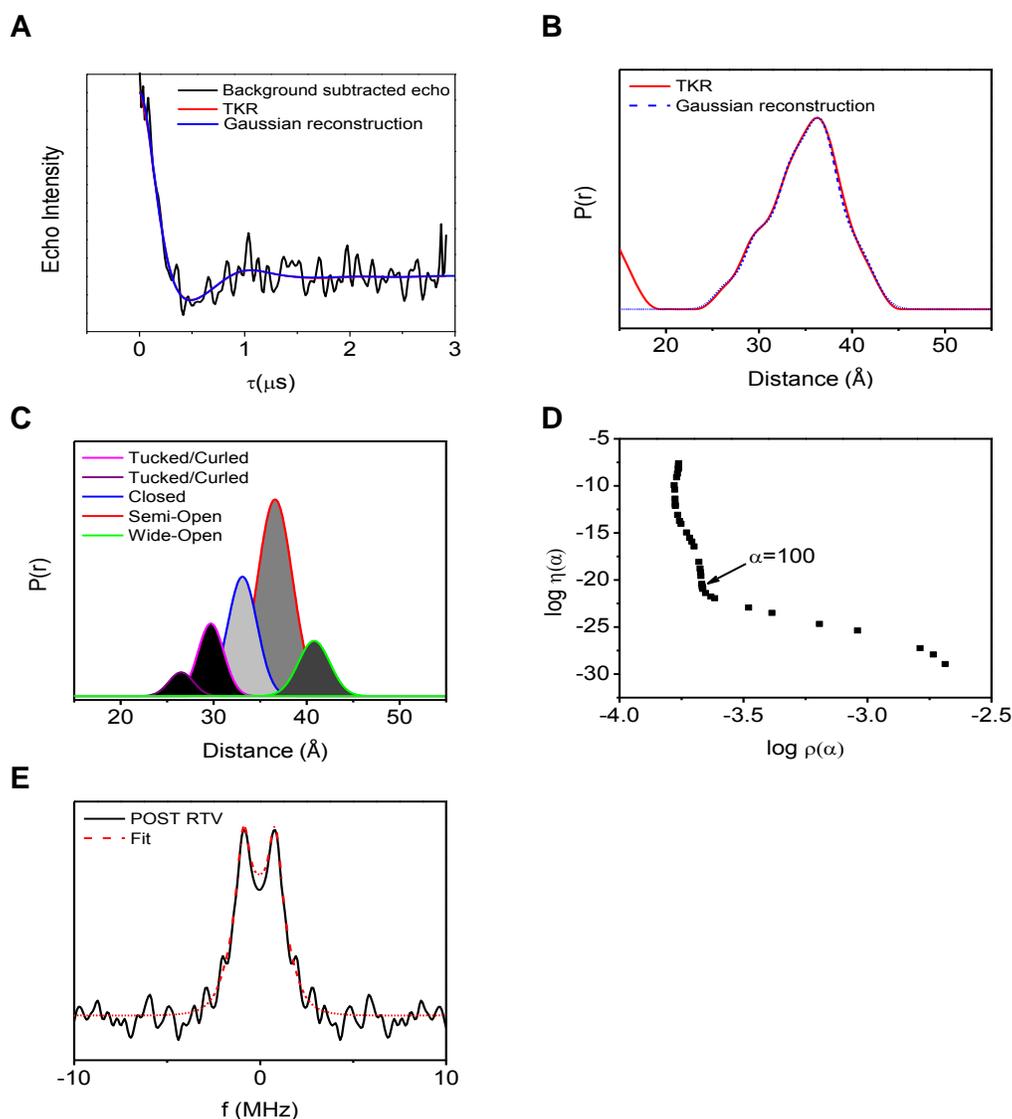


Figure 5-20. DEER data analysis for POST RTV. A) Background subtracted dipolar echo curve (black) overlaid with the TKR-generated dipolar modulation (red) and Gaussian reconstructed dipolar modulation (blue), B) distance distribution profile from TKR analysis (red) overlaid with the sum of the populations obtained by Gaussian reconstruction (blue dashed), C) individual Gaussian populations, D) L-curve ($\alpha = 100$), and E) Pake dipolar pattern.

Table 5-12. Individual population profiles for POST RTV.

	Center (Å) (± 0.3 Å)	FWHM (Å) (± 0.5 Å)	Relative Population (%) ($\pm 5\%$)
Tucked/Curled 1	26.5	3.1	4
Tucked/Curled 2	29.7	3.3	13
Closed	33.1	3.7	24
Semi-Open	36.6	4.4	47
Wide-Open	40.8	4.0	12

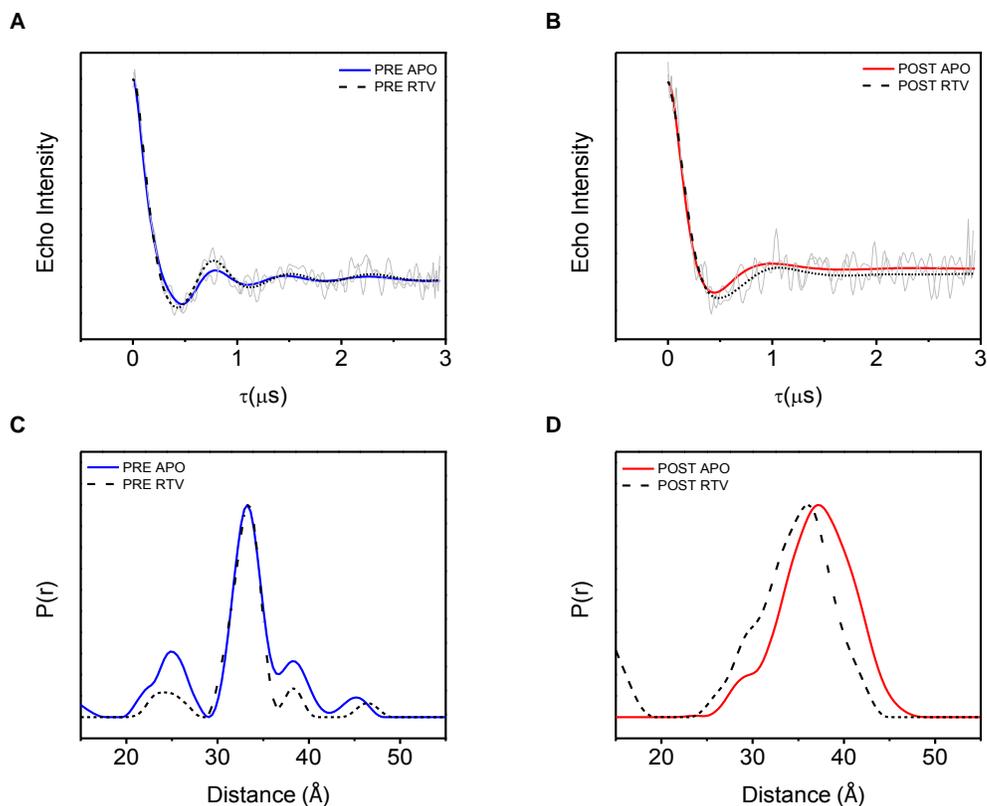


Figure 5-21. Overlay of background subtracted dipolar echo curve (gray) with TKR-generated dipolar evolution curves of A) PRE apo (blue) and PRE RTV (black dashed) and B) POST apo (red) and POST RTV (black dashed) and overlay of distance distribution profiles of C) PRE apo (blue) and PRE RTV (black dashed) and D) POST apo (red) and POST RTV (black dashed).

Table 5-13. Comparison of relative percentages of the individual populations for PRE and POST HIV-1 PR constructs in the absence and presence of RTV.

	Distance Assignment (Å)	Relative Population (%) ($\pm 5\%$)			
		PRE APO	PRE RTV	POST APO	POST RTV
Tucked/Curled	25-30	22	24	8	17
Closed	33	56	65	18	24
Semi-open	36-38	17	7	45	47
Wide-open	40-46	5	4	29	12

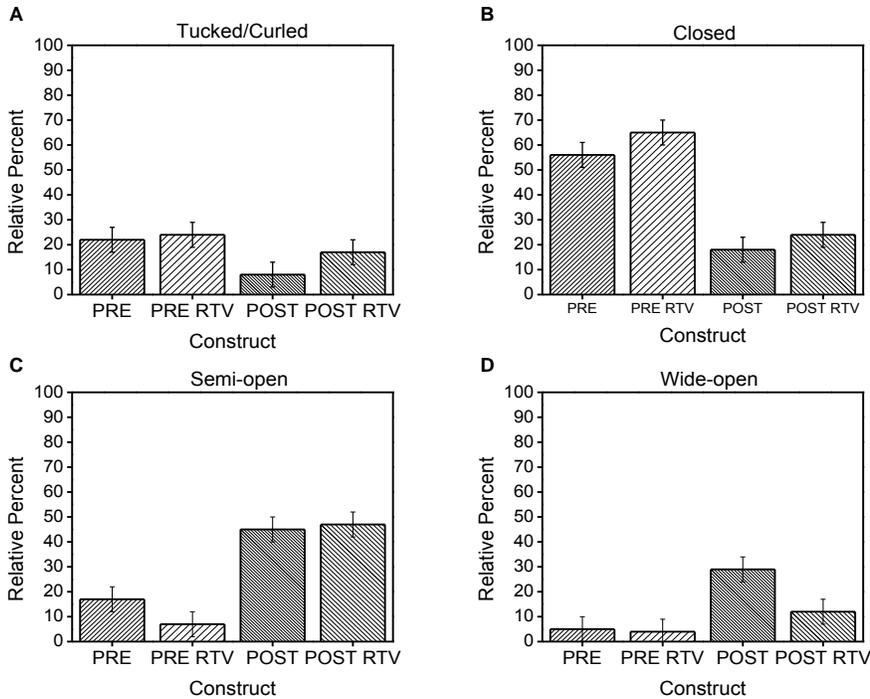


Figure 5-22. Individual plots showing relative percentage of A) tucked/curled, B) closed, C) semi-open and D) wide-open conformations of PRE and POST HIV-1 PR constructs in the absence and presence of RTV. Error is $\pm 5\%$.

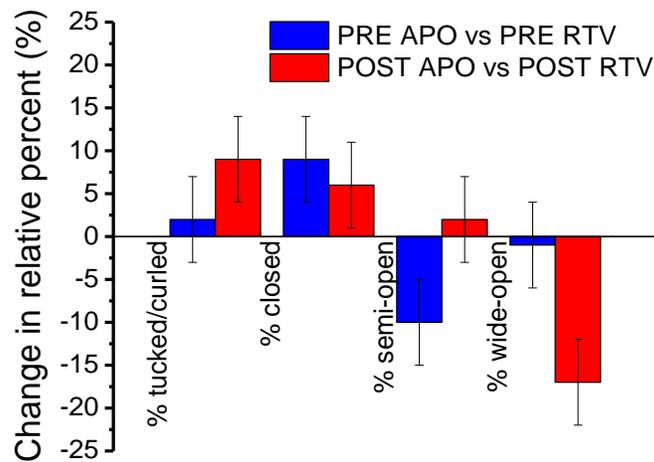


Figure 5-23. Change in relative percent of tucked/curled, closed, semi-open, and wide-open conformations of PRE RTV relative to PRE apo (blue) and POST RTV relative to POST apo (red). Error is $\pm 5\%$.

CHAPTER 6 CONCLUSIONS AND FUTURE WORK

Conclusions

This work investigated different subtype B HIV-1 PR variants by Double Electron-Electron Resonance (DEER) spectroscopy. In particular, DEER results for the pre-therapy (PRE_i) and post-therapy (POST_i) constructs were compared to determine the effect of drug-pressure selected mutations on flap conformation and flexibility. Other mutants were also constructed by reverting sites in POST_i to PRE_i amino acid residues. The PRE_i, POST_i, and POST_i-A82V constructs were also compared to B_{si}, which is the inactive and stabilized form of subtype B LAI, and B_{si}-I63P constructs.

In Chapter 3, the cloning and mutagenesis of all HIV-1 PR constructs were presented. Because the codon-optimized genes, Pre_i-K55C and Post_i-K55C DNA, were purchased in a vector that is not suitable for expression, cloning had to be performed to insert the target genes into the pET23a vector. The results have shown that PRE and POST DNA were successfully cloned into pET23a. Mutagenesis experiments were performed to obtain the following constructs containing drug-pressure selected mutations: 1) pET23_PMUT1, 2) pET23_PMUT3, 3) pET23_PMUT2, and 4) pET23_PMUT5. Based on the DNA sequencing results, all of the mutants were successfully constructed. The success of cloning and mutagenesis experiments was important as the DNA obtained was used for expression of HIV-1 PR in BL21(DE3)pLysS *E. coli* cells.

Results for the purification and characterization of HIV-1 PR were presented in Chapter 4. The following purified and spin labeled inactive HIV-1 PR constructs: 1) PRE, 2) POST, 3) PMUT1, 4) PMUT2, 5) PMUT3, 6) PMUT5, 7) B_{si}, and 8) B_{si}-I63P

(See Table 3-3 for the naming convention) were characterized by SDS-PAGE, CD spectroscopy, MS, and CW-EPR for subsequent analysis by DEER spectroscopy. SDS-PAGE analysis was done on samples to determine the purity of HIV-1 PR. The HIV-1 PR constructs were determined to be >95% pure after the refolding and buffer exchange process. CD spectroscopy was performed to verify the secondary structure of the HIV-1 PR constructs. Results showed that the purified and spin labeled proteins were correctly folded. MS analysis was carried out to confirm the MW of the spin labeled HIV-1 PR and to check for presence of any unlabeled protein. Results showed that some of the HIV-1 PR constructs, particularly PMUT3, PMUT5, and B_{si}, were less than 80% labeled and may not be suitable for DEER analysis. Based on the CW-EPR spectra, all HIV-1 PR constructs have good SNR. However, determination of the spin label concentrations for the proteins revealed that PMUT3, PMUT5, and B_{si} may not be efficiently labeled, which is consistent with MS results.

Results for the DEER analysis of the PRE, POST, PMUT1, B_{si}, and B_{si}-I63P apo constructs were presented in Chapter 5. By detecting shifts in the inter-flap distances, DEER spectroscopy is useful in determining changes in the conformational ensemble of HIV-1 PR. DEER results have shown variations in the population of conformational states of different subtypes and variants of HIV-1 PR. However, the accuracy of the assignment of populations is limited by the quality of the data. The signal-to-noise ratio (SNR) has to be sufficiently high to accurately determine lowly populated conformational states. Nevertheless, the technique is useful in comparing the predominant conformations of different apo HIV-1 PR constructs. PRE and B_{si}-I63P have predominantly closed conformations, indicating a limited flexibility in the range of

HIV-1 PR flap motion, whereas POST, POST-A82V, and B_{si} have the expected predominant semi-open conformation. DEER results also showed that addition of ritonavir (RTV) to the PRE and POST constructs did not significantly shift the conformational ensemble of HIV-1 PR. The results demonstrate that certain amino acids in the PRE construct, particularly at site 63, and the combination of mutations in the POST construct, have a major impact on the flap conformation and flexibility.

Future Work

DEER analysis needs to be performed on the purified and spin labeled PRE, POST, and PMUT1 samples, which were reported to have good spin labeling efficiency in Chapter 4, to obtain DEER data with high SNR. Data needs to be recollected for these constructs in order to verify DEER results presented in this work and to determine the reproducibility of analyzing efficiently spin labeled samples and DEER data with good SNR. DEER analysis on the PRE and POST constructs in the presence of RTV also needs to be repeated because of low SNR. The other POST mutants constructed in Chapters 3 and 4 will also be analyzed. Proper folding of constructs needs to be verified by further CD analysis to compare the % β -sheet structure of the constructs containing mutations with PRE, POST, and B_{si}. Additional DEER studies will be performed by determining the effects of substrates (CA/p2 and p2/p7) and FDA-approved inhibitors on the flap conformations of all HIV-1 PR constructs in this work. For constructs in the presence of inhibitor, inhibitor binding needs to be verified using other techniques such as ITC and NMR titration experiments.

APPENDIX A
PRIMER CHARACTERISTICS AND PCR PARAMETERS FOR MUTAGENESIS
EXPERIMENTS

Table A-1. Primers used in incorporating mutations into pET23_ POST to produce the pET23_ PMUT1 construct.

Mutation	Primer 5'-3'	T _m (°C)	MW	GC content (%)
A82V				
Forward	GTGGGTCCAAC TCCGGTAAACATTATCGGCCG	66.1	9,825.4	56.2
Reverse	CGGCCGATAATGTTTACCGGAGTTGGACCCAC	66.1	9,825.4	56.2

Table A-2. Primers used in incorporating mutations into pET23_ POST to produce the pET23_ PMUT3 construct.

Mutation	Primer 5'-3'	T _m (°C)	MW	GC content (%)
A82V				
Forward	GTGGGTCCAAC TCCGGTAAACATTATCGGCCG	66.1	9,825.4	56.2
Reverse	CGGCCGATAATGTTTACCGGAGTTGGACCCAC	66.1	9,825.4	56.2
I10L				
Forward	GCTGTGGCAACGTCCTGTCACCATCAAAG	67.0	9,785.4	56.2
Reverse	CTTTGATGGTGACAAGCGGACGTTGCCACAGC	67.0	9,865.4	56.2
V15I				
Forward	CGCTTGTCACCATCAAATTTGGCGGTCAGCTGAAAG	66.1	11,069.2	50.0
Reverse	CTTTCAGCTGACCCCAATTTTGTGATGGTGACAAGCG	66.1	11,051.2	50.0

Table A-3. Primers used in incorporating mutations into pET23_ POST to produce the pET23_ PMUT2 construct.

Mutation	Primer 5'-3'	T _m (°C)	MW	GC content (%)
A54I				
Forward	GGCATCGGCGGTTTTATCTGCGTCCGTGAGTAC	67.2	10,167.6	57.5
Reverse	GTACTCACGGACGCAGATAAAAACCGCCGATGCC	67.2	10,101.6	57.5
E58Q				
Forward	GGTTTTATCTGCGTCCGTGTCAGTACGACCAGGTGCCG	68.4	11,059.2	58.3
Reverse	CGGCACCTGGTCGTACTGACGGACGCAGATAAAAACC	68.4	11,064.2	58.3

Table A-4. Primers used in incorporating mutations into pET23_ POST to produce the pET23_ PMUT5 construct.

Mutation	Primer 5'-3'	T _m (°C)	MW	GC content (%)
A54I				
Forward	GGCATCGGCGGTTTTATCTGCGTCCGTGAGTAC	67.2	10,167.6	57.5
Reverse	GTACTCACGGACGCAGATAAAAACCGCCGATGCC	67.2	10,101.6	57.5
E58Q				
Forward	GGTTTTATCTGCGTCCGTGAGTACGACCAGGTGCCG	68.4	11,059.2	58.3
Reverse	CGGCACCTGGTCTGACTGACGGACGCAGATAAAAACC	68.4	11,064.2	58.3
Q34E				
Forward	CGGACGATACCGTTCTGGAGGAAATCAACTTGAC	63.8	10,475.8	50.0
Reverse	GTCAAGTTGATTTCTCCAGAACGGTATCGTCCG	63.8	10,408.8	50.0
I36M				
Forward	GTTCTGGAGGAAATGAACTTGACGGGTGCGC	63.9	9,327.1	53.3
Reverse	GCGACCCGTCAAGTTCATTTCTCCAGAAC	63.9	9,086.9	53.3
N37T				
Forward	GTTCTGGAGGAAATGACCTTGACGGGTGCGCTG	66.0	9,936.5	56.2
Reverse	CAGCGACCCGTCAAGTTCATTTCTCCAGAAC	66.0	9,714.3	56.2

Table A-5. Primers used in incorporating mutations into pET23_ B_{si} to produce the pET23_ B_{si}-I63P construct.

Mutation	Primer 5'-3'	T _m (°C)	MW	GC content (%)
I63P				
Forward	CAGTACGACCAGATCCCTATCGAAATCGCCGGC	66.5	10,052.6	57.5
Reverse	GCCGGCGATTTTCGATAGGGATCTGGTCTGACTG	66.5	10,216.6	57.5

Table A-6. Components of the PCR mixture and their corresponding volume for site-directed mutagenesis.

Order of Addition	Component	Volume PCR Mixture 1 (μL)	Volume PCR Mixture 2 (μL)	Volume Control (μL)
1	Water	34.5	38	38
2	DMSO	2.5	0	0
3	dNTP Mix (10mM)	1	1	1
4	Primers			
	Forward	2	1.5	2
	Reverse	2	1.5	2
5	Template DNA	2	2	2
6	<i>Pfu</i> buffer	5	5	5
7	<i>Pfu</i> polymerase	1	1	0
	Total Volume	50	50	50

Table A-7. Thermal cycling parameters for HIV-1 protease site-directed mutagenesis reactions.

Segment	Cycles	Temperature (°C)	Time
1	1	95	30 sec
2	18	95	30 sec
		55	1 min
		68	6 min

APPENDIX B
ADDITIONAL INFORMATION ON THE PURIFICATION AND CHARACTERIZATION
OF HIV-1 PR CONSTRUCTS

Table B-1. Buffers used in the purification of HIV-1 PR.

Buffer	pH	Component
Resuspension Buffer (Stored at 25°C)	7.5	20 mM Tris-HCl 1 mM EDTA ^a 10 μM BME ^b
Wash Buffer 1 (Stored at 4°C)	7.0	25 mM Tris-HCl 2.5 mM EDTA 0.5 M NaCl 1mM Gly-Gly 50 μM BME
Wash Buffer 2 (Prepared fresh)	7.0	25 mM Tris-HCl 2.5 mM EDTA 0.5 M NaCl 1mM Gly-Gly 50 μM BME 1M Urea
Wash Buffer 3 (Stored at 4°C)	7.0	25 mM Tris-HCl 1.0 mM EDTA 0.5 M NaCl 1mM Gly-Gly 50 μM BME
Inclusion Body (IB) Resuspension Buffer (Prepared fresh)	Adjust depending on the pI of protein (not >0.5 lower than pI)	25 mM Tris-HCl 2.5 mM EDTA 0.5 M NaCl 1mM Gly-Gly 50 μM BME 9M Urea

^aEDTA: Ethylenediaminetetraacetic acid; ^bBME: β-mercapthoethanol

Table B-2. Optical density at 280 nm and estimated protein concentration of HIV-1 PR samples used for CD data collection.

HIV-1 PR Construct	OD ₂₈₀	Estimated Protein Concentration (μM)
PRE	0.251	20.1
POST	0.332	26.6
PMUT1	0.262	21.0
PMUT3	0.325	26.0
PMUT2	0.266	21.3
PMUT5	0.308	24.7
B _{si}	0.256	20.5
B _{si} -I63P	0.253	20.3

Table B-3. Optical density at 280 nm and estimated concentration of HIV-1 PR samples used for EPR data collection and concentration of spin label for the HIV-1 PR monomer determined from CW-EPR.

HIV-1 PR Construct	OD ₂₈₀	Estimated Protein Concentration (μM)	Estimated Concentration of MTSL (μM)
PRE	1.136	91.0	18.1
POST	1.364	109.2	23.2
PMUT1	1.305	104.5	24.2
PMUT3	1.423	113.9	13.9
PMUT2	1.207	96.6	19.7
PMUT5	1.701	136.2	19.6
B _{si}	1.717	137.5	19.6
B _{si} -I63P	1.025	82.1	16.9

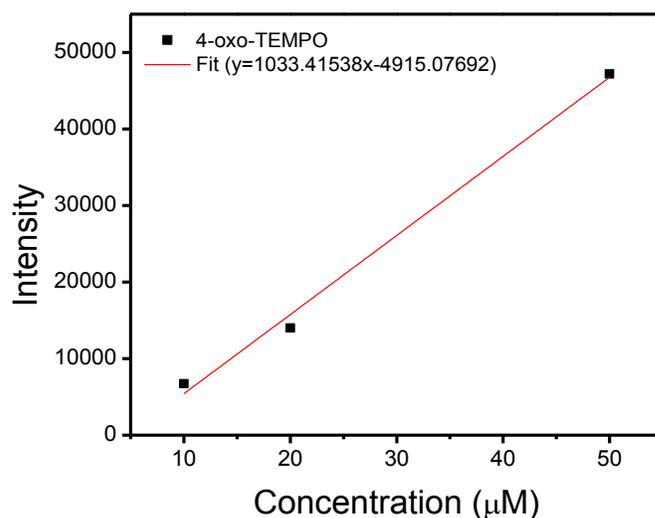


Figure B-1. Calibration curve of 4-oxo-TEMPO standard determined by CW-EPR. EPR intensities of three calibration points (10 μM, 20 μM and 50 μM) were measured. The linear fit equation as indicated was used to estimate the spin label concentration of HIV-1 PR samples.

The DEER results presented in Chapter 5 correspond to older samples of the PRE, POST and PMUT1 HIV-1 PR constructs. The spin labeling efficiency of these samples was verified by MS and CW-EPR and results are presented here (Figures B-2 to B-5). To verify the certainty of the DEER results obtained, the three constructs were repurified and spin-labeled and results for the characterization are presented in Chapter 4. However, DEER analysis was not performed on the repurified PRE, POST and PMUT1 samples because of technical problems associated with the DEER instrument.

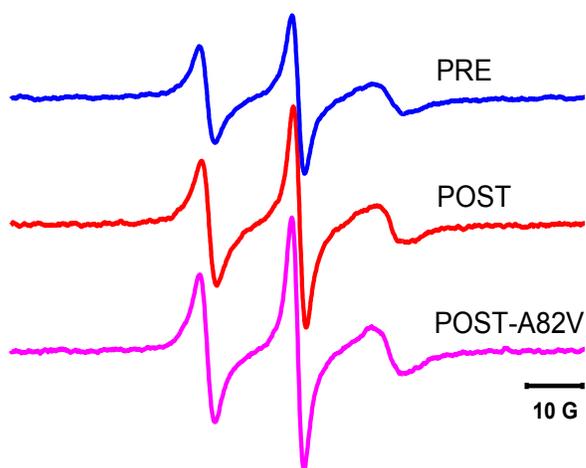


Figure B-2. CW-EPR spectra for spin labeled HIV-1 PR PRE (blue), POST (red), and PMUT1 (magenta).

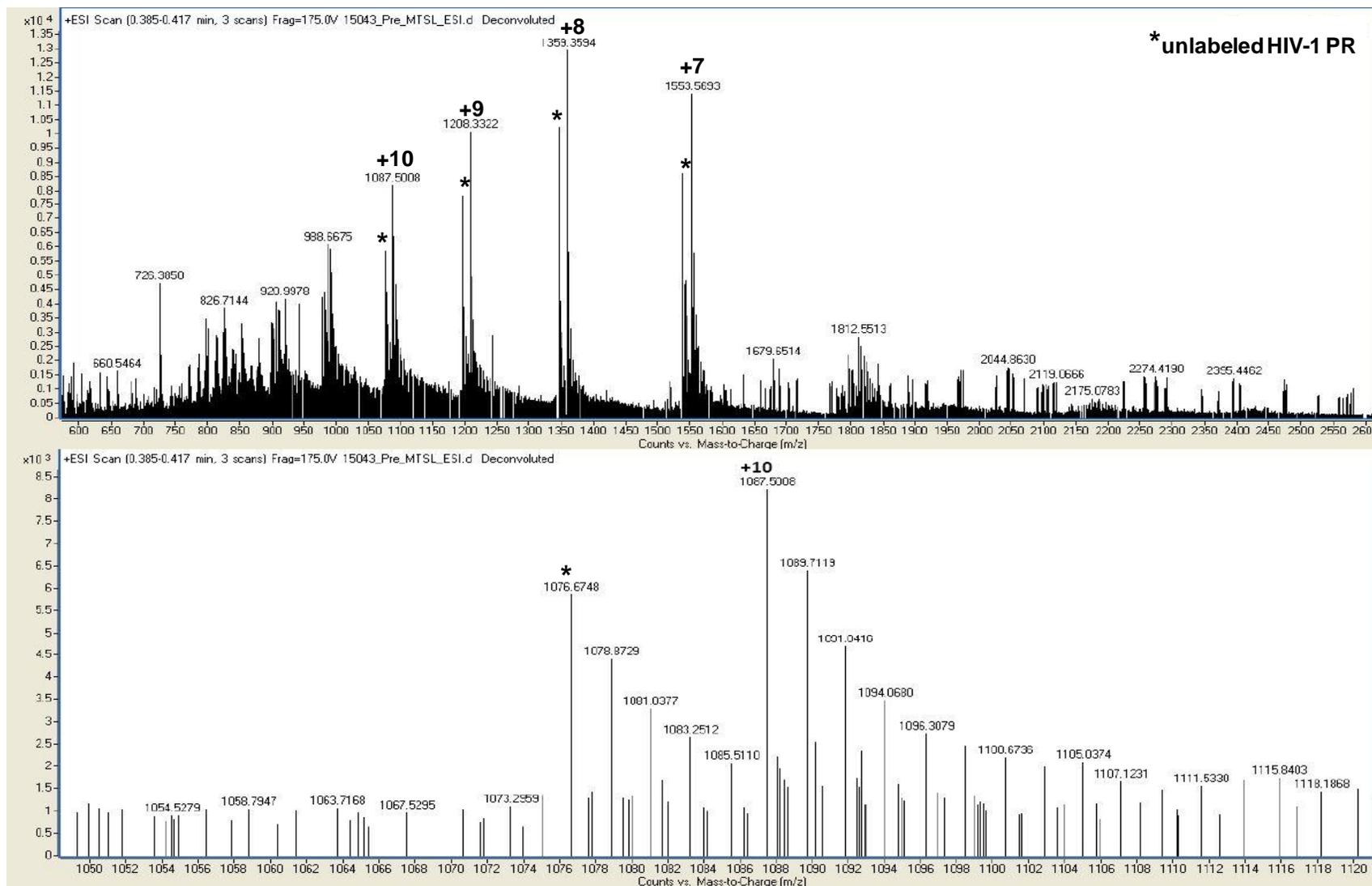


Figure B-3. Mass spectra of spin labeled PRE. The upper spectrum shows the major peaks with the charge states indicated. The lower spectrum is an enlarged view of the m/z peak with +10 charge state.

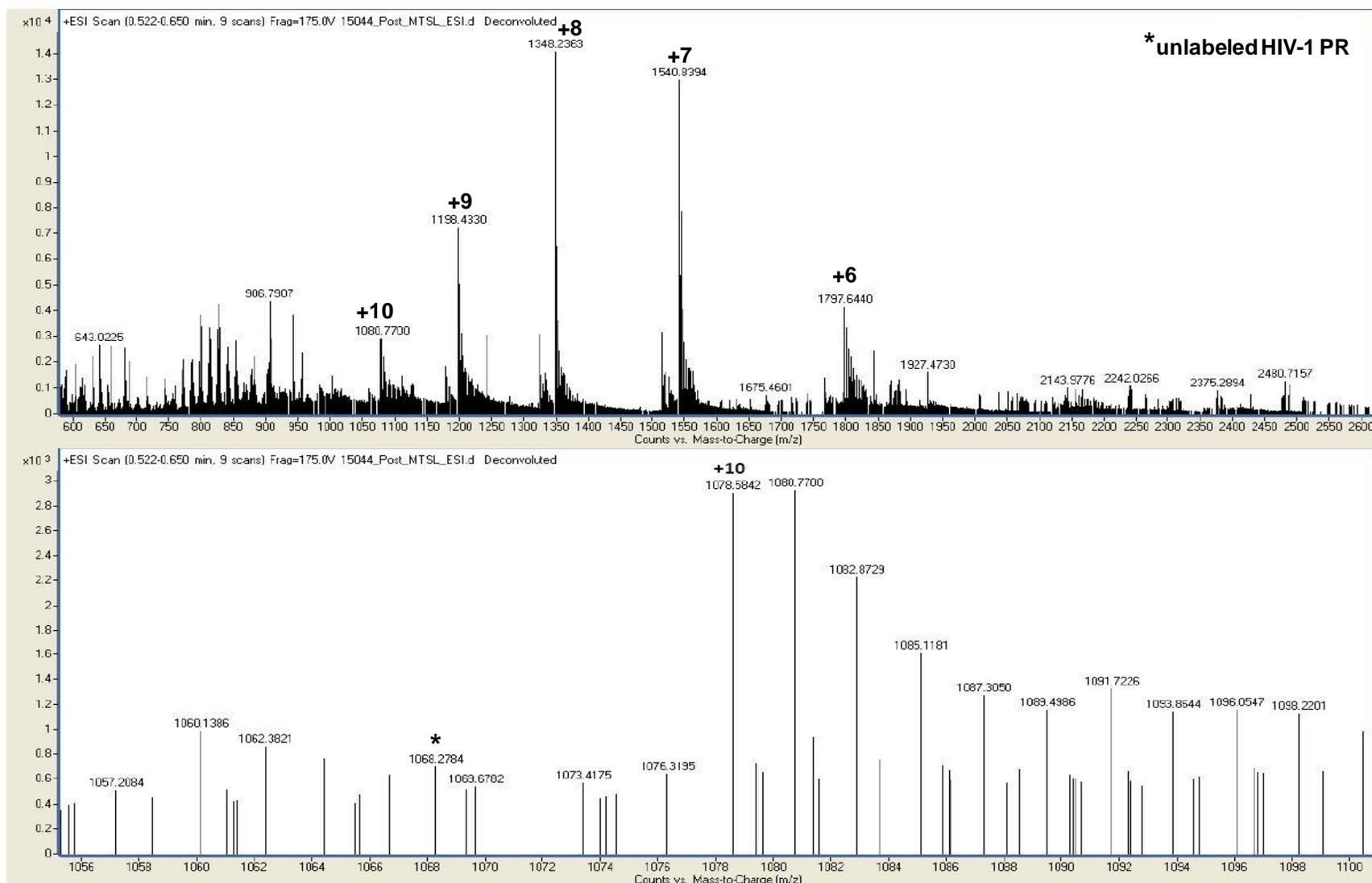


Figure B-4. Mass spectra of spin labeled POST. The upper spectrum shows the major peaks with the charge states indicated. The lower spectrum is an enlarged view of the m/z peak with +10 charge state.

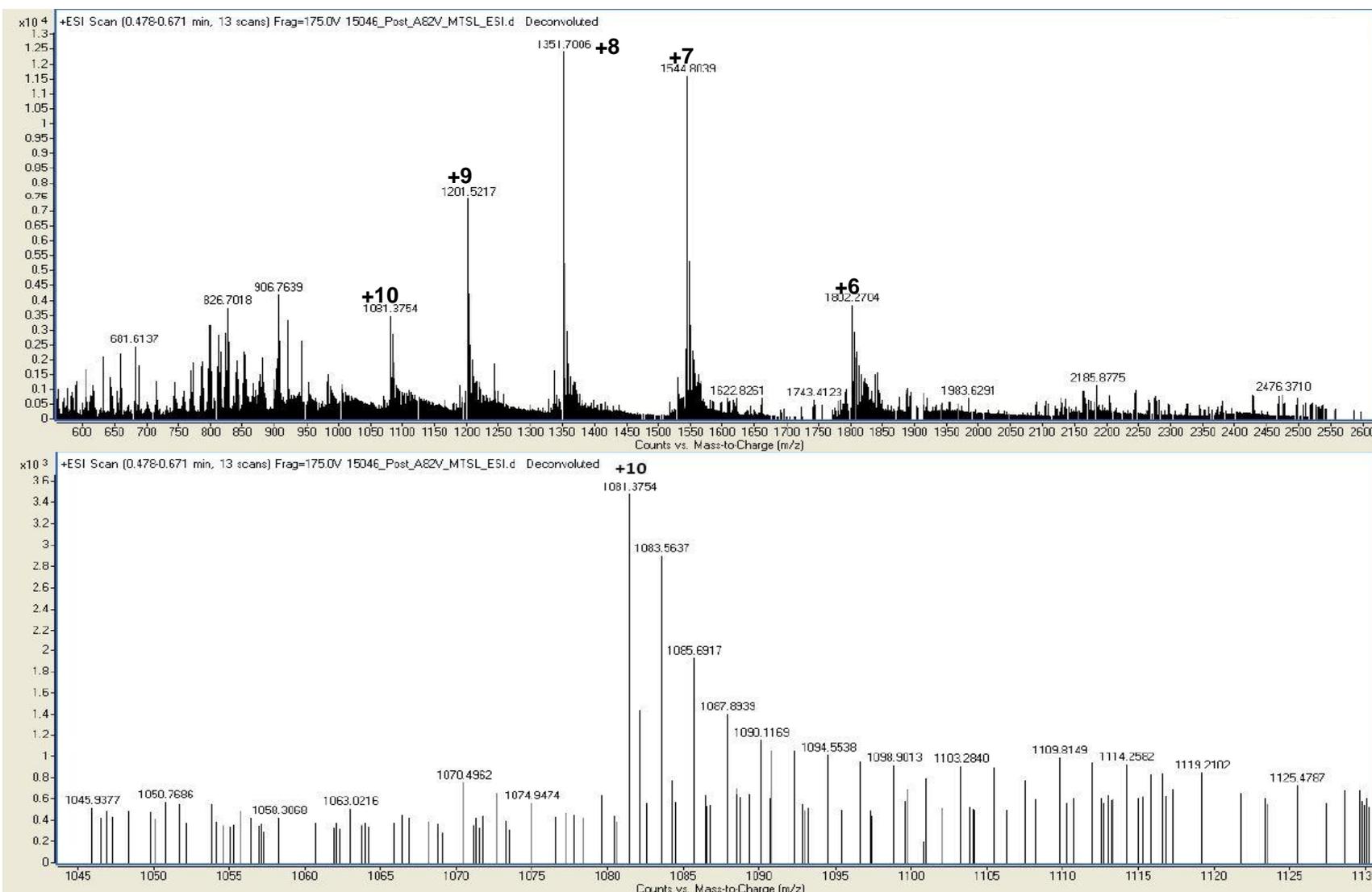


Figure B-5. Mass spectra of spin labeled PMUT1. The upper spectrum shows the major peaks with the charge states indicated. The lower spectrum is an enlarged view of the m/z peak with +10 charge state.

APPENDIX C
POPULATION VALIDATION OF SELECTED HIV-1 PR CONSTRUCTS

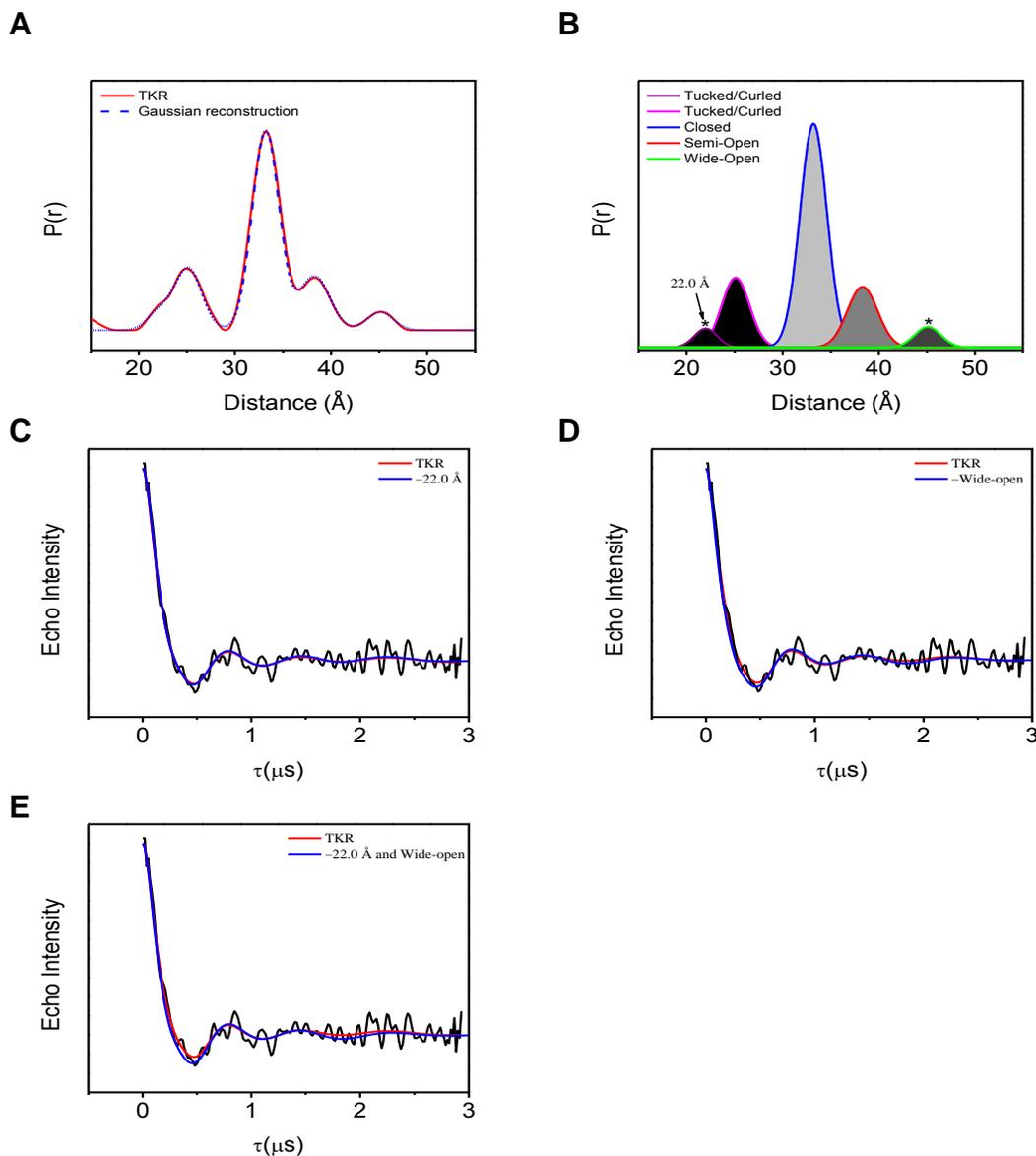


Figure C-1. Population validation of PRE apo. A) Distance distribution profile from TKR analysis (red) overlaid with the sum of the populations obtained by Gaussian reconstruction (blue dashed), B) individual Gaussian populations (suppressed populations are indicated by an asterisk), and C-E) background subtracted dipolar echo curve (black) overlaid with the TKR-generated dipolar modulation (red) and Gaussian reconstructed dipolar modulation generated by subtracting the suppressed populations (blue).

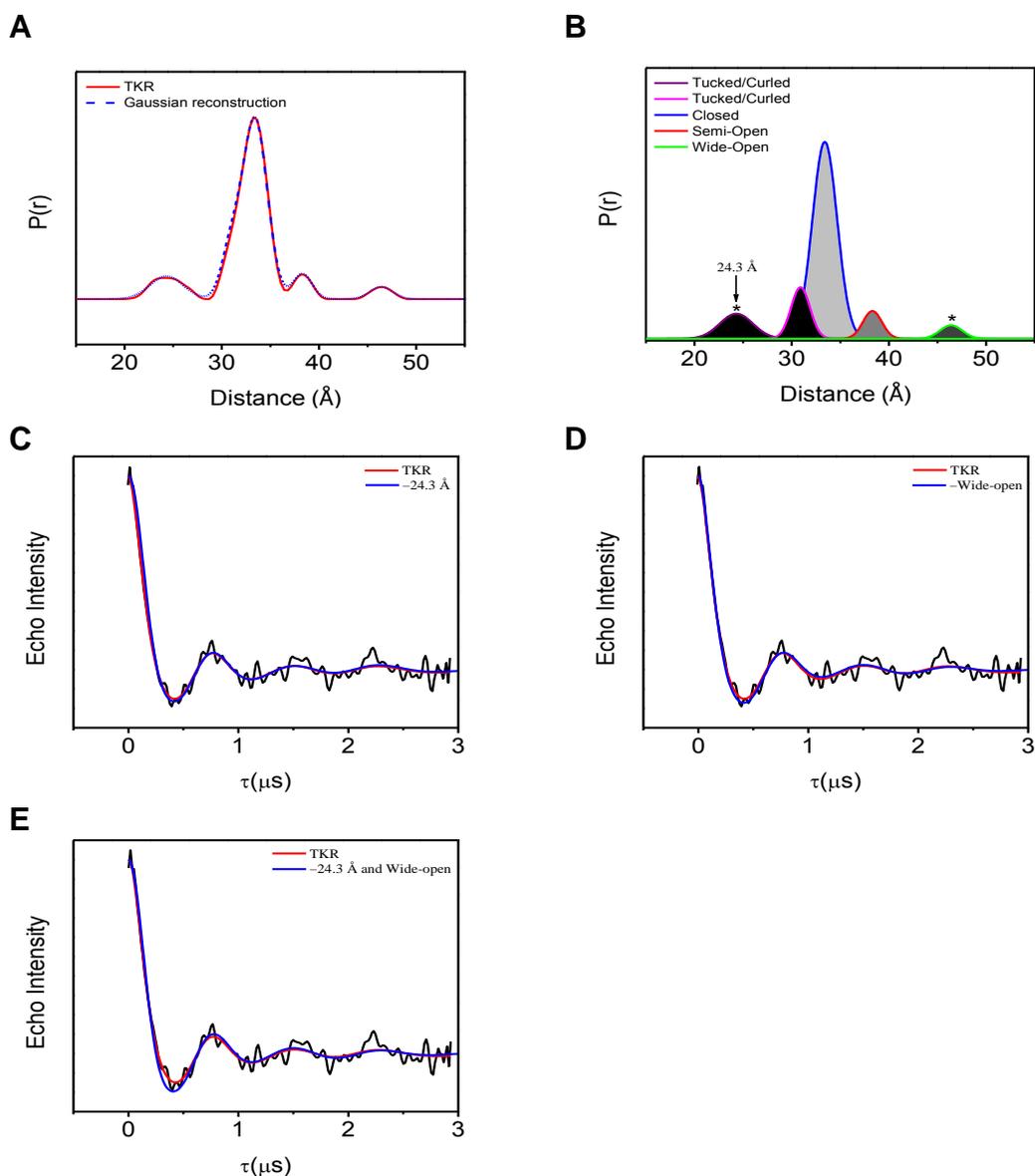


Figure C-2. Population validation of PRE RTV. A) Distance distribution profile from TKR analysis (red) overlaid with the sum of the populations obtained by Gaussian reconstruction (blue dashed), B) individual Gaussian populations (suppressed populations are indicated by an asterisk), and C-E) background subtracted dipolar echo curve (black) overlaid with the TKR-generated dipolar modulation (red) and Gaussian reconstructed dipolar modulation generated by subtracting the suppressed populations (blue).

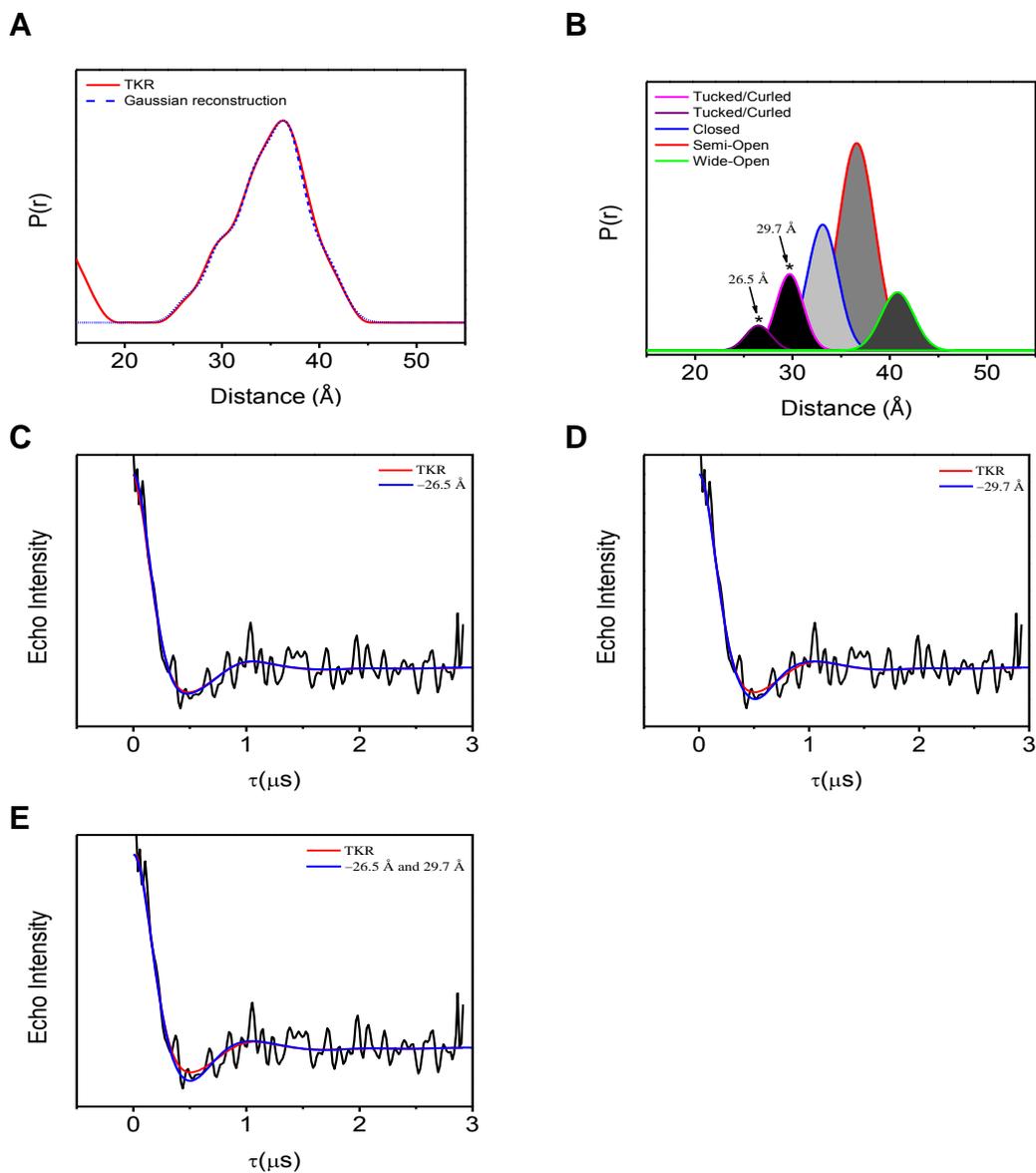


Figure C-3. Population validation of POST RTV. A) Distance distribution profile from TKR analysis (red) overlaid with the sum of the populations obtained by Gaussian reconstruction (blue dashed), B) individual Gaussian populations (suppressed populations are indicated by an asterisk), and C-E) background subtracted dipolar echo curve (black) overlaid with the TKR-generated dipolar modulation (red) and Gaussian reconstructed dipolar modulation generated by subtracting the suppressed populations (blue).

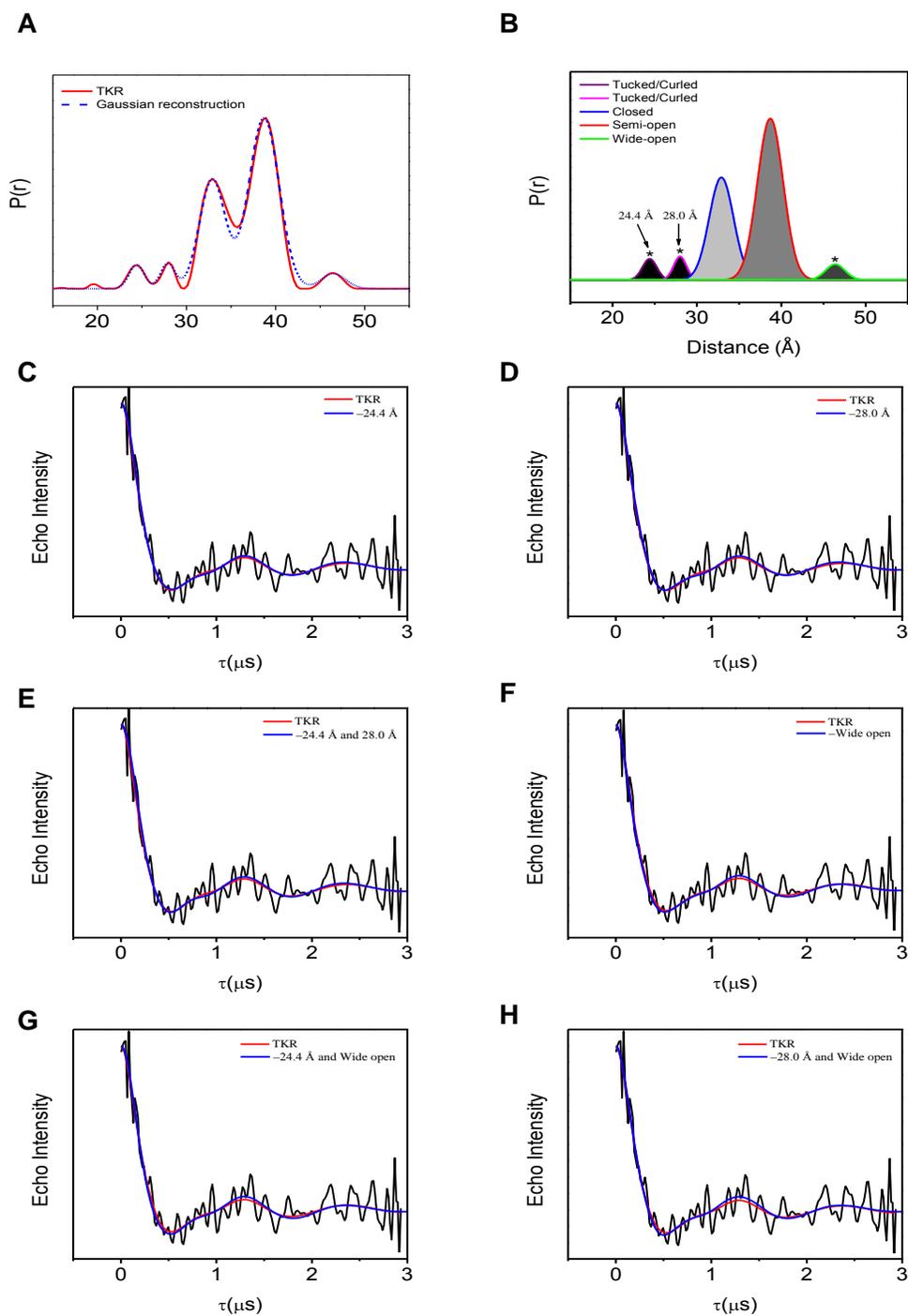


Figure C-4. Population validation of PMUT1 apo. A) Distance distribution profile from TKR analysis (red) overlaid with the sum of the populations obtained by Gaussian reconstruction (blue dashed), B) individual Gaussian populations (suppressed populations are indicated by an asterisk), and C-H) background subtracted dipolar echo curve (black) overlaid with the TKR-generated dipolar modulation (red) and Gaussian reconstructed dipolar modulation generated by subtracting the suppressed populations (blue).

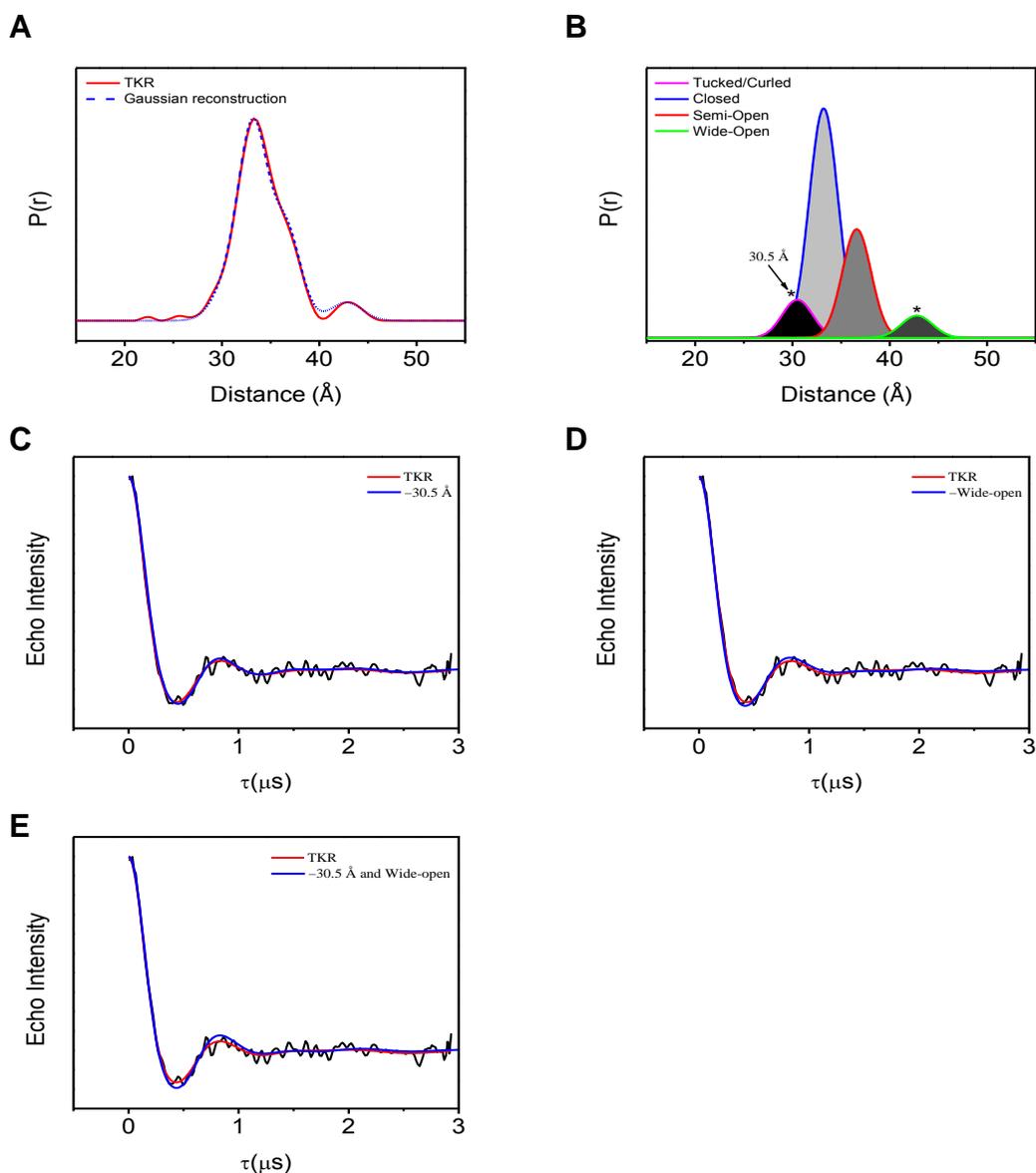


Figure C-5. Population validation of B_{Si}-I63P apo. A) Distance distribution profile from TKR analysis (red) overlaid with the sum of the populations obtained by Gaussian reconstruction (blue dashed), B) individual Gaussian populations (suppressed populations are indicated by an asterisk), and C-E) background subtracted dipolar echo curve (black) overlaid with the TKR-generated dipolar modulation (red) and Gaussian reconstructed dipolar modulation generated by subtracting the suppressed populations (blue).

LIST OF REFERENCES

1. UNAIDS. (2010) 2010 Report on the Global AIDS Epidemic, UNAIDS.
2. Levy, J. A. (1993) Pathogenesis of human immunodeficiency virus infection, *Microbiol. Rev.* 57, 183-289.
3. Andersson, S. (2001) HIV-2 and the immune response, *AIDS Rev.* 3, 11-23.
4. Azevedo-Pereira, J. M., Santos-Costa, Q., and Moniz-Pereira, J. (2005) HIV-2 infection and chemokine receptors usage – clues to reduced virulence of HIV-2, *Curr. HIV Res.* 3, 3-16.
5. Levy, J. A. (2007) *HIV and the Pathogenesis of AIDS*, 3rd ed., ASM Press, Washington, D.C.
6. HIV Sequence Compendium 2010 Kuiken, C., Leitner, T., Hahn, B., Mullins, J., Wolinsky, S., Foley, B., Apetrei, C., Mizrachi, I., Rambaut, A., and Korber, B., Eds. Published by Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, NM, LA-UR 10-03684.
7. HIV Replication Cycle (2010) National Institute of Allergy and Infectious Diseases, National Institutes of Health (NIH), Bethesda, Maryland. Accessed from <http://www.niaid.nih.gov> on January 2011.
8. Ashorn, P., McQuade, T. J., Thaisrivongs, S., Tomasselli, A. G., Tarpley, W. G., and Moss, B. (1990) An inhibitor of the protease blocks maturation of human and simian immunodeficiency viruses and spread of infection, *Proc. Natl. Acad. Sci. U.S.A.* 87, 7472-7476.
9. Condra, J. H., Schleif, W. A., Blahy, O. M., Gabryelski, L. J., Graham, D. J., Quintero, J. C., Rhodes, A., Robbins, H. L., Roth, E., Shivaprakash, M., Titus, D., Yang, T., Teppler, H., Squires, K. E., Deutsch, P. J., and Emini, E. A. (1995) *In vivo* emergence of HIV-1 variants resistant to multiple protease inhibitors, *Nature* 374, 569-571.
10. Barrett, A. J., Rawlings, N. D., and Woessner, J. F. (2003) *The Handbook of Proteolytic Enzymes*, Academic Press, London.
11. Shafer, R. W., Stevenson, D., and Chan, B. (1999) Human immunodeficiency virus reverse transcriptase and protease sequence database, *Nucleic Acids Res.* 27, 348-352.
12. Andreeva, N., and Gustchina, A. (1979) On the supersecondary structure of acid proteases, *Biochem. Biophys. Res. Commun.* 87, 32-42.

13. Andreeva, N. (1991) A consensus template of the aspartic proteinase fold, in *Structure and Function of the Aspartic Proteinases* (Dunn, B., Ed.), pp 559-572, Plenum Press, New York.
14. Wlodawer, A., and Gustchina, A. (2000) Structural and biochemical studies of retroviral proteases, *Biochim. Biophys. Acta* 1477, 16-34.
15. Dunn, B. M., Goodenow, M. M., Gustchina, A., and Wlodawer, A. (2002) Retroviral proteases, *Genome Biology* 3, reviews3006.3001-3006.3007.
16. Ali, A., Bandaranayake, R. M., Cai, Y., King, N. M., Kolli, M., Mittal, S., Murzycki, J. F., Nalam, M. N. L., Nalivaika, E. A., Özen, A., Prabu-Jeyabalan, M. M., Thayer, K., and Schiffer, C. A. (2010) Molecular basis for drug resistance in HIV-1 protease, *Viruses* 2, 2509-2535.
17. The PyMOL Molecular Graphics System, Version 1.1r1, Schrödinger, LLC.
18. Davies, D. R. (1990) The structure and function of the aspartic proteinases, *Annu. Rev. Biophys. Biophys. Chem.* 19, 189-215.
19. Navia, M. A., Fitzgerald, P. M., McKeever, B. M., Leu, C. T., Heimbach, J. C., Herber, W. K., Sigal, I. S., Darke, P. L., and Springer, J. P. (1989) Three-dimensional structure of aspartyl protease from human immunodeficiency virus HIV-1, *Nature* 337, 615-620.
20. Lapatto, R., Blundell, T., Hemmings, A., Overington, J., Wilderspin, A. F., Wood, S., Merson, J. R., Whittle, P. J., Danley, D. E., Geoghegan, K. F., Hawrylik, S. J., Lee, S. E., Scheld, K. G., and Hobart, P. M. (1989) X-ray analysis of HIV-1 proteinase at 2.7 Å resolution confirms structural homology among retroviral enzymes, *Nature* 342, 299-302.
21. Wlodawer, A., Miller, M., Jaskolski, M., Sathyanarayana, B. K., Baldwin, E., Weber, I. T., Selk, L. M., Clawson, L., Schneider, J., and Kent, S. B. H. (1989) Conserved folding in retroviral proteases: Crystal structure of a synthetic HIV-1 protease *Science* 245, 616-621.
22. Spinelli, S., Liu, Q. Z., Alzari, P. M., Hirel, P. H., and Poljak, R. J. (1991) The three-dimensional structure of the aspartyl protease from the HIV-1 isolate BRU, *Biochimie* 73, 1391-1396.
23. Rose, R. B., Craik, C. S., and Stroud, R. M. (1998) Domain flexibility in retroviral proteases: Structural implications for drug resistant mutations, *Biochemistry* 37, 2607-2621.

24. Gustchina, A., and Weber, I. T. (1990) Comparison of inhibitor binding in HIV-1 protease and in non-viral aspartic proteases: The role of the flap, *FEBS Lett.* 269, 269-272.
25. Tomasselli, A. G., Howe, W. J., Sawyer, T. K., Wlodawer, A., and Heinrikson, R. L. (1991) The complexities of AIDS: An assessment of the HIV protease as a therapeutic target, *Chim. Oggi* 9, 6-27.
26. Rick, S. W., Erickson, J. W., and Burt, S. K. (1998) Reaction path and free energy calculations of the transition between alternate conformations of HIV-1 protease, *Proteins* 32, 7-16.
27. Miller, M., Schneider, J., Sathyanarayana, B. K., Toth, M. V., Marshall, G. R., Clawson, L., Selk, L., Kent, S. B. H., and Wlodawer, A. (1989) Structure of complex of synthetic HIV-1 protease with a substrate-based inhibitor at 2.3 Å resolution, *Science* 246, 1149-1152.
28. Ishima, R., Freedberg, D. I., Wang, Y.-X., Louis, J. M., and Torchia, D. A. (1999) Flap opening and dimer-interface flexibility in the free and inhibitor-bound HIV protease, and their implications for function, *Structure* 7, 1047-1055.
29. Freedberg, D. I., Ishima, R., Jacob, J., Wang, Y.-X., Kustanovich, I., Louis, J. M., and Torchia, D. A. (2002) Rapid structural fluctuations of the free HIV protease flaps in solution: Relationship to crystal structures and comparison with predictions of dynamics calculations, *Protein Sci.* 11, 221-232.
30. Ishima, R., Ghirlando, R., Tözsér, J., Gronenborn, A. M., Torchia, D. A., and Louis, J. M. (2001) Folded monomer of HIV-1 protease, *J. Biol. Chem.* 276, 49110-49116.
31. Ishima, R., Torchia, D. A., Lynch, S. M., Gronenborn, A. M., and Louis, J. M. (2003) Solution structure of the mature HIV-1 protease monomer: Insight into the tertiary fold and stability of a precursor, *J. Biol. Chem.* 278, 43311-43319.
32. Louis, J. M., Ishima, R., Nesheiwat, I., Pannell, L. K., Lynch, S. M., Torchia, D. A., and Gronenborn, A. M. (2003) Revisiting monomeric HIV-1 protease. Characterization and redesign for improved properties, *J. Biol. Chem.* 278, 6085-6092.
33. Todd, M. J., Semo, N., and Freire, E. (1998) The structural stability of the HIV-1 protease, *J. Mol. Biol.* 283.
34. Todd, M. J., and Freire, E. (1999) The effect of inhibitor binding on the structural stability and cooperativity of the HIV-1 protease, *Proteins* 36, 147-156.

35. Todd, M. J., Luque, I., Velázquez-Campoy, A., and Freire, E. (2000) Thermodynamic basis of resistance to HIV-1 protease inhibition: Calorimetric analysis of the V82F/I84V active site resistant mutant, *Biochemistry* 39, 11876-11883.
36. Velázquez-Campoy, A., Todd, M. J., and Freire, E. (2000) HIV-1 protease inhibitors: Enthalpic versus entropic optimization of the binding affinity, *Biochemistry* 39, 2201-2007.
37. Hornak, V., Okur, A., Rizzo, R. C., and Simmerling, C. (2006) HIV-1 protease flaps spontaneously open and reclose in molecular dynamics simulations, *Proc. Natl. Acad. Sci. U.S.A.* 103, 915-920.
38. Collins, J. R., Burt, S. K., and Erickson, J. W. (1995) Flap opening in HIV-1 protease simulated by 'activated' molecular dynamics, *Nat. Struct. Biol.* 2, 334-338.
39. Ding, F., Layten, M., and Simmerling, C. (2008) Solution structure of HIV-1 protease flaps probed by comparison of molecular dynamics simulation ensembles and EPR experiments, *J. Am. Chem. Soc.* 130, 7184-7185.
40. Hornak, V., Okur, A., Rizzo, R. C., and Simmerling, C. (2006) HIV-1 protease flaps spontaneously close to the correct structure in simulations following manual placement of an inhibitor into the open state, *J. Am. Chem. Soc.* 128, 2812-2813.
41. Scott, W. R. P., and Schiffer, C. A. (2000) Curling of flap tips in HIV-1 protease as a mechanism for substrate entry and tolerance of drug resistance, *Structure* 8, 1259-1265.
42. Tóth, G., and Borics, A. (2006) Flap opening mechanism of HIV-1 protease, *J. Mol. Graph. Model.* 24, 465-474.
43. Nicholson, L. K., Yamazaki, T., Torchia, D. A., Grzesiek, S., Bax, A., Stahl, S. J., Kaufman, J. D., Wingfield, P. T., Lam, P. Y., and Jadhav, P. K. (1995) Flexibility and function in HIV-1 protease, *Nat. Struct. Biol.* 2, 274-280.
44. Heaslet, H., Rosenfeld, R., Giffin, M., Lin, Y.-C., Tam, K., Torbett, B. E., Elder, J. H., McRee, D. E., and Stout, C. D. (2007) Conformational flexibility in the flap domains of ligand-free HIV protease, *Acta Crystallogr. D Biol. Crystallogr.* 63(Pt 8), 866-875.
45. Galiano, L. M., Bonora, M., and Fanucci, G. E. (2007) Inter-flap distances in HIV-1 protease determined by pulsed EPR measurements, *J. Am. Chem. Soc.* 129, 11004-11005.

46. Galiano, L. M., Ding, F., Veloro, A. M., Blackburn, M. E., Simmerling, C., and Fanucci, G. E. (2009) Drug pressure selected mutations in HIV-1 protease alter flap conformations, *J. Am. Chem. Soc.* *131*, 430-431.
47. Galiano, L. M. 2008. Site-directed spin labeling EPR studies of conformational dynamics in the flap region of HIV-1 protease. Ph.D. dissertation. University of Florida, Gainesville, Florida.
48. Blackburn, M. E., Veloro, A. M., and Fanucci, G. E. (2009) Monitoring inhibitor-induced conformational population shifts in HIV-1 protease by pulsed EPR spectroscopy, *Biochemistry* *48*, 8765-8767.
49. Blackburn, M. E. 2010. Monitoring polymorphism and inhibitor-induced conformational ensemble shifts in HIV-1 protease via pulsed electron paramagnetic resonance. Ph.D. dissertation. University of Florida, Gainesville, Florida.
50. Kear, J. L., Blackburn, M. E., Veloro, A. M., Dunn, B. M., and Fanucci, G. E. (2009) Subtype polymorphisms among HIV-1 protease variants confer altered flap conformations and flexibility, *J. Am. Chem. Soc.* *131*, 14650-14651.
51. Kear, J. L. 2010. An electron paramagnetic resonance study of HIV-1 protease and the development of a soluble expression system for prorenin. Ph.D. dissertation. University of Florida, Gainesville, Florida.
52. Prabu-Jeyabalan, M., Nalivaika, E. A., and Schiffer, C. A. (2002) Substrate shape determines specificity of recognition for HIV-1 protease: Analysis of crystal structures of six substrate complexes, *Structure* *10*, 369-381.
53. Schechter, I., and Berger, A. (1967) On the size of the active site in proteases, *Biochem. Biophys. Res. Commun.* *27*, 157-162.
54. Xiang, Y., Ridky, T. W., Krishna, N. K., and Leis, J. (1997) Altered Rous sarcoma virus gag polyprotein processing and its effects on particle formation, *J. Virol.* *71*, 2083-2091.
55. Barrie, K. A., Perez, E. E., Lamers, S. L., Farmerie, W. G., Dunn, B. M., Sleasman, J. W., and Goodenow, M. M. (1996) Natural variation in HIV-1 protease, gag p7 and p6, and protease cleavage sites within gag/pol polyproteins: Amino acid substitutions in the absence of protease inhibitors in mother and children infected by human immunodeficiency virus type 1, *Virology* *219*, 407-416.

56. Pettit, S. C., Sheng, N., Tritch, R., Erickson-Viitanen, S., and Swanstrom, R. (1998) The regulation of sequential processing of HIV-1 gag by the viral protease, in *Aspartic Proteinases, Retroviral and Cellular Enzymes* (James, M. N. G., Ed.), pp 15-25, Plenum Press, New York.
57. Pettit, S. C., Moody, M. D., Wehbie, R. S., Kaplan, A. H., Nantermet, P. V., Klein, C. A., and Swanstrom, R. (1994) The p2 domain of human immunodeficiency virus type 1 gag regulates sequential proteolytic processing and is required to produce fully infectious virions, *J. Virol.* 68, 8017–8027.
58. Tie, Y., Boross, P. I., Wang, Y.-F., Gaddis, L., Liu, F., Chen, X., Tozser, J., Harrison, R. W., and Weber, I. T. (2005) 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 Journal* 272, 5265-5277.
59. Goodenow, M. M., Bloom, G., Rose, S. L., Pomeroy, S. M., O'Brien, P. O., Perez, E. E., Sleasman, J. W., and Dunn, B. M. (2002) Naturally occurring amino acid polymorphisms in human immunodeficiency virus type 1 (HIV-1) gag p7^{NC} and the C-cleavage site impact gag-pol processing by HIV-1 protease, *Virology* 292, 137-149.
60. Anderson, J., Schiffer, C., Lee, S.-K., and Swanstrom, R. (2009) Viral Protease Inhibitors, in *Handbook of Experimental Pharmacology* (Kräusslich, H.-G., and Bartenschlager, R., Eds.), Springer-Verlag Berlin, Heidelberg.
61. Menéndez-Arias, L. (2010) Molecular basis of human immunodeficiency virus drug resistance: An update, *Antiviral Res.* 85, 210-231.
62. Tözsér, J., Weber, I. T., Gustchina, A., Blaha, I., Copeland, T. D., Louis, J. M., and Oroszlan, S. (1992) Kinetic and modeling studies of S3-S3' subsites of HIV proteinases, *Biochemistry* 31, 4793-4800.
63. Dunn, B. M., Gustchina, A., Wlodawer, A., and Kay, J. (1994) Subsite preferences of retroviral proteinases, *Methods Enzymol.* 241, 254-278.
64. Wlodawer, A., and Vondrasek, J. (1998) Inhibitors of HIV-1 protease: A major success of structure-assisted drug design, *Annu. Rev. Biophys. Biomol. Struct.* 27, 249-284.
65. Wensing, A. M. J., van Maarseveen, N. M., and Nijhuis, M. (2010) Fifteen years of HIV protease inhibitors: Raising the barrier to resistance, *Antiviral Res.* 85, 59-74.
66. Tomasselli, A. G., and Heinrikson, R. L. (2000) Targeting the HIV-protease in AIDS therapy: A current clinical perspective, *Biochim. Biophys. Acta* 1477, 189-214.

67. Kozal, M. J., Shah, N., Shen, N., Yang, R., Fucini, R., Merigan, T. C., Richman, D. D., Morris, D., Hubbell, E., Chee, M., and Gingeras, T. R. (1996) Extensive polymorphisms observed in HIV-1 clade B protease gene using high-density oligonucleotide arrays, *Nature* 2, 753-759.
68. Los Alamos HIV Database (2011) Los Alamos National Laboratory, Los Alamos, New Mexico. Accessed from <http://hiv.lanl.gov> on January 2011.
69. Mansky, L. M., and Temin, H. M. (1995) Lower *in vivo* mutation rate of human immunodeficiency virus type 1 than that predicted from the fidelity of purified reverse transcriptase, *J. Virol.* 69, 5087-5094.
70. Kantor, R., Shafer, R. W., and Katzenstein, D. (2005) The HIV-1 non-subtype workgroup: An international collaboration for the collection and analysis of HIV-1 non-subtype B data, *J. Int. AIDS Soc.* 7, 71.
71. Le Vu, S., Le Strat, Y., Barin, F., Pillonel, J., Cazein, F., Bousquet, V., Brunet, S., Thierry, D., Semaille, C., Meyer, L., and Desenclos, J.-C. (2010) Population-based HIV-1 incidence in France, 2003-08: A modelling analysis, *The Lancet Infectious Diseases* 10, 682-687.
72. Wong-Staal, F. (1985) Name for AIDS virus, *Nature* 314, 574.
73. Shafer, R. W., Rhee, S. Y., Pillay, D., Miller, V., Sandstrom, P., Schapiro, J. M., Kuritzkes, D. R., and Bennett, D. (2007) HIV-1 protease and reverse transcriptase mutations for drug resistance surveillance, *AIDS* 21, 215-223.
74. HIV Drug Resistance Database (2011) Stanford University, Palo Alto, California. Accessed from <http://hivdb.stanford.edu> on January 2011.
75. Wu, T. D., Schiffer, C. A., Gonzales, M. J., Taylor, J., Kantor, R., Chou, S., Israelski, D., Zolopa, A. R., Fessel, W. J., and Shafer, R. W. (2003) Mutation patterns and structural correlates in human immunodeficiency virus type 1 protease following different protease inhibitor treatments, *J. Virol.* 77, 4836-4847.
76. Rhee, S.-Y., Taylor, J., Fessel, W. J., Kaufman, D., Towner, W., Troia, P., Ruane, P., Hellinger, J., Shirvani, V., Zolopa, A., and Shafer, R. W. (2010) HIV-1 protease mutations and protease inhibitor cross-resistance, *Antimicrob. Agents Chemother.* 54, 4253-4261.
77. Johnson, V. A., Brun-Vézinet, F., Clotet, B., Günthard, H. F., Kuritzkes, D. R., Pillay, D., Schapiro, J. M., and Richman, D. D. (2010) Update of the drug resistance mutations in HIV-1: December 2010, *Top HIV Med.* 18, 156-163.

78. Gulnik, S. V., Suvorov, L. I., Liu, B., Yu, B., Anderson, B., Mitsuya, H., and Erickson, J. W. (1995) Kinetic characterization and cross-resistance patterns of HIV-1 protease mutants selected under drug pressure, *Biochemistry* 34, 9282-9287.
79. Sugiura, W., Matsuda, Z., Yokomaku, Y., Hertogs, K., Larder, B., Oishi, T., Okano, A., Shiino, T., Tatsumi, M., Matsuda, M., Abumi, H., Takata, N., Shirahata, S., Yamada, K., Yoshikura, H., and Nagai, Y. (2002) Interference between D30N and L90M in selection and development of protease inhibitor-resistant human immunodeficiency virus type 1, *Antimicrob. Agents Chemother.* 46, 708-715.
80. Jacobsen, H., Yasargil, K., Winslow, D. L., Craig, J. C., Kröhn, A., Duncan, I. B., and Mous, J. (1995) Crystal structure of human immunodeficiency virus type 1 mutants with decreased sensitivity to proteinase inhibitor Ro 31-8959, *Virology* 206, 527-534.
81. Molla, A., Korneyeva, M., Gao, Q., Vasavanonda, S., Schipper, P. J., Mo, H.-M., Markowitz, M., Chernyavskiy, T., Niu, P., Lyons, N., Hsu, A., Granneman, G. R., Ho, D. D., Boucher, C. A. B., Leonard, J. M., Norbeck, D. W., and Kempf, D. J. (1996) Ordered accumulation of mutations in HIV protease confers resistance to ritonavir, *Nature Med.* 2, 760-766.
82. Menéndez-Arias, L., Martínez, M. A., Quiñones-Mateu, M. E., and Martínez-Picado, J. (2003) Fitness variations and their impact on the evolution of antiretroviral drug resistance, *Curr. Drug Targets-Infect. Disord.* 3, 355-371.
83. Muzammil, S., Ross, P., and Freire, E. (2003) A major role for a set of non-active site mutations in the development of HIV-1 protease drug resistance, *Biochemistry* 42, 631-638.
84. Clemente, J. C., Hemrajani, R., Blum, L. E., Goodenow, M. M., and Dunn, B. M. (2003) Secondary mutations M36I and A71V in the human immunodeficiency virus type 1 protease can provide an advantage for the emergence of the primary mutation D30N, *Biochemistry* 42, 15029-15035.
85. Mahalingam, B., Boross, P. I., Wang, Y. F., Louis, J. M., Fischer, C. C., Tozser, J., Harrison, R. W., and Weber, I. T. (2002) Combining mutations in HIV-1 protease to understand mechanisms of resistance, *Proteins* 48, 107-116.
86. Mahalingam, B., Louis, J. M., Reed, C. C., Adomat, J. M., Krouse, J., Wang, Y.-F., Harrison, R. W., and Weber, I. T. (1999) Structural and kinetic analysis of drug resistant mutants of HIV-1 protease, *Eur. J. Biochem.* 263, 238-245.

87. Lech, W. J., Wang, G., Yang, Y. L., Chee, Y., Dorman, K., McCrae, D., Lazzeroni, L. C., Erickson, J. W., Sinsheimer, J. S., and Kaplan, A. H. (1996) *In vivo* sequence diversity of the protease of human immunodeficiency virus type 1: Presence of protease inhibitor resistant variants in untreated subjects, *J. Virol.* **70**, 2038-2043.
88. Markowitz, M., Mo, H., Kempf, D. J., Norbeck, D. W., Bhat, T. N., Erickson, J. W., and Ho, D. D. (1995) Selection and analysis of human immunodeficiency virus type 1 variants with increased resistance to ABT-538, a novel protease inhibitor, *J. Virol.* **69**, 701-706.
89. Prabu-Jeyabalan, M., Nalivaika, E. A., King, N. M., and Schiffer, C. A. (2003) Viability of a drug-resistant human immunodeficiency virus type 1 protease variant: Structural insights for better antiviral therapy, *J. Virol.* **77**, 1306-1315.
90. Clemente, J. C., Moose, R. E., Hemrajani, R., Whitford, L. R. S., Govindasamy, L., Reutzel, R., McKenna, R., Agbandje-McKenna, M., Goodenow, M. M., and Dunn, B. M. (2004) Comparing the accumulation of active- and nonactive-site mutations in the HIV-1 protease, *Biochemistry* **43**, 12141-12151.
91. Piana, S., Carloni, P., and Rothlisberger, U. (2002) Drug resistance in HIV-1 protease: Flexibility-assisted mechanism of compensatory mutations *Protein Sci.* **11**, 2393-2402.
92. Foulkes-Murzycki, J. E., Scott, W. R. P., and Schiffer, C. A. (2007) Hydrophobic sliding: A possible mechanism for drug resistance in human immunodeficiency virus type 1 protease, *Structure* **15**, 225-233.
93. King, N. M., Prabu-Jeyabalan, M., Nalivaika, E. A., and Schiffer, C. A. (2004) Combating susceptibility to drug resistance: Lessons from HIV-1 protease, *Chem. Biol.* **11**, 1333-1338.
94. Zhang, Y. M., Imamichi, H., Imamichi, T., Lane, H. C., Falloon, J., Vasudevachari, M. B., and Salzman, N. P. (1997) Drug resistance during indinavir therapy is caused by mutations in the protease gene and its gag substrate cleavage sites, *J. Virol.* **71**, 6662-6670.
95. Bally, F., Martinez, R., Peters, S., Sudre, P., and Telenti, A. (2000) Polymorphism of HIV type 1 gag p7/p1 and p1/p6 cleavage sites: Clinical significance and implications for resistance to protease, *AIDS Res. Hum. Retroviruses* **16**, 1209-1213.
96. Mammano, F., Petit, C., and Clavel, F. (1998) Resistance-associated loss of viral fitness in human immunodeficiency virus type 1: Phenotypic analysis of protease and gag coevolution in protease inhibitor-treated patients, *J. Virol.* **72**, 7632-7637.

97. Doyon, L., Payant, C., Brakier-Gingras, L., and Lamarre, D. (1998) Novel gag-pol frameshift site in human immunodeficiency virus type 1 variants resistant to protease inhibitors, *J. Virol.* **72**, 6146-6150.
98. Maguire, M. F., Guinea, R., Griffin, P., Macmanus, S., Elston, R. C., Wolfram, J., Richards, N., Hanlon, M. H., Porter, D. J., Wrin, T., Parkin, N., Tisdale, M., Furfine, E., Petropoulos, C., Snowden, B. W., and Kleim, J. P. (2002) Changes in human immunodeficiency virus type 1 gag at positions L449 and P453 are linked to I50V protease mutants *in vivo* and cause reduction of sensitivity to amprenavir and improved viral fitness *in vitro*, *J. Virol.* **76**, 7398-7406.
99. Nijhuis, M., van Maarseveen, N. M., Lastere, S., Schipper, P., Coakley, E., Glass, B., Rovenska, M., de Jong, D., Chappey, C., Goedegebuure, I. W., Heilek-Snyder, G., Dulude, D., Cammack, N., Brakier-Gingras, L., Konvalinka, J., Parkin, N., Krausslich, H. G., Brun-Vezinet, F., and Boucher, C. A. B. (2007) A novel substrate-based HIV-1 protease inhibitor drug resistance mechanism, *PLoS Med.* **4**, e36.
100. Ho, S. K., Coman, R. M., Bunger, J. C., Rose, S. L., O'Brien, P., Munoz, I., Dunn, B. M., Sleasman, J. W., and Goodenow, M. M. (2008) Drug-associated changes in amino acid residues in gag p2, p7^{NC}, and p6^{Gag}/p6^{Pol} in human immunodeficiency virus type 1 (HIV-1) display a dominant effect on replicative fitness and drug response, *Virology* **378**, 272-281.
101. Rose, J. R., Salto, R., and Craik, C. S. (1993) Regulation of autoproteolysis of the HIV-1 and HIV-2 proteases with engineered amino acid substitution, *J. Biol. Chem.* **268**, 11939-11945.
102. 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., Heinrikson, R. L., and Tomasselli, A. G. (1994) The HIV-1 protease as enzyme and substrate: Mutagenesis of autolysis sites and generation of a stable mutant with retained kinetic properties, *Biochemistry* **33**, 9405-9413.
103. Logsdon, B. C., Vickrey, J. F., Martin, P., Proteasa, G., Koepke, J. I., Terlecky, S. R., Wawrzak, Z., Winters, M. A., Merigan, T. C., and Kovari, L. C. (2004) Crystal structures of a multidrug-resistant human immunodeficiency virus type 1 protease reveal an expanded active-site cavity, *J. Virol.* **78**, 3123-3132.
104. Coman, R. M., Robbins, A. H., Fernandez, M. A., Gilliland, C. T., Sochet, A. A., Goodenow, M. M., McKenna, R. M., and Dunn, B. M. (2008) The contribution of naturally occurring polymorphisms in altering the biochemical and structural characteristics of HIV-1 subtype C protease, *Biochemistry* **47**, 731-743.

105. Ohtaka, H., Schon, A., and Freire, E. (2003) Multi-drug resistance to HIV-1 protease inhibition requires cooperative coupling between distal mutations, *Biochemistry* 42, 13659-13666.
106. Voet, D., and Voet, J. G. (2004) *Biochemistry*, Vol. 1, 3rd ed., John Wiley & Sons, Inc., Hoboken, New Jersey.
107. Higgins, S. J., and Hames, B. D., (Eds.) (1999) *Protein Expression: A Practical Approach*, Oxford University Press, Inc., New York.
108. Howe, C. (2007) *Gene Cloning and Manipulation*, 2nd ed., Cambridge University Press, Cambridge.
109. Welch, M., Govindarajan, S., Ness, J. E., Villalobos, A., Gurney, A., Minshull, J., and Gustafsson, C. (2009) Design parameters to control synthetic gene expression in *Escherichia coli*, *PLoS ONE* 4, e7002.
110. Studier, F. W., and Moffatt, B. A. (1986) Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes, *J. Mol. Biol.* 189, 113-130.
111. Rosenberg, A. H., Lade, B. N., Chui, D. S., Lin, S. W., Dunn, J. J., and Studier, F. W. (1987) Vectors for selective expression of cloned DNAs by T7 RNA polymerase, *Gene* 56, 125-135.
112. Studier, F. W., Rosenberg, A. H., Dunn, J. J., and Dubendorff, J. W. (1990) Use of T7 RNA polymerase to direct expression of cloned genes, *Methods Enzymol.* 185, 60-89.
113. Studier, F. W. (1991) Use of bacteriophage T7 lysozyme to improve an inducible T7 expression system, *J. Mol. Biol.* 219, 37-44.
114. pET System Manual (2003) Novagen, Gibbstown, New Jersey. Accessed from <http://www.emdchemicals.com/life-science-research> on January 2011.
115. Phillips, T. A., VanBogelen, R. A., and Neidhardt, F. C. (1984) *lon* gene product of *Escherichia coli* is a heat-shock protein, *J. Bacteriol.* 159, 283-287.
116. Grodberg, J., and Dunn, J. J. (1988) *ompT* encodes the *Escherichia coli* outer membrane protease that cleaves T7 RNA polymerase during purification, *J. Bacteriol.* 170, 1245-1253.
117. Weiner, M. P., Anderson, C., Jerpseth, B., Wells, S., and Johnson-Browne, B. (1994) *Strategies* 7, 41-43.

118. QuikChange Site-Directed Mutagenesis Kit Instruction Manual (2005) Agilent Technologies Stratagene Products Division, Santa Clara, California. Accessed from <http://genomics.agilent.com> on January 2011.
119. Zheng, L., Baumann, U., and Reymond, J.-L. (2004) An efficient one-step site-directed and site-saturation mutagenesis protocol, *Nucleic Acids Res.* 32, e115:111-115.
120. Buxbaum, E. (2007) *Fundamentals of Protein Structure and Function*, Springer Science, New York.
121. Berg, J. M., Tymoczko, J. L., and Stryer, L. (2002) *Biochemistry*, W.H. Freeman and Co., New York.
122. Protein Electrophoresis: Technical Manual (1999) GE Healthcare Life Sciences, Pittsburgh, Pennsylvania. Accessed from <http://www.gelifesciences.com> on January 2011.
123. Boyer, R. (2000) *Modern Experimental Biochemistry*, Addison Wesley Longman, Inc., San Francisco, California.
124. Barrett, G. C., and Elmore, D. T. (2004) *Amino Acids and Peptides*, Cambridge University Press, Cambridge.
125. Fändrich, M., and Dobson, C. M. (2002) The behaviour of polyamino acids reveals an inverse side chain effect in amyloid structure formation, *The EMBO Journal* 21, 5682-5690.
126. Greenfield, N. J. (2007) Using circular dichroism spectra to estimate protein secondary structure, *Nature Protocols* 1, 2876 - 2890.
127. Whitmore, L., and Wallace, B. A. (2007) Protein secondary structure analyses from circular dichroism spectroscopy: Methods and reference databases, *Biopolymers* 89, 392-400.
128. Mikkelsen, S. R., and Corton, E. (2004) *Bioanalytical Chemistry*, John Wiley & Sons, Inc., Hoboken, New Jersey.
129. Matsuo, T., and Seyama, Y. (2000) Introduction to modern biological mass spectrometry, *J. Mass Spectrom.* 35, 114-130.
130. Weil, J. A., and Bolton, J. R. (1972) *Electron Spin Resonance: Elementary Theory and Practical Applications*, John Wiley and Sons, Inc., Hoboken, New Jersey.

131. Poole Jr., C. P. (1983) *Electron Spin Resonance. A Comprehensive Treatise on Experimental Techniques*, 2nd ed., John Wiley and Sons, New York.
132. Mchaourab, H. S., Lietzow, M. A., Hideg, K., and Hubbell, W. L. (1996) Motion of spin-labeled side chains in T4 lysozyme. Correlation with protein structure and dynamics, *Biochemistry* 35, 7692-7704.
133. Columbus, L., Kálai, T., Jekö, J., Hideg, K., and Hubbell, W. L. (2001) Molecular motion of spin labeled side chains in α -helices: Analysis by variation of side chain structure, *Biochemistry* 40, 3828–3846.
134. Langen, R., Oh, K. J., Cascio, D., and Hubbell, W. L. (2000) Crystal structures of spin-labeled T4 lysozyme mutants: Implications for the interpretation of EPR spectra in terms of structure, *Biochemistry* 39, 8396-8405.
135. Fanucci, G. E., and Cafiso, D. S. (2006) Recent advances and applications of site-directed spin labeling, *Curr. Opin. Struct. Biol.* 16, 644-653.
136. Hubbell, W. L., and Altenbach, C. (1994) Investigation of structure and dynamics in membrane proteins using site-directed spin labeling, *Curr. Opin. Struct. Biol.* 4, 566-573.
137. Hubbell, W. L., Cafiso, D. S., and Altenbach, C. (2000) Identifying conformational changes with site-directed spin labeling, *Nat. Struct. Biol.* 7, 735-739.
138. Columbus, L., and Hubbell, W. L. (2002) A new spin on protein dynamics, *Trends Biochem. Sci.* 27, 288-295.
139. Hubbell, W. L., Gross, A., Langen, R., and Lietzow, M. A. (1998) Recent advances in site-directed spin labeling of proteins, *Curr. Opin. Struct. Biol.* 8, 649-656.
140. Mchaourab, H. S., Kálai, T., Hideg, K., and Hubbell, W. L. (1999) Motion of spin-labeled side chains in T4 lysozyme: Effect of side chain structure, *Biochemistry* 38, 2947–2955.
141. Crane, J. M., Mao, C., Lilly, A. A., Smith, V. F., Suo, Y., Hubbell, W. L., and Randall, L. L. (2005) Mapping of the docking of secA onto the chaperone secB by site-directed spin labeling: Insight into the mechanism of ligand transfer during protein export, *J. Mol. Biol.* 353, 295-307.
142. Jeschke, G., and Polyhach, Y. (2007) Distance measurements on spin-labelled biomacromolecules by pulsed electron paramagnetic resonance, *Phys. Chem. Chem. Phys.* 9, 1895-1910.

143. Pannier, M., Veit, S., Godt, A., Jeschke, G., and Spiess, H. W. (2000) Dead-time free measurement of dipole-dipole interactions between electron spins, *J. Magn. Reson.* *142*, 331-340.
144. Rabenstein, M. D., and Shin, Y.-K. (1995) Determination of the distance between two spin labels attached to a macromolecule, *Proc. Natl. Acad. Sci. U.S.A.* *92*, 8239-8243.
145. Jeschke, G., Chechik, V., Ionita, P., Godt, A., and Zimmerman, H. (2006) DeerAnalysis2006 - A comprehensive software package for analyzing pulsed ELDOR data, *Appl. Magn. Reson.* *30*, 473-498.
146. XL1-Blue Competent Cells (2004) Agilent Technologies Stratagene Products Division, Santa Clara, California. Accessed from <http://www.genomics.agilent.com> on January 2011.
147. Velazquez-Campoy, A., Vega, S., and Freire, E. (2002) Amplification of the effects of drug resistance mutations by background polymorphisms in HIV-1 protease from African subtypes, *Biochemistry* *41*, 8613-8619.
148. Lippincott, J., and Apostol, I. (1999) Carbamylation of cysteine: A potential artifact in peptide mapping of hemoglobins in the presence of urea, *Anal. Biochem.* *267*, 57-64.
149. Shao, W., Everitt, L., Manchester, M., Loeb, D. D., Hutchison, C. A., and Swanstrom, R. (1997) Sequence requirements of the HIV-1 protease flap region determined by saturation mutagenesis and kinetic analysis of flap mutants, *Proc. Natl. Acad. Sci. U.S.A.* *94*, 2243-2248.
150. Yamazaki, T., Hinck, A. P., Wang, Y.-X., Nicholson, L. K., Torchia, D. A., Wingfield, P., Stahl, S. J., Kaufman, J. D., Chang, C.-H., Domaille, P. J., and Lam, P. Y. S. (1996) Three-dimensional solution structure of the HIV-1 protease complexed with DMP323, a novel cyclic urea-type inhibitor, determined by nuclear magnetic resonance spectroscopy, *Protein Sci.* *5*, 495-506.

BIOGRAPHICAL SKETCH

Estrella Garlit Gonzales, also known as Star, was born in 1981 in Southeastern Mindanao in the Philippines. She studied high school at the Philippine Science High School-Southern Mindanao Campus in Davao City. She moved to Manila and attended undergraduate school at the Ateneo de Manila University, where she obtained her B.S. Chemistry degree in March 2001. She worked in the same institution as an instructor and research assistant from 2002-2008. In August 2008, she moved to the United States and attended the Department of Chemistry graduate program at the University of Florida. She eventually joined the research group of Dr. Gail E. Fanucci. She obtained her M.S. Chemistry degree in August 2011.