Crystallographic Study of HIV Protease Drug Resistant Variants and Green Fluorescent Protein Based Calcium Sensor CatchER

Ying Zhang

Follow this and additional works at: http://scholarworks.gsu.edu/chemistry_diss

Recommended Citation
Zhang, Ying, "Crystallographic Study of HIV Protease Drug Resistant Variants and Green Fluorescent Protein Based Calcium Sensor CatchER." Dissertation, Georgia State University, 2013.
http://scholarworks.gsu.edu/chemistry_diss/83
CRYSTALLOGRAPHIC STUDY OF HIV PROTEASE DRUG RESISTANT VARIANTS AND GREEN FLUORESCENT PROTEIN BASED CALCIUM SENSOR CATCHER

by

YING ZHANG

Under the Direction of Dr. Irene T. Weber

ABSTRACT

HIV-1 protease (PR) is an important enzyme for the maturation of infectious virions and has been an effective drug target for HIV/AIDS for about two decades. PR inhibitors (PIs) were successfully designed based on the structural data for AIDS therapy. Nonetheless, the drug resistant PR variants are selected rapidly during therapy. Crystal structures of HIV-1 PR variants and their complexes with PIs have been studied so as to discover the mechanisms for drug resistance and assist in design of more effective drugs for AIDS therapy. Mutation L76V is associated with drug resistance and shows opposite effects on different protease inhibitors. Kinetics and stability of PR<sub>L76V</sub> containing L76V mutation were studied and high-resolution X-ray crystal structures of complexes of PR<sub>L76V</sub> with inhibitors, darunavir and saquinavir, were solved at resolutions of 1.46-1.45 Å to identify structural changes. HIV-1 PR<sub>P51</sub> variant is a multiple mutant selected for resistance to darunavir in the laboratory in vitro and it is of utility to investigate the mechanisms of HIV-1 resistance to darunavir. The crystal structures of an inactive form of PR<sub>P51</sub> have been determined at 1.50-1.66 Å resolution: a darunavir bound structure and a ligand free
structure. The kinetics and crystal structures of these drug resistant mutants provide the information to understand drug resistance mechanisms and give hints for design of novel inhibitors.

Ca$^{2+}$ plays an important role in the regulation of numerous biological functions in the human body as a ubiquitous signaling molecule. Calcium biosensor CatchER was designed by site-directed mutagenesis in the fluorescent sensitive location of chromophore. CatchER was studied via X-ray crystallography to investigate the calcium binding site and mechanisms of chromophore response to Ca$^{2+}$. Crystal structures of CatchER in the absence of Ca$^{2+}$ (apo form), complexed with Ca$^{2+}$ and also with Gd$^{3+}$ were determined at resolutions of 1.78-1.20 Å. Metal ions were identified in the designed calcium-binding site and structures showed the metal ion induced changes related to changes in optical properties. The structural information can be useful for the optimization of the sensor and assist the design of protein-ligand interaction based biosensors for the detection of various physiological molecules.

INDEX WORDS: X-ray crystallography, Crystal structure, Kinetics, HIV-1 protease, Protease inhibitor, Drug resistance, Saquinavir, Darunavir, L76V, Green fluorescence protein, Biosensor, Calcium signaling.
CRYSTALLOGRAPHIC STUDY OF HIV PROTEASE DRUG RESISTANT VARIANTS AND GREEN FLUORESCENT PROTEIN BASED CALCIUM SENSOR CATCHER

by

YING ZHANG

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

Doctor of Philosophy

in the College of Arts and Sciences

Georgia State University

2013
CRYSTALLOGRAPHIC STUDY OF HIV PROTEASE DRUG RESISTANT VARIANTS AND GREEN FLUORESCENT PROTEIN BASED CALCIUM SENSOR CATCHER

by

YING ZHANG

Committee Chair: Irene T. Weber

Committee: Jenny J. Yang
Robert W. Harrison
Stuart A. Allison

Electronic Version Approved:

Office of Graduate Studies
College of Arts and Sciences
Georgia State University
December 2012
ACKNOWLEDGEMENTS

I would like to give my special thanks and deep appreciation to my advisor, Dr. Irene T. Weber for giving the opportunity to work in her lab on several attractive projects. During the 6 years study in Dr. Weber’s lab, she not only inspired me with wonderful ideas and valuable suggestions in the research and training me to be a qualified research scientist. Her encouragement, advice and support have provided me the confidence, knowledge and techniques to perform scientific research. Without her guidance, I would not enter into the fantastic world of protein crystallography in my Ph.D career. Not only she is a distinguished mentor for me, she is also a great friend in personal life.

I also want to thank Dr. Jenny J. Yang, who has shared one of her interesting projects and collaborated with our lab so I can have an outstanding publication with high profile. During the collaboration period, she provided me a great help in understanding their research direction and collecting experimental information. Plus, she is also one of my committee members. I would like to acknowledge my two other committee members: Dr. Robert Harrison and Dr. Stuart Allison. Dr. Harrison is a professor in our joint group meeting and he has provided me supportive guidance and opinion on my research. Dr. Allison has given me a lot of useful suggestions for my research. With their help, I can fulfill my research work more completely and smoothly.

Also I want to express my sincere appreciation to my coworkers: Yuanfang Wang, Dr. Johnson Agniswamy, Dr. Eric Chang, Dr. Guoxing Fu, Dr. Yunfeng Tie, Dr. Bin Fang, Chen-Hsiang Shen, Hongmei Zhang, Tingyi Chiu, Xiaia Yu, Xiaodan Du, Maria Casanova, and Shaquita Greene from the Weber lab, and Dr. Shen Tang, Florence Reddish and You Zhuo from Dr. Jenny Yang’s lab for their support and concern in my research work. Thanks also should be sent to all the staff at the SER-CAT beamline at Argonne National Laboratory for their assistance during X-ray data collection for their great help.
Financially, I have to thank the Center for Diagnostics and Therapeutics (CDT) program for the fellowship to support me in recent years, and also this work is supported by NIH award R01 GM062920. Finally, I would love to especially thank my husband, Wen Zhang, who has always been supportive and encouraged me in both daily life and academic career.
# TABLE OF CONTENTS

ACKNOWLEDGEMENTS .................................................................................................................. iv

LIST OF TABLES .................................................................................................................................. x

LIST OF FIGURES ................................................................................................................................. xi

LIST OF ABBREVIATIONS ..................................................................................................................... xiv

1 INTRODUCTION ................................................................................................................................. 1

1.1 Overcoming Drug Resistance by Targeting HIV Protease Mutants ............................................. 1

1.1.1 *HIV Life Cycle* ............................................................................................................................. 1

1.1.2 *HIV Protease and Structures* ....................................................................................................... 6

1.1.3 *Substrate Specificity of HIV PR* .................................................................................................. 9

1.1.4 *Catalytic Mechanism of HIV-1 Protease* .................................................................................. 12

1.1.5 *Interactions between Protease and Substrates or Inhibitors* ..................................................... 15

1.1.6 *Structure-Guided Design of Protease Inhibitors* ................................................................. 18

1.1.7 *Drug Resistance for Protease Inhibitors in AIDS Therapy* .................................................... 21

1.1.8 *Specific Aims and Brief Results of Studying HIV-1 Drug Resistant Mutants PR<sub>L76V</sub>

    and PR<sub>P51</sub>* .......................................................................................................................... 27

1.2 Design of Green Fluorescent Protein Based Calcium Sensor .................................................... 30

1.2.1 *The Green Fluorescent Protein* ................................................................................................. 30

1.2.2 *Role of Calcium in Biological System* ...................................................................................... 38

1.2.3 *Calcium Measurement* .............................................................................................................. 40
1.2.4 Specific Aims and Brief Results of Studying GFP-based Calcium Sensor CatchER... 45

2 THE L76V DRUG RESISTANCE MUTATION DECREASES THE DIMER STABILITY AND RATE OF AUTOPROCESSING OF HIV-1 PROTEASE BY REDUCING INTERNAL HYDROPHOBIC CONTACTS

47

2.1 Abstract .......................................................................................................................... 47

2.2 Introduction ..................................................................................................................... 48

2.3 Experimental Procedures ............................................................................................. 50

2.3.1 Protein Expression and Purification ................................................................. 50

2.3.2 Enzyme Assays ......................................................................................................... 50

2.3.3 Calorimetry .................................................................................................................. 51

2.3.4 Autoprocessing of TFR-PR$_{L76V}$ Precursor ......................................................... 52

2.3.5 Crystallographic Analysis ......................................................................................... 53

2.3.6 Protein Data Bank Entries ....................................................................................... 54

2.4 Results and Discussion ................................................................................................. 54

2.4.1 Properties and Stability of Mature PR$_{L76V}$......................................................... 54

2.4.2 Inhibitor Binding ........................................................................................................ 54

2.4.3 Autoprocessing of PR$_{L76V}$ and PR$_{M46/L76V}$ Precursors .................................. 56

2.4.4 PR$_{L76V}$-Inhibitor Crystal Structures ..................................................................... 57

2.4.5 Structural Changes at the Site of Mutation .............................................................. 58

2.4.6 Protease-Inhibitor Interactions .................................................................................. 60

2.4.7 Implications for Drug Resistance .............................................................................. 61
3 STRUCTURES OF HIGHLY DARUNAVIR-RESISTANT HIV-1 PROTEASE MUTANT REVEAL WIDE
OPEN FLAPS AND ATYPICAL BINDING SITE FOR DARUNAVIR ........................................ 72

3.1 Abstract .......................................................................................................................... 72
3.2 Introduction ..................................................................................................................... 72
3.3 Results and Discussion .................................................................................................. 75
  3.3.1 Crystal Structures of Ligand-free and DRV Bound PR_{P51} ........................................ 75
  3.3.2 PR_{P51}/DRV Exhibits Unusual Binding Conformations of DRV .............................. 76
  3.3.3 Flaps of PR_{P51}/DRV and PR_{P51} Display Different Intersubunit Interactions .......... 78
  3.3.4 Multiple Mutations Contribute to the Structural Changes in PR_{P51} ....................... 79
3.4 Methods ......................................................................................................................... 82
  3.4.1 Construction, Expression and Purification ................................................................. 82
  3.4.2 Crystallization and Data Collection .......................................................................... 82
  3.4.3 Data Processing and Structure Determination ......................................................... 83
3.5 Acknowledgments ......................................................................................................... 83

4 STRUCTURAL BASIS FOR HAND-LIKE SITE IN CALCIUM SENSOR CATCHER WITH FAST
KINETICS ............................................................................................................................. 95

4.1 Abstract .......................................................................................................................... 95
4.2 Introduction ..................................................................................................................... 95
4.3 Experiments and Methods ........................................................................................... 98
  4.3.1 Expression and Purification ..................................................................................... 98
4.3.2 Crystallization, X-ray Data Collection and Structure Determination .................................. 99

4.3.3 Absorbance Spectrum of CatchER and Metal Binding Affinity via Fluorescence Spectroscopy .......................................................................................................................... 100

4.4 Results and Discussion ............................................................................................................ 102

4.4.1 Metal Binding Properties of CatchER ................................................................................. 102

4.4.2 Crystallographic Analysis of CatchER Structures ................................................................. 102

4.4.3 Identification of Metal Ions in Designed Binding Site of CatchER ..................................... 104

4.4.4 Structural Changes around the Chromophore ..................................................................... 106

4.4.5 Relationship between Mutations of the Novel Metal Binding Site and Optical Properties ................................................................................................................................. 109

4.5 Conclusions ............................................................................................................................ 110

4.6 Acknowledgements ................................................................................................................ 110

REFERENCES ........................................................................................................................................ 127

APPENDICES ........................................................................................................................................ 146

Appendix A: List of Crystal Structures ............................................................................................ 146

Appendix B: List of Publications ......................................................................................................... 146
LIST OF TABLES

Table 1.1 Protease cleavage sites in HIV-1 Gag and Gag-Pol polyproteins........................................... 11
Table 1.2 Selected major classes of GFPs.................................................................................................. 35
Table 2.1 Dimer dissociation, urea and thermal denaturation of PR_{L76V}.............................................. 64
Table 2.2 Crystallographic data collection and refinement statistics......................................................... 65
Table 3.1 Crystallographic data collection and refinement statistics....................................................... 84
Table 4.1 Relative amounts of CatchER-Gd complex formation with corresponding ratios ... 112
Table 4.2 Crystallographic data and refinement statistics ................................................................. 113
Table 4.3 CatchER-metal interactions within proposed coordination site .............................................. 114
Table 4.4 Data statistics vs resolution ..................................................................................................... 115
LIST OF FIGURES

Figure 1.1 Drawing of mature HIV virion surrounded by ribbon representation of HIV protein and protein fragment structures ................................................................. 3
Figure 1.2 General schemes of HIV-1 life cycle .......................................................... 4
Figure 1.3 Composition of Gag, Gag-Pol and Env polyproteins ................................... 5
Figure 1.4 Overall structure of HIV-1 PR/inhibitor complex in cartoon representation .... 8
Figure 1.5 Schematic diagram of a substrate (P4-P3’) bound to HIV-1 PR (S4-S3’) subsites 10
Figure 1.6 General scheme of proposed reaction pathway ........................................... 14
Figure 1.7 PR structures with/without inhibitor or substrate ........................................ 16
Figure 1.8 Hydrogen bond interactions between PR and substrates or inhibitors .......... 17
Figure 1.9 Currently approved FDA drugs for HIV PR with year of approval .................... 20
Figure 1.10 Mutations in the protease gene associated with resistance to protease inhibitors ................................................................................................................... 25
Figure 1.11 Sites of drug resistance mutations ............................................................. 26
Figure 1.12 Proposed mechanism for the intramolecular biosynthesis of chromophore ... 34
Figure 1.13 Fluorescence excitation and emission spectra and the chromophore structures 36
Figure 1.14 Schemes of the interactions between the chromophore and the neighboring residues and water molecules in wild type GFP and S65T mutant .............................. 37
Figure 1.15 The calcium pathways in the cell ............................................................... 39
Figure 1.16 Scheme of EGFP-based Ca^{2+} biosensor ................................................. 44
Figure 2.1 HIV-1 PR dimer structure bound to darunavir ............................................. 66
Figure 2.2 Urea denaturation of PR\textsubscript{L76V} and wild-type PR and kinetic determination of PR\textsubscript{L76V} dimer dissociation ........................................................................................................ 67
Figure 2.3 Thermochemical data for interactions of inhibitors with PR\textsubscript{L76V} ........... 68
Figure 2.4 Autocatalytic maturation of precursors TFR-PR_{L76V} and TFR-PR_{M46I/L76V} at pH 4.5 and its inhibition ................................................................................................................................. 69
Figure 2.5 Hydrophobic interactions of residue 76 in the crystal structures of PR_{L76V} and wild type PR ........................................................................................................................................... 70
Figure 2.6 Protease-inhibitor hydrogen bond interactions ........................................................................................................................................................................ 71
Figure 3.1 PR_{PS1} mutations ........................................................................................................................................................................................................ 85
Figure 3.2 DRV binds to different sites in PR and PR_{PS1} ..................................................................................................................................................... 86
Figure 3.3 Unique binding site for DRV in PR_{PS1}. ......................................................................................................................................................... 87
Figure 3.4 DRV has distinct hydrogen bonding interactions with wild type PR and PR_{PS1} ........................................................................................................ 88
Figure 3.5 Comparison of dimers of PR_{PS1}/DRV and wild type PR ................................................................................................................................................ 89
Figure 3.6 Comparison of two structures of PR_{PS1} ....................................................................................................................................................... 90
Figure 3.7 Comparison of dimers of PR_{PS1} and PR20/p2-NC_open ............................................................................................................................................. 91
Figure 3.8 Interactions of Leu89 in PR/DRV and Met89 in PR_{PS1}/DRV with neighboring residues .................................................................................................................. 92
Figure 3.9 DRV interactions with symmetry-related monomer of PR_{PS1} ............................................................................................................................................. 93
Figure 3.10 Comparison of PR_{PS1}, PR20 and wild type PR structures ................................................................................................................................. 94
Figure 4.1 Structure of CatchER indicating the locations of mutated residues S147E, S202D, Q204E, F223E and T225E, and chromophore CRO66 ...................................................................................................................... 117
Figure 4.2 Absorbance spectra of CatchER and metal binding analysis via fluorescence spectroscopy .............................................................................................................................................................................. 118
Figure 4.3 Selected F_{o}-F_{c} omit map in CatchER-Ca^{2+} CatchER-Gd^{3+} and CatchER(apo) .................................................................................................................. 119
Figure 4.4 Hydrogen bond interactions at the designed Ca^{2+} binding site in CatchER structures and superposition of CatchER-Ca^{2+} and CatchER-Gd^{3+} .................................................................................................................. 120
Figure 4.5 Protonation states of the GFP chromophore with the corresponding absorbance wavelength.................................................................................................................................. 121

Figure 4.6 Scheme of the hydrogen bond interactions around chromophore in CatchER structures, EGFP and GFP .................................................................................................................................. 122

Figure 4.7 Comparison of EGFP and CatchER(apo) at the chromophore and nearby residues .................................................................................................................................. 123

Figure 4.8 Fluorescence response of CatchER to Ca$^{2+}$ and Gd$^{3+}$ excited at 395 nm and fitted normalized intensity .................................................................................................................................. 124

Figure 4.9 Flexible loop region comprising residues 155-159 .................................................................................................................................. 125

Figure 4.10 Structure validation by Ramachandran plots for the three CatchER structures... 126
# LIST OF ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Å</td>
<td>Angstrom</td>
</tr>
<tr>
<td>ALA</td>
<td>alanine</td>
</tr>
<tr>
<td>ARG</td>
<td>arginine</td>
</tr>
<tr>
<td>ASP</td>
<td>aspartic acid</td>
</tr>
<tr>
<td>ASN</td>
<td>asparagine</td>
</tr>
<tr>
<td>BFP</td>
<td>blue fluorescent protein</td>
</tr>
<tr>
<td>CaM</td>
<td>calmodulin</td>
</tr>
<tr>
<td>CatchER</td>
<td>calcium sensor for detecting high concentration in the <strong>ER</strong></td>
</tr>
<tr>
<td>Cα</td>
<td>alpha carbon</td>
</tr>
<tr>
<td>CFP</td>
<td>cyan fluorescent protein</td>
</tr>
<tr>
<td>CYS</td>
<td>cysteine</td>
</tr>
<tr>
<td>DDT</td>
<td>dithiothreitol</td>
</tr>
<tr>
<td>DMSO</td>
<td>dimethylsulfoxide</td>
</tr>
<tr>
<td>DRV</td>
<td>darunavir (3R,3AS,6AR)-hexahydrofuro [2,3-B]furan-3- yl(1S,2R)-3-[<a href="isobutyl">4-aminophenyl)sulfonyl</a> amino]- 1-benzyl-2- hydroxypropylcarbamate</td>
</tr>
<tr>
<td>EGFP</td>
<td>enhanced green fluorescent protein</td>
</tr>
<tr>
<td>EGTA</td>
<td>ethylene glycol tetraacetic acid</td>
</tr>
<tr>
<td>ER</td>
<td>endoplasmic reticulum</td>
</tr>
<tr>
<td>FRET</td>
<td>fluorescent resonance energy transfer</td>
</tr>
<tr>
<td>GAG</td>
<td>group antigen</td>
</tr>
<tr>
<td>GFP</td>
<td>green fluorescent protein</td>
</tr>
<tr>
<td>Acronym</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>--------------------------------------------------</td>
</tr>
<tr>
<td>GLU</td>
<td>glutamic acid</td>
</tr>
<tr>
<td>GLN</td>
<td>glutamine</td>
</tr>
<tr>
<td>GLY</td>
<td>glycine</td>
</tr>
<tr>
<td>HIS</td>
<td>histidine</td>
</tr>
<tr>
<td>HIV-1</td>
<td>Human Immunodeficiency Virus 1</td>
</tr>
<tr>
<td>ILE</td>
<td>isoleucine</td>
</tr>
<tr>
<td>IPTG</td>
<td>isopropyl-D-thio-galactopyranoside</td>
</tr>
<tr>
<td>LB</td>
<td>Luria-Bertani</td>
</tr>
<tr>
<td>LEU</td>
<td>leucine</td>
</tr>
<tr>
<td>LYS</td>
<td>lysine</td>
</tr>
<tr>
<td>mM</td>
<td>millimolar</td>
</tr>
<tr>
<td>M</td>
<td>molar</td>
</tr>
<tr>
<td>MET</td>
<td>methionine</td>
</tr>
<tr>
<td>MR</td>
<td>molecular replacement</td>
</tr>
<tr>
<td>PDB</td>
<td>Protein Data Bank</td>
</tr>
<tr>
<td>PEG</td>
<td>polyethylene glycol</td>
</tr>
<tr>
<td>PHE</td>
<td>phenylalanine</td>
</tr>
<tr>
<td>POL</td>
<td>polymerase</td>
</tr>
<tr>
<td>PR</td>
<td>wild type HIV-1 protease</td>
</tr>
<tr>
<td>PR&lt;sub&gt;M46I/L76V&lt;/sub&gt;</td>
<td>PR with M46I and L76V mutation</td>
</tr>
<tr>
<td>PR&lt;sub&gt;L76V&lt;/sub&gt;</td>
<td>PR with L76V mutation</td>
</tr>
<tr>
<td>PR&lt;sub&gt;P51&lt;/sub&gt;</td>
<td>PR with 14 mutations from HIV strains at passaged 51</td>
</tr>
<tr>
<td>PRO</td>
<td>proline</td>
</tr>
<tr>
<td>RMSD</td>
<td>root mean square deviation</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>SER</td>
<td>serine</td>
</tr>
<tr>
<td>SQV</td>
<td>saquinavir cis-N-tert-butyl-decahydro-2-{2(R)-hydroxy-4-phenyl-3(S)-[[N-2-quinolylcarbonyl-L-asparaginy]amino]butyl}-(4AS)-isoquinoline-3(S)-carboxamide</td>
</tr>
<tr>
<td>SR</td>
<td>sarcoplasmic reticulum</td>
</tr>
<tr>
<td>THR</td>
<td>threonine</td>
</tr>
<tr>
<td>TRP</td>
<td>tryptophan</td>
</tr>
<tr>
<td>TYR</td>
<td>tyrosine</td>
</tr>
<tr>
<td>μl</td>
<td>microliter</td>
</tr>
<tr>
<td>VAL</td>
<td>valine</td>
</tr>
<tr>
<td>YFP</td>
<td>yellow fluorescent protein</td>
</tr>
</tbody>
</table>
1 INTRODUCTION

1.1 Overcoming Drug Resistance by Targeting HIV Protease Mutants

1.1.1 HIV Life Cycle

AIDS is a global pandemic and approximately 35.3 million people worldwide are living with HIV as of 2012 according to the UNAIDS report 2013 (UNAIDS, 2013). It has been well established that HIV (Human Immunodeficiency Virus) is the cause of AIDS (Acquired Immunodeficiency Deficiency Syndrome) for about three decades (Gallo et al., 1984). HIV is a member of retrovirus family (Levy et al., 1985). Therefore, its genetic information carrier is RNA that is the characteristic of all retroviruses (Barre-Sinoussi et al., 1983, Levy et al., 1985). Figure 1.1 shows the general construction of the mature HIV virion and ribbon representations of known structures of the viral proteins and protein fragments (Turner et al., 1999). The surface of virus is a lipid bilayer in which envelope proteins derived from the membrane of its host are embedded by surface unit glycoprotein (SU). In the middle of the spherical viral particle is a matrix shell composed by matrix protein (MA) and a capsid core that consists of the capsid protein (CA). Inside the capsid particle are two copies of the viral genome RNA and three most essential viral enzymes: protease (PR), reverse transcriptase (RT) and integrase (IN).

HIV life cycle is a process of viral infection and replication in a host cell, which involves four major steps: cell fusion and entry, reverse transcription of the viral RNA to DNA, integration into the host cell DNA and maturation of virus (Figure 1.2). Currently, the four key steps in HIV replication have been employed as anti-HIV drug targets. The virus fuses and enters the host cell via the recognition of the viral envelope glycoprotein gp120 and gp41 by CD4 receptors and other coreceptors on the host cells (Clapham et al., 1997, Maddon et al., 1986, Moore, 1997). The viral genetic material, two copies of single-strand RNA, and viral enzymes are released into the cytosol of the host cell. The viral RNA is first re-
verse transcribed into pro-viral DNA by the catalysis of viral reverse transcriptase and then the pro-viral DNA is integrated into the genome of the host cell by the catalytic action of integrase. The host cell provides the machine for the viral genome to replicate and to be translated into accessory proteins and three major polyproteins (Figure 1.3): Env, Gag, and Gag-pol. Env contains two exterior proteins SU (surface unit glycoprotein) and TM (transmembrane envelope). Then, viral RNA and proteins accumulate at the host cell surface to form an immature particle. Mature protease is formed by autoprocessing of Gag-Pol through folding and dimerization (Louis et al., 1999a, Louis et al., 2000). Finally, as protease catalyzes the cleavage of Gag and Gag-Pol to produce structural proteins including MA, CA and NC and functional enzymes such as RT, IN and PR, the virus becomes mature and infectious (Emini, 2002).
Figure 1.1 Drawing of mature HIV virion surrounded by ribbon representation of HIV protein and protein fragment structures. (Turner et al., 1999)
Figure 1.2 General schemes of HIV-1 life cycle. (Turner et al., 1999)
**Figure 1.3 Composition of Gag, Gag-Pol and Env polyproteins.** MA is matrix, CA is capsid, NC is nucleocapsid, TF is transframe, RT is reverse transcriptase, RH is RNase H, IN is integrase, p1 and p2 are spacer peptides, p6 is peptide at 3’ region of the Gag precursor, SU is surface unit glycoprotein and TM is transmembrane envelope.
1.1.2 **HIV Protease and Structures**

HIV protease (PR) is an enzyme that is responsible for catalyzing the cleavage of two polyproteins—Gag and Gag-pol into structural and functional proteins for the maturation of virus. Thus, PR is essential for production of infectious virus (Weber *et al.*, 2009b). It is encoded in the 5’ end of the Pol gene, expressed as part of the Gag-Pol polyprotein produced from a ribosomal frameshift (Louis *et al.*, 2000) and become active through autoprocessing (Louls *et al.*, 1994). Protease inhibitors can block the proteolytic activity of PR and the cleavage of Gag and Gag-Pol is interrupted, resulting in failure of viral replication (Kohl *et al.*, 1988, Seelmeier *et al.*, 1988). The three dimensional structures of HIV PR in the absence and presence of inhibitor have been determined by X-ray crystallography (Lapatto *et al.*, 1989, Vondrasek *et al.*, 1997, Wlodawer *et al.*, 1989, Wlodawer *et al.*, 1993). As shown in Figure 1.4, HIV-1 PR is a homodimer with 99 amino acid residues in each subunit, and the residues are numbered 1-99 in one subunit and 1’-99’ in the other subunit. Folding of the two subunits forms the active site where substrates bind and hydrolysis occurs. PR is a member of the aspartic protease family, of which the active site shares the catalytic triad Asp25-Thr26-Gly27. PR inhibitors are designed to competitively bind in the active site to inhibit the binding and cleavage of natural substrates. The secondary structures of each subunit of PR contain one short α-helix near the C terminus and two β sheets, which are connected by turns. Two antiparallel β-strands composed by residues 44-57 from one subunit create the flexible flap. Residues 8, 23-30, 32, 45-50, 53, 56, 76, 80-82 and 84 form the substrate binding site. PR residues 1-3, 5-9, 23-27, 29, 47-52, 54, 67, 81, 87 and 90-99 are engaged in the dimer interface contacts, including non-covalent interactions, hydrophobic packing of side chains and interactions. Additionally, the stabilization of each subunit is achieved by the contacts of aliphatic residues in a hydrophobic core.

The glycine-rich loop from residues 44-57 is known as the flaps. The two flaps in the dimer exist in the dynamic equilibrium of fully open, semi-open and closed conformations as analyzed by NMR and molecular dynamic simulations (Hornak *et al.*, 2006, Freedberg *et al.*, 2002). The fully open flaps allow
the entry of substrate and the closed conformation will occur upon binding inhibitors or peptides, while the semi-open form is observed in the structures of unliganded PR (Nicholson et al., 1995, Freedberg et al., 2002). Mutation of residues on the flaps has been studied to identify the role of side chains for PR activity (Shao et al., 1997). Based on the side chains, the residues on the flaps can be divided into three groups: Met46, Phe53 and Lys55 with side chains directed into solvent show tolerance to substitutions; Ile47, Ile50, Ile54 and Val56 with side chains directed into the protein interior and tolerate a few conservative substitutions; and Gly48, Gly49, Gly51 and Gly52 in the Gly-rich region which is sensitive to substitutions (Shao et al., 1997). Different degrees of separation of the flaps have been observed in many structures of wild type PR and PR variants, such as apo wild type PR, PR variants MDR769, PR20 and in solution studies using Double Electron-Electron Resonance (DEER) spectroscopy (Martin et al., 2005, Liu et al., 2006, Wlodawer et al., 1989, de Vera et al., 2012, Spinelli et al., 1991, Lapatto et al., 1989, Navia et al., 1989, Heaslet et al., 2007, Logsdon et al., 2004, Agniswamy et al., 2012). The flexibility of flaps to undergo movements is critical for the enzyme to bind and cleave substrate and disruption of these movements could lead to the inhibition of the enzyme (Gustchina et al., 1990, Scott et al., 2000, Kurt et al., 2003).
Figure 1.4 Overall structure of HIV-1 PR/inhibitor complex in cartoon representation. The two subunits are colored in grey and cyan, respectively. Flaps and 80’s loops are colored in red and green, respectively. The Asp25/Asp25’ are shown as balls and sticks in each subunit. The inhibitor darunavir (DRV) is represented by spheres and yellow sticks. The double-sided arrow indicates the dimer interface. The two one-sided arrows indicate the two flaps and the two 80’s loops.
1.1.3 Substrate Specificity of HIV PR

HIV PR hydrolyzes the peptide bond in several different cleavage sites in the Gag and Gag-Pol precursors (Figure 1.3). It is important that knowing the substrate specificity of HIV PR can assist in the design of new anti-viral drugs. Despite the symmetric characteristic of PR dimer, the substrate binds in an asymmetric manner (Figure 1.5). The residues of substrate peptide at carbonyl side of peptide bond are defined as P1, P2, P3 and P4, and P1’, P2’, P3’ and P4’ in the amino side. To accommodate P4’-P4, the residues of PR are correspondingly defined as forming subsites S4’-S4. The substrate specificity is determined by the shape and chemical structure of side chains of amino acids around the cleavage site (Konvalinka et al., 1990, Tozser et al., 1992, Tozser et al., 1991). Table 1.1 lists the cleavage site sequences on Gag and Gag-Pol polyproteins (Louis et al., 2000). The sequences are quite diverse, but still show some common features (Pettit et al., 1991, Poorman et al., 1991, Tozser et al., 1992, Louis et al., 2000). Two major types of cleavage sites at P1 were proposed for HIV-1 PR: type 1-Tyr(Phe)-Pro at P1-P1’; type 2-hydrophobic residues (excluding Pro) at P1 and P1’. Hydrophobic and small polar residues typically occupy P2 and P2’, while P3 is glutamine or basic amino acids and P4 is normally a small amino acid. Enzyme kinetics studies using oligopeptides have showed that PR cannot cleave peptides with β-branched amino acids and Pro, Ser or Gly substituted in the P1 position (Tozser et al., 1992, Phylip et al., 1990). And also charged amino acids at P1 or P1’ positions are not preferred in substrates (Konvalinka et al., 1990, Cameron et al., 1993). P3 and P3’ can be occupied by various residues while the P2’ preference is dependent on the type of amino acid at P1 and P1’ positions (Cameron et al., 1993, Konvalinka et al., 1990, Tozser et al., 1992).
Figure 1.5 Schematic diagram of a substrate (P4-P3') bound to HIV-1 PR (S4-S3') subsites. The scissile bond is indicated by a red X. (Courtesy of Dr Irene T. Weber)
### Table 1.1 Protease cleavage sites in HIV-1 Gag and Gag-Pol polyproteins. *(Louis et al., 2000)*

<table>
<thead>
<tr>
<th>Location of the cleavage site</th>
<th>P4</th>
<th>P3</th>
<th>P2</th>
<th>P1</th>
<th>P1’</th>
<th>P2’</th>
<th>P3’</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>In Gag</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MA-CA</td>
<td>Ser</td>
<td>Gln</td>
<td>Asn</td>
<td>Tyr</td>
<td>Pro</td>
<td>Ile</td>
<td>Val</td>
</tr>
<tr>
<td>CA-p2</td>
<td>Ala</td>
<td>Arg</td>
<td>Val</td>
<td>Leu</td>
<td>Ala</td>
<td>Glu</td>
<td>Ala</td>
</tr>
<tr>
<td>p2-NC</td>
<td>Ala</td>
<td>Thr</td>
<td>Ile</td>
<td>Met</td>
<td>Met</td>
<td>Gln</td>
<td>Arg</td>
</tr>
<tr>
<td>NC-p1</td>
<td>Arg</td>
<td>Gln</td>
<td>Ala</td>
<td>Phe</td>
<td>Phe</td>
<td>Leu</td>
<td>Gly</td>
</tr>
<tr>
<td>p1-p6</td>
<td>Pro</td>
<td>Gln</td>
<td>Asn</td>
<td>Leu</td>
<td>Leu</td>
<td>Gln</td>
<td>Ser</td>
</tr>
<tr>
<td>in p6</td>
<td>Lys</td>
<td>Glu</td>
<td>Leu</td>
<td>Pro</td>
<td>Pro</td>
<td>Leu</td>
<td>Thr</td>
</tr>
<tr>
<td><strong>In Pol</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TFP-p6</td>
<td>Asp</td>
<td>Leu</td>
<td>Ala</td>
<td>Phe</td>
<td>Leu</td>
<td>Gln</td>
<td>Gly</td>
</tr>
<tr>
<td>p6</td>
<td>Ser</td>
<td>Phe</td>
<td>Asn</td>
<td>Phe</td>
<td>Pro</td>
<td>Gln</td>
<td>Ile</td>
</tr>
<tr>
<td>PR-RT</td>
<td>Thr</td>
<td>Leu</td>
<td>Asn</td>
<td>Phe</td>
<td>Pro</td>
<td>Ile</td>
<td>Ser</td>
</tr>
<tr>
<td>p66-p51</td>
<td>Ala</td>
<td>Glu</td>
<td>Thr</td>
<td>Phe</td>
<td>Tyr</td>
<td>Val</td>
<td>Asp</td>
</tr>
<tr>
<td>RT-IN</td>
<td>Arg</td>
<td>Lys</td>
<td>Ile</td>
<td>Leu</td>
<td>Phe</td>
<td>Leu</td>
<td>Asp</td>
</tr>
</tbody>
</table>
1.1.4 Catalytic Mechanism of HIV-1 Protease

As a member of the aspartic protease family, the HIV-1 PR contains a conserved triplet, Asp25-Thr26-Gly27 at the active site. The mutation of Asp25 results in the complete inactivation of the HIV-1 PR and thus the Asp25 is a critical residue in the reaction to hydrolyze the peptide bond (Seelmeier et al., 1988, Kohl et al., 1988). Although several proteolytic mechanisms have been proposed for HIV PR, the detailed catalytic mechanism is still not fully understood. A commonly accepted basic mechanism is the general acid-base mechanism for aspartic PR in despite of some difference on details among the various possible mechanisms. As shown in Figure 1.6, Hyland proposed the mechanism based on solvent isotope effects and structure data (Hyland et al., 1991, Brik et al., 2003). In this mechanism, the water molecule is activated by the unprotonated Asp25’ and becomes nucleophilic. Then the hydroxyl of water attacks the carbonyl C of the substrate and the proton of one of the carboxylate O of Asp25 is transferred to carbonyl O of the substrate while the photon of water is transferred to the Asp25’. A diol tetrahedral intermediate is formed and then the newly gained proton of Asp25’ then is transferred to the amide N of the substrate; simultaneously (Hyland et al., 1991, Mitra et al., 1996) or subsequently (Okimoto et al., 1999) a proton from hydroxyl group of the carbonyl C donates to the Asp25 (Piana et al., 2002). Among those mechanisms, the nucleophilic water is critical and should be located between the two Asp25/25’. However, crystal structures of HIV PR complexed with inhibitors reveal that there is no water molecule or other atoms bigger than a proton at this position due to the lack of space.

Three consecutive steps in the proteolytic reaction of HIV-1 PR have been captured in crystal structures of PR and other mutants (Shen et al., 2012, Kumar et al., 2005, Kovalevsky et al., 2007). Structural analysis reveals the gem-diol tetrahedral intermediate, which has also been reported in previous work on other aspartic proteinases (Veerapandian et al., 1992, Holladay et al., 1985). Besides the agreement with the general features of the reaction pathway with three stages, Gly27 was proposed to have a potential role in recycling from the product complex to the ligand-free enzyme (Shen et al., 2012).
Despite variation in details of the proposed mechanisms, they agree with involvement of the nucleophilic water and the key role of the two aspartic residues.
Figure 1.6 General scheme of proposed reaction pathway. (Brik et al., 2003)
1.1.5 *Interactions between Protease and Substrates or Inhibitors*

The substrate binding site mainly consists of two characteristic catalytic triplets, two flexible flaps and 80’s loop (Figure 1.4). When substrates or inhibitors enter the active site, the flaps are expected to fold down to bind substrate and exclude water molecule from the active site, changing from the open conformation to the closed form (Figure 1.7) (Louis *et al.*, 2000, Wlodawer *et al.*, 1993). As shown in Figure 1.8A and B, one buried water molecule that connects the ligand and Ile50/50’ NH groups of the flaps is conserved in most PR-inhibitor complexes including substrate analogs (Harrison *et al.*, 1994, Gustchina *et al.*, 1990). Mutation of catalytic residues at the active site or disruption of the dimer interface contacts of active HIV-1 protease monomers will lead to the elimination of the catalytic activity of the PR and thus block the viral maturation. Inhibitors or substrates are recognized in the active site of PR by different types of interactions including hydrogen bonds, van der Waals contacts and CH...O interactions depending on the shape and chemistry of the groups or side chains at each position (Figure 1.8A and B). A set of hydrogen bonds interactions is conserved between the main chain amides and carbonyl oxygens of PR, involving the amide and carboxylate oxygen of Asp29, the carbonyl oxygen of Gly27 and the amide and carbonyl of Gly48 from both subunits. Clinical drugs have smaller chemical structure and fewer interactions with PR compared to the substrate analogs.
Figure 1.7 PR structures with/without inhibitor (or substrate). Red one presents PR dimer with bound inhibitor (or substrate) and blue for PR without inhibitor (or substrate).
Figure 1.8 Hydrogen bond interactions between PR and substrates or inhibitors. **A.** PR and substrate analog p2/NC. **B.** PR and inhibitor darunavir. PR and ligands are in stick representation; water molecules are indicated by red spheres; hydrogen bonds are shown as black dotted lines. (Tie et al., 2005, Tie et al., 2004)
1.1.6 Structure-Guided Design of Protease Inhibitors

Protease inhibitors (PIs) have been successful in treating HIV infection to provide HIV-infected individuals with a longer life span and improve the quality of life (Wlodawer et al., 1998). Currently 9 FDA (Food and Drugs Administration) approved drugs include the PIs: saquinavir, ritonavir, indinavir, nelfinavir, lopinavir, atazanavir, fosamprenavir, tipranavir, and darunavir (Figure 1.9). Except for tipranavir, the design of all other PIs is based on mimicking the substrate analog (Roberts et al., 1990). Structure-guided drug design is one of the powerful approaches in drug development if molecular structures of targets are available or predicted accurately. The initial design of anti-viral PIs was based on X-ray crystal structures of HIV-1 PR complexed with non-hydrolysable substrate analogs in which the peptide bond at the cleavage site is reduced from carbonyl to methylene or hydroxymethylene group (Miller et al., 1989). Structural analysis of crystal structures of HIV-1 PR with the peptidic inhibitors shows that main chain-main chain interactions including hydrogen bonds and hydrophobic interactions connect peptide inhibitor with PR residues 25-29 and 48-50 (Figure 1.8A) (Gustchina et al., 1990, Tie et al., 2005). However, there are several disadvantages in applying the substrate analogs in the clinics, such as large size, difficult penetration into cells, easy digestion by other proteinases. Therefore, the following design of new inhibitors utilizes smaller and less peptidic organic molecules.

In 1995, saquinavir, which was designed based on the crystal structures of HIV-1 PR complexed with substrate analogs, was the first PR inhibitor approved for AIDS therapy. Saquinavir still retains many peptide-like features, such as the peptidic backbone (Figure 1.7) maintaining the hydrogen bonds with main chain of PR (Tie et al., 2007, Krohn et al., 1991). The first generation of clinical inhibitors also contains large hydrophobic groups to achieve high affinity for PR. Nevertheless, drug resistant mutations that can alter the interactions between PR and inhibitors can reduce the affinity for these drugs (Tie et al., 2007, Liu et al., 2005). Structural comparison of PR and its variants demonstrated minimal deviation in the backbone atoms around the active site and some conserved residues that are essential to main-
tain the critical PR function in the viral replication. Therefore, this structural information provided insights into the design strategy of maximizing inhibitor interactions with PR backbones in the S2–S2’ sub-sites and conserved residues in the active site for the next generation of inhibitor (Ghosh et al., 2012). The second generation of drugs was designed based on Dr. Ghosh’s backbone-binding concept to target drug resistant PR as exemplified by darunavir (Figure 1.7), which is the latest PR inhibitor approved by FDA. Darunavir shows exceptional performance in cells infected with HIV bearing drug resistant PR mutations (Koh et al., 2003). A bis-tetrahydrofuran (THF) in the P2 group was used to introduce more non-peptide hydrogen bonds with main chain atoms and conserved residues of PR (Tie et al., 2004). Not only was DRV bound in the regular active site with key interactions with the backbone of PR in wild type and resistant PR mutants but a unique second binding site on one flap region was observed in high resolution X-ray structures which agreed with its kinetic study result of a “mixed-type competitive-uncompetitive inhibition” model (Tie et al., 2004, Liu et al., 2008, Kovalevsky et al., 2008). Since darunavir has demonstrated remarkable anti-viral efficiency in AIDS therapy for combating HIV drug resistance, further design based on the backbone scaffold of DRV provided a series of new inhibitors. The design of new inhibitors varies the chemical groups at P2, P1’ and P2’, most of which exhibit inhibition of PR in picomolar enhanced interaction with the conserved residues (Ghosh et al., 2012). However, the persistent emergence of drug resistance viral strains necessitates the design of another generation of inhibitor as a long term challenge and the backbone-binding strategy will be useful to guide future development of anti-AIDS agents.
Figure 1.9 Currently approved FDA drugs for HIV PR with year of approval.
1.1.7 Drug Resistance for Protease Inhibitors in AIDS Therapy

However, the most severe concern in AIDS therapy is the emergence of drug resistance after drug administration even when the patients achieve the viral suppression initially (Johnson et al., 2013). The rapid development of mutations in the virus is caused by multiple factors. Since the genetic information carrier of virus is RNA, reverse transcriptase (RT) is used to produce DNA from viral RNA and RT lacks 3’-5’ exonuclease proofreading function, therefore mutations during viral replication are intrinsically prone to occur. During the reverse transcription, roughly one base mutation is introduced in each viral life cycle with the high error rate of RT (Preston et al., 1988, Mansky, 1998). Another factor is the high rate of replication of the virus with as many as $10^9$ virus particles produced daily in a fully infected patient (Ho et al., 1995). Drug pressure causes the rapid selection of drug resistant strains in the antiviral therapy. The RT errors and rapid replication make selection of resistant variants faster in HIV. In the presence of antiretroviral drugs, drug resistant strains are selected and replace the wild-type virus (Drake, 1993, Wei et al., 1995).

Due to the polymorphous gene of HIV-1 PR even in the absence of inhibitors, as many as 50 different residues mutate even in the absence of PIs and mutations of more than 30 different residues have been reported to be associated with resistance to PIs (Figure 1.10) (Johnson et al., 2013). The mutations in PR upon long exposure to the drugs are selected to confer drug resistance to PIs. Multiple mutations in the PR emerged in ordered, stepwise fashion in response to PIs (Molla et al., 1996). Single or multiple mutations can cause multidrug resistance and cross-resistance (Figure 1.10). Drug resistant mutations in PR can change the enzyme properties like lowering catalytic activity or stability, while different combinations of single mutations in PR or the cleavage sites of Gag and Gag-Pol substrates were observed in the resistant clinical isolates which can compensate for the changes caused by the initial PR mutations (Zhang et al., 1997, Cote et al., 2001, Rhee et al., 2003, Shafer et al., 1999). Drug therapy selects for resistant mutations that may tend to compromise proper protein folding, structural stability,
and catalytic activity on various substrates (Erickson et al., 1996, Weber et al., 2009a). To overcome the problem of drug resistance and develop new antiviral therapies, it is important to investigate the drug resistance mechanisms.

Many crystal structures of drug resistant PR variants have been solved in complex with the corresponding PIs, such as darunavir, saquinavir, indinavir and nelfinavir. Resistance mechanisms could depend on the specific type of mutations and drugs in therapy. Drug resistant mutations can be simply divided into substitutions in the active site (inhibitor binding site), the dimer interface and other distal regions of PR (Figure 1.11): mutations at the active site can lead to the reduced affinity of the mutated PR to inhibitor; mutations at the dimer interface can cause altered structural stability of PR; and mutations of distal mutations can show a variety of effects (Weber et al., 2009a). PR variants with a single drug resistant mutation were studied to better understand the effect of an individual mutation rather than a multiply mutated variant. Substitution of residues in the active site can diminish direct interactions with inhibitor. Mutations of residues Leu23, Asp30, Val32, Met46, Ile47, Gly48, Ile50, Val82, and Ile84 that form interactions with substrates or inhibitors in the active site cavity have been reported in drug resistance (Johnson et al., 2013). The loss of interactions is observed in the crystal structures of PR variants with PIs. For example, in the complex of I84V variant with darunavir, the mutation resulted in loss of two van der Waals contacts between the residue 84 and the inhibitor due to reduced size of valine compared to isoleucine and this agrees with the Ki value of this variant for darunavir showing a 3-fold increase compared with the value for wild type PR (Tie et al., 2004). The flap mutation G48V differs from many other drug resistant mutations due to the increased size of the amino acid side chain. G48V is selected by saquinavir treatment and structural studies of a double mutant bearing G48V/L90M revealed the loss of an important hydrogen bond between carbonyl oxygen of residue 48 and the amide of saquinavir at P2 (Mahalingam et al., 1999). Mutations of residues Leu24, Ile50 and Phe53 that lie at the interface between the two subunits have been shown to affect the PR stability (Liu et al., 2005, Liu et al.,
The dimer stability studies were based on the sensitivity to urea and the value of dimer dissociation. PR with the single mutation of L24I, I50V or F53L has significantly decreased stability as assessed by urea-denaturation and increased dimer dissociation value compared to wild type PR. Reduced subunit-subunit interactions appear in structures of variants L24I and I50V with indinavir. I50V is also a mutation within the active site and PR with I50V has lost four hydrophobic contacts with between residues 50 from the two subunits and indinavir compared with the 9 hydrophobic contacts in wild type PR. Mutations at the distal regions can alter residues outside of the active site cavity and they occur at the PR locations that far from inhibitors or dimer interface. Although they do not directly interact with inhibitors, the mutated side chains perturb interactions in the PR dimer through an extended structural network and can indirectly interfere with the binding of the inhibitor. The distal mutations of I54M/V, L90M, and G73S have shown diverse effects to contribute to the drug resistance: mutation of residue 54 induces changes in residues 80-82 (the 80’s loop) that interact with inhibitors (Liu et al., 2008); L90M perturbs the catalytic triplet (Asp25-Thr26-Gly27) and further alters the affinity for inhibitors (Mahalingam et al., 2004, Kozisek et al., 2007); G73S altered specificity for substrate peptides and caused the changes in internal network of hydrogen bonds (Liu et al., 2005). Based on these studies on PR with single mutations, diverse drug resistance mechanisms have been proposed: the loss of interactions with inhibitors or peptides; loss of intersubunit contacts and stability (Weber et al., 2009a).

In clinical isolates, multiple mutations accumulate together in PR in response to PIs and combine to contribute to drug resistance from all individual mutations (Ohtaka et al., 2003, Henderson et al., 2012). A mutant protease PR20 containing 20 substitutions of Q7K, L10F, I13V, I15V, D30N, V32I, L33F, E35D, M36I, S37N, I47V, I54L, Q58E, I62V, L63P, A71V, I84V, N88D, L89T and L90M exhibits more than 1000-fold reduced susceptibility compared to wild type PR for darunavir (DRV) and saquinavir (SQV) (Louis et al., 2011a). Crystal structure of PR20 with DRV and SQV revealed loss of hydrophobic contacts between the mutant and inhibitors attributed to the mutations D30N, V32I, I47V, and I84V, which alter
alter the size, shape and charge of the inhibitor binding site (Agniswamy et al., 2012). A larger separation of 3 Å increase between the closest side chain atoms of residues 47 and 84 where the P2’ of SQV was bound expands the S2/S2’ subsites. The open-conformation flap of the unliganded PR20 has a separation of 6 Å between the tips of the two flaps. In another study, a multi drug resistant variant MDR769 with 10 mutations exhibits 2000-fold weaker inhibition by darunavir compared to wild type PR in agreement with the structural data of an expanded active site at S1/S1’ caused by V82I and I84V and also at S3/S3’ and the more separated flaps by ~12 Å compared the number of ~4 Å in the wild type enzyme (Logsdon et al., 2004, Martin et al., 2005). Meanwhile, the relationship between changes in flap closure and drug potency for PR variant MDR769 was investigated through the solution method using Double Electron-Electron Resonance (DEER) spectroscopy (de Vera et al., 2012). However, since the inactivating catalytic mutation D25N was introduced to prevent self-proteolysis in MDR769 variant, it may influence the expanded active site and the unusual separation of the flaps. Studies of multiple mutants demonstrated that multiple mutations coordinate in highly drug resistant mutants like PR20, and MDR769 by causing widely separated flaps and reducing interactions with peptide or inhibitors with expanded inhibitor binding subsites, which is considered as the drug resistance mechanism of multiple mutations, and thus provide the ideas for design of new inhibitors by adding interactions with flap residues to stabilize the variable flaps or Introducing larger P2/P2’ groups to fit better in enlarged S2/S2’ subsites of PR20.
Figure 1.10 Mutations in the protease gene associated with resistance to protease inhibitors. (Johnson et al., 2013)
Figure 1.11 Sites of drug resistance mutations. The mutations located in the active site cavity are colored in red while the flap mutations are in orange and the mutations distal from the active site are indicated in blue color.
1.1.8 Specific Aims and Brief Results of Studying HIV-1 Drug Resistant Mutants $PR_{L76V}$ and $PR_{P51}$

The specific aims of studying HIV-1 drug resistant mutants $PR_{L76V}$ and $PR_{P51}$ is to investigate the molecular basis of drug resistance and further assist in the design of new anti-viral drugs for AIDS therapy. $PR_{L76V}$ and $PR_{P51}$ are two drug resistant variants with a single mutation and multiple mutations, respectively, and thus, it is important to study the effect of a single mutation and multiple mutations in drug resistance. The L76V mutation is a drug resistant mutation selected, interestingly, with decreased susceptibility to LPV, DRV, APV, and IDV but similar or increased susceptibility to ATV, NFV, SQV or TPV (Nijhuis et al., 2009a, Mueller et al., 2004, Vermeiren et al., 2007, Tartaglia et al., 2009, Young et al., 2010). The single L76V mutant has displayed reduced viral replication capacity severely. Therefore, in my study of $PR_{L76V}$, the stability, activity, structures with inhibitors of $RP_{L76V}$, and the autocatalytic processing of the precursor with the single mutation L76V were explored in order to study drug resistance mechanism of this mutation (Louis et al., 2011b). DRV and SQV are chosen to co-crystallize with $PR_{L76V}$ due to the opposite effects in susceptibility that L76V mutation produced. The residue 76 is located in a distal region where the residues have no direct contacts with inhibitors or substrates (Figure 1.11). Kinetic assays of HIV-1 $PR_{L76V}$ revealed that $PR_{L76V}$ has similar catalytic activity as wild type PR. The experimental result from thermal, urea-induced denaturation and dimer dissociation indicate that $PR_{L76V}$ has lower stability than wild type PR. Crystal structures at high resolution of $PR_{L76V}$ with inhibitors, DRV and SQV show the loss of internal van der Waals contacts of Val76 compared to Leu76 in wild type PR within hydrophobic core consistent with lower stability, which is considered as a unique molecular mechanism due to absence of contact with inhibitor or other subunit. Analysis of crystal structures showed that $PR_{L76V}$ has lost two hydrogen bonds with DRV in agreement with lower susceptibility to DRV and improved interactions with SQV consistent with increased susceptibility and thus helps explain the opposite effect of the L76V mutation on PR susceptibility to DRV and SQV. The unique molecular mechanism
for drug resistance provides the insights into the design of new anti-viral drugs to combat drug resistance. L76V is also studied with new inhibitor GRL02031 designed based on DRV from structural and inhibition aspects (Chang et al., 2012). Furthermore, the L76V mutation produces a severe defect in autoprocessing. The detailed studies including experimental methods and results on PR<sub>L76V</sub> are described in Chapter 2 of this dissertation and (Louis et al., 2011b).

PR<sub>PS1</sub> is an example of drug resistant mutant that contains multiple mutations evolving to give high level resistance as generally happens when HIV replicates in the presence of drugs (Condra et al., 1999). It is of utility to explore the molecular basis of the multiple mutations and further aid the design of new anti-viral drugs. The highly DRV-resistant HIV-1 variant HIV-1<sub>MIX</sub><sup>PS1</sup> (viral population at passage 51) was obtained by the propagation in the laboratory of a mixture of 8 highly DRV-susceptible HIV-1 clinical isolates. The mixture contains 9 to 14 PI-resistant mutations in the presence of DRV and replicated well at the concentration of 5 μM DRV (Koh et al., 2010). The HIV-1<sub>MIX</sub><sup>PS1</sup> protease (PR<sub>PS1</sub>) has 14 amino acid substitutions (Koh et al., 2010). The studies of physical and biochemical properties of PR<sub>PS1</sub> showed that the affinity for DRV and SQV was about 7400-fold and 135-fold weaker, respectively, than the corresponding values for wild-type PR. The autoprocessing of the TFR-PR<sub>PS1</sub> precursor was uninhibited by DRV and marginally inhibited by SQV, at 150 μM PI concentration (Louis et al., 2011a). These properties of PR<sub>PS1</sub> are consistent with the high antiviral resistance to DRV measured for virus bearing this variant (>300-fold increased EC<sub>50</sub>) relative to wild type (Koh et al., 2010). In my study of PR<sub>PS1</sub>, crystallographic study was performed to investigate the DRV-bound PR<sub>PS1</sub> complex and unliganded PR<sub>PS1</sub> so as to understand the molecular basis of the high resistance to DRV of this mutant and how the multiple mutations influence the overall structure and changes in specific region of PR<sub>PS1</sub> and further aid the design of an anti-viral drug to target multiple drug resistant mutants. Two crystal structures of PR<sub>PS1</sub> bearing the inactivating mutation D25N to avoid autoproteolysis were determined: a DRV bound structure and a ligand free structure. The crystal structure shows that DRV was bound in an unusual position com-
pared to the position of DRV in wild type PR within the active site of the protein parallel to dimer interface and also interacts with residues from two symmetry related dimers. The interactions between the variant and DRV mainly depend on van der Waals interaction and fewer hydrogen bonds compared with DRV bound in wild type PR. In both the unliganded PR$_{PS1}$ and DRV bound PR$_{PS1}$ structures, the flaps are more separated from each other than seen in wild type PR. The multiple mutations effect the structural changes in PR$_{PS1}$/DRV. The structural information suggests that the large separation of flaps may be a common mechanism for resistance to PIs and the unique binding site for DRV provides the hint for designing novel types of antiviral inhibitors that capture the open, inactive conformation of the protease. The description of studies and findings on the crystal structures PR$_{PS1}$ is provided in Chapter 3 of this dissertation.
1.2 Design of Green Fluorescent Protein Based Calcium Sensor

1.2.1 The Green Fluorescent Protein

The green fluorescent protein (GFP) is a chemiluminescent protein first isolated from *Aequorea* jellyfish (Shimomura *et al*., 1962). The absorbance spectrum and fluorescence quantum yield of GFP were measured and the result showed that aequorin could transfer its luminescence energy to GFP (Morise *et al*., 1974). Chromophore is the luminescence generator and is centrally located in the molecule of GFP (Shimomura *et al*., 1979). The chemical name of the chromophore is 4-(p-hydroxybenzylidene)imidazolidin-5-one and the chemical structure is shown in Figure 1.12. The fact that the GFP gene can be cloned and expressed in other organisms and shows absorbance changes demonstrated that the posttranslational synthesis of the chromophore is not jellfish-specific (Prasher *et al*., 1992, Inouye *et al*., 1994). *Aequorea* GFP and *Renilla* GFP have been biochemically well characterized. The differences are that *Renilla* GFP has a higher extinction coefficient, and is more likely to dimerize and is more resistant to pH-induced conformational changes and denaturation than *Aequorea* GFP (Ward *et al*., 1982). However, the *Aequorea* GFP genes are the only GFP genes that have been cloned successfully (Tsien, 1998).

There are at least five currently known isoforms of the gene sequence of *Aequorea* GFP and the gene has been introduced with altered codons or sequences so as to improve the translational initiation (Tsien, 1998). The chromophore is formed from residues Ser65-Tyr66-Gly67 in the protein through the three steps of folding-cyclization-dehydration in the currently accepted mechanism (Figure 1.12) (Heim *et al*., 1994, Cubitt *et al*., 1995, Reid *et al*., 1997). First, the folding of GFP leads to a nearly native conformation, then the amide of Gly67 as nucleophile attacks the carbonyl of residue 65 to form the imidazolinone, followed by dehydration. Finally, the α-β bond of residue 66 is dehydrogenated resulting in the conjugation of its aromatic group with the imidazolinone. The three dimensional structure of GFP, composed of 238 amino acids, was first solved in 1996 independently by Ormo (Ormo *et al*., 1996), with PDB
accession number 1EMA, and by Yang (Yang et al., 1996), with PDB ID 1GFL. The secondary structure of GFP has 11 β-strands that create a cylindrical beta-can structure and one α-helix running up the axis of the cylinder. The chromophore is deeply buried in the center of the hydrophobic core of GFP.

Based on the component of the chromophore derived from residue 66, GFP can be classified into seven classes: class 1, a mixture of anionic phenolate and neutral phenol; class 2, the phenolate anion; class 3, neutral phenol; class 4, phenolate anion with stacked π-electron system; class 5, indole; class 6, imidazole; and class 7, phenyl (Tsien, 1998) (Figure 1.13). The chromophore of classes 1-4 contains Tyr at position 66, while the ones of classes 5-7 include Trp, His, and Phe at that position, respectively (Table 1.2). Class 1 is the wild type GFP with a major excitation peak at 395 nm and a minor peak at 475 nm, where the amplitude of the major peak is three times higher than the minor one. Excitation at 395 nm leads to an emission maximum at 508 nm, whereas excitation at 475 nm gives a peak at 503 nm (Heim et al., 1994). The generally accepted interpretation is that the protonated or neutral chromophore is responsible for the 395-nm peak and the deprotonated or anionic chromophore is responsible for the 475-nm peak (Heim et al., 1994, Cubitt et al., 1995). Since the coexistence of neutral and anion chromophores produces two excitation peaks in the spectrum, wild type GFP has many disadvantages. GFPs with phenolate anions in the chromophore have been widely used for cell biological applications because they combine the simple excitation and emission peaking with the high brightness. The significant mutation causing ionization of phenol is Ser65 to Thr65, and thus the 395-nm excitation peak is suppressed and the 470 to 475-nm peak is enhanced and shifted to 489-490 nm with five- to six fold increase in amplitude (Cubitt et al., 1995, Cheng et al., 1996). To pursue an obvious interest for protein expression at 37 °C, additional mutations like F64L and V163A have been commonly used to improve the folding efficiency and have little effect on the brightness (fluorescence intensity) of properly folded molecules (Cormack et al., 1996).
Crystal structures of wild type GFP and GFP with the substitution of S65T have revealed the probable mechanism for the preference of neutral or anionic chromophore (Yang et al., 1996, Ormo et al., 1996, Brejc et al., 1997) (Figure 1.13). Ser65 can serve as a hydrogen bond donor to the buried side chain of Glu222 to induce the ionization of the carboxylate group. The carboxylate group is located within 3.7 Å away from the chromophore and creates an anionic environment, and thus the electrostatic repulsion is unfavorable for the anionic state of the chromophore and favorable for the neutral state of the chromophore. However, when threonine replaces the serine at position 65, this larger side chain cannot provide the correct conformation in the crowded interior of protein to maintain the protonated form of the carboxylate group of Glu222 (Ormo et al., 1996, Brejc et al., 1997). Therefore, the ionization state of the Glu222 side chain is important for the existence of different forms of the chromophore. The effect of mutation E222G is the same as that of mutations of Ser65 on changes in the spectral shape and wavelengths (Ehrig et al., 1995). Thr203 is another one of the most essential residues. The main chain carbonyl oxygen of Thr203 can form a hydrogen bond interaction with the phenol of the chromphore in wild type GFP, and the side chain hydroxyl group can be hydrogen-bonded to the phenolic oxygen of the chromophore in both wild type GFP and the phenolate class of GFP such as EGFP (enhanced green fluorescent protein) with S65T and F64L. Mutation T203I in wild type GFP results in the lack of the 475 nm absorbance peak attributed to the anionic chromophore, which is due to the loss of a hydrogen bond between the threonine side chain and the phenolate (Heim et al., 1994, Ehrig et al., 1995). The replacement of Thr by Ile could prevent ionization of the chromophore.

Applications of GFPs in cell biology are mainly divided into tagging or labeling and indication of the change of its environment. As a tag, GFP can be used to detect gene expression when it is fused into the host protein (Cheng et al., 1996, Muldoon et al., 1997). For another use as a tag, GFP has been successfully applied to monitoring the localization and fate of host proteins without interfering in the function (Tsien, 1998). A series of fluorescent proteins with emission spectra peaks from UV to infrared
wavelengths has been created randomly or by direct mutagenesis (Shaner et al., 2004). GFP can act as an indicator in response to the change of the environment. Fluorescence resonance energy transfer (FRET) is a general way to explore the environment by using linked GFPs of different colors (Miyawaki et al., 1997). FRET is quantum-mechanical phenomenon that is produced when one of two fluorophores, as the donor, has an emission spectrum overlapping the excitation spectrum of the other fluorophore, the acceptor, and they less than 100 Å apart. The efficiency of FRET can be modulated by changes of the spatial relationship between the fluorophores. The change in ratio of acceptor to donor emissions is indicative of chemical signal. The first-demonstrated application of FRET between two GFPs is to monitor the protease action by inserting a protease-sensitive spacer between blue-emitting GFP mutant (i.e. BFP short for blue fluorescent protein) and a phenolate-containing GFP. Upon the addition of protease that can cleave the spacer and cut the linkage of the two GFPs, FRET is disrupted with increased ratio of blue to green emissions (Heim et al., 1996, Mitra et al., 1996). Currently, an interesting wide application of FRET is to monitor Ca$^{2+}$ by incorporating Ca$^{2+}$ sensitive region between two GFPs and this will be described in the following section of 1.2.3.
Figure 1.12 Proposed mechanism for the intramolecular biosynthesis of chromophore. (Cubitt et al., 1995)
<table>
<thead>
<tr>
<th>Mutation</th>
<th>Common name</th>
<th>Excitation wavelength</th>
<th>Emission wavelength</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class 1, wild-type None or Q80R F99S, M153T, V163A</td>
<td>Wild type Cycle 3</td>
<td>395-397/470-475 397/475</td>
<td>504 506</td>
</tr>
<tr>
<td>Class 2, phenolate anion S65T F64L, S65T</td>
<td>EGFP</td>
<td>489</td>
<td>509-511  507-509  511</td>
</tr>
<tr>
<td>F64L, S65T, V163A S65T, S72A, N149K, M153T, I167T</td>
<td>Emerald</td>
<td>487</td>
<td>509</td>
</tr>
<tr>
<td>Class 3, neutral phenol S202F, T203I T203I, S72A, Y145F</td>
<td>H9 H9-40</td>
<td>399 399</td>
<td>511 511</td>
</tr>
<tr>
<td>Class 4, phenolate anion with π stacked electron system S65G, S72A, T203F S65G, S72A, T203H</td>
<td>10C Q69K 10C</td>
<td>512 514</td>
<td>522 527</td>
</tr>
<tr>
<td>Class 5, indole in chromophore (cyan fluorescence proteins) Y66W Y66W, N146I, M153T, V163A</td>
<td>W7 W1B or ECFP</td>
<td>436 434/452 434/452</td>
<td>485 476/505 476/505</td>
</tr>
<tr>
<td>Class 6, imidazole in chromophore (blue fluorescence proteins) Y66H Y66H, Y145F</td>
<td>BFP P4-3</td>
<td>384 382</td>
<td>448 446</td>
</tr>
<tr>
<td>F64L, Y66H, Y145F</td>
<td>EBFP</td>
<td>380-383</td>
<td>440-447</td>
</tr>
<tr>
<td>Class 7, phenyl in chromophore Y66F</td>
<td></td>
<td>360</td>
<td>442</td>
</tr>
</tbody>
</table>
Figure 1.13 Fluorescence excitation and emission spectra and the chromophore structures. (a) wild-type, (b) Emerald, (c) H9-40, (d) Topaz, (e) W1B, and (f) P4-3 as described in Table 1.2; solid and dashed lines represent excitation and emission, respectively; spectra have been normalized to a maximum amplitude of 1. (Tsien, 1998)
Figure 1.14 Schemes of the interactions between the chromophore and the neighboring residues and water molecules in wild type GFP (a) and S65T mutant (b). (Brejc et al., 1997)
1.2.2 Role of Calcium in Biological System

Calcium acts as a second messenger in regulating many biological functions, such as heart beat, neural function, muscle contraction and cell proliferation and apoptosis (Berridge, 1998). Calcium is a ubiquitous signaling molecule interacting with calcium binding proteins, and it can flux between intracellular and extracellular space and subcellular compartments (Berridge et al., 2003). The intracellular concentration of calcium ranges from submicromolar to millimolar levels and the time scale of cytosolic calcium transients varies from microseconds for neuron transmitter release, 0.1 milliseconds for muscle contraction to days for cell development. It has been reported that the periodic change of calcium concentration can regulate the cardiac relaxation and contraction in cardiomyocytes (Dibb et al., 2007) and heart failure can be caused by the abnormal calcium signaling (Kawase et al., 2008). Upon binding to Ca$^{2+}$, A variety of calcium binding proteins, such as calmodulin (CaM), troponin C (TnC) and other ion channels will be induced with conformational change on binding calcium, and can further influence numerous cellular processes (Nelson et al., 1998).

The endo/sarcoplasmic reticulum (ER/SR) lumen is the main store for >90% of intracellular Ca$^{2+}$ and is critical in controlling Ca$^{2+}$ signaling, although it occupies less than 10% of the total cell volume (Hogan et al., 2010, Clapham, 2007, Berridge, 2007). The ER/SR Ca$^{2+}$ store is pivotal for the regulation of Ca$^{2+}$ signaling and maintenance of Ca$^{2+}$ gradients. The fluxing of Ca$^{2+}$ across the cellular compartments is regulated by ion channels, receptors and Ca$^{2+}$ buffer proteins. The Ca$^{2+}$ receptors and pumps can be activated by the agonists such as ATP, ionomycin, histamine, and glutamine, resulting in the ER Ca$^{2+}$ release from ER into cytosol (Bers, 2003, Galione et al., 2002, Giovannucci et al., 2002). Changes of the free [Ca$^{2+}$] in the cytosol modulate the activity of these channels and receptors (Babu et al., 1989). Sarcoplasmic Reticulum Ca$^{2+}$ release is very important in regulating skeletal muscle contraction (Ebashi, 1961, Ebashi et al., 1962). Therefore, it is crucial to study Ca$^{2+}$ signaling in ER/SR for their importance in Ca$^{2+}$ related biological functions.
Figure 1.15 The calcium pathways in the cell. (Courtesy of Dr. Shen Tang from Dr. Yubin Zhou)
1.2.3 Calcium Measurement

Due to the importance of Ca\(^{2+}\) in biological systems, the attempt to measure \([\text{Ca}^{2+}]\) has continued since 1920. Currently, numerous techniques or methods have been developed for analyzing cellular or subcellular Ca\(^{2+}\) activity. Currently used calcium sensors are mainly divided into two classes: chemical fluorescent indicators generated by organic synthesis based on BAPTA (1,2-bis(o-aminophenoxy)ethane-N,N,N',N'-tetraacetic acid) and bioluminescent calcium indicators containing photoproteins (Tsien et al., 1985, McCombs et al., 2008). BAPTA is the first organically synthesized small molecule designed on the basis of EGTA (ethylene glycol tetraacetic acid), which exhibits high metal selectivity for Ca\(^{2+}\) (Tsien, 1980). BAPTA became the essential component of several calcium dyes such as Indo 1, Fluo 3 and Oregon green BAPTA (Gryniewicz et al., 1985, Minta et al., 1989). The chemical fluorescent indicators can be excited at different wavelength ranges: UV wavelength excitation, such as Indo 1, Fura 2 and their derivatives (Gryniewicz et al., 1985, Naraghi, 1997, Etter et al., 1994, Etter et al., 1996), and visible wavelength excitation, such as Fluo 3, calcium green, dextran conjugates (Gryniewicz et al., 1985, Minta et al., 1989). Even though the UV-exitable Ca\(^{2+}\) indicators are still used due to their quantitative ratiometric property, they are known to be more cytotoxic (Brakenhoff et al., 1996). Therefore, visible wavelength Ca\(^{2+}\) sensors have many advantages over UV-based Ca\(^{2+}\) indicators like less cytotoxicity, the ability to monitor change of \([\text{Ca}^{2+}]\) with UV-sensitive compounds and emitting light within the region of the electromagnetic spectrum with less severe background scattering (Takahashi et al., 1999). The majority of these indicators can produce greater dynamic range of signals like fluo 3 that can undergo 40-200 fold increase in fluorescence upon binding Ca\(^{2+}\) (Minta et al., 1989, Harkins et al., 1993). However, the big disadvantage of these synthesized Ca\(^{2+}\) sensors is their non-specific targeting of membrane and easy leakage from cell (McCombs et al., 2008). In addition, Ca\(^{2+}\) in denser or thicker tissues cannot be detected due to the difficulty in loading these Ca\(^{2+}\) dyes and also, the emitted light signal is attenuated in denser material.
Bioluminescent Ca\(^{2+}\) indicators mainly include two types of Ca\(^{2+}\) sensors: Ca\(^{2+}\)-binding photoproteins and GFP-based indicators. Ca\(^{2+}\)-binding photoproteins, such as aequorin and obelin etc, offer simplicity for instrumentation by emitting visible bioluminescence through an intramolecular reaction upon binding Ca\(^{2+}\) (Shimomura, 1984, Campbell, 1974). Aequorin is the most widely used bioluminescent Ca\(^{2+}\) sensor which is composed of three parts: apoaequorin protein of molecular mass of 21 kDa, the luminophore coeleterazine and molecular oxygen (Inouye et al., 1985, Inouye et al., 1989, Shimomura et al., 1988). The aequorin contains three Ca\(^{2+}\)-binding sites and produces emission with Ca\(^{2+}\) binding to at least two sites (Shimomura et al., 1963). Upon binding Ca\(^{2+}\) ions, the molecular oxygen in aequorin is released and the luminophore coeleterazine is oxidized to coelenteramide, emitting blue light (465 nm) which increases as [Ca\(^{2+}\)] increases between 10\(^{-7}\) and 10\(^{-4}\) M (Shimomura et al., 1963). However, its inability to get into organelles limits the distribution of aequorin in cell (Shimomura et al., 1962). Therefore, recombinant aequorins have become quite useful probes for calcium due to their ability to enter into various kinds of organelles without interfering with the physiological condition (Rizzuto et al., 1992).

Obelin is a Ca\(^{2+}\)-activated photoprotein that binds at least three molecules of Ca\(^{2+}\) to emit bioluminescence (Campbell, 1974). Compared to aequorin, obelin has faster onset of the bioluminescence in response to binding Ca\(^{2+}\) (3 ms by obelin vs. 10 ms by aequorin) but less [Ca\(^{2+}\)] sensitivity (Moisescu et al., 1975). The GFP-based Ca\(^{2+}\) biosensor is a promising indicator that exhibits high specificity for location and provides high accuracy for the measurement of subcellular Ca\(^{2+}\) signaling (Inouye et al., 1994). The basic strategy for the design of GFP-based Ca\(^{2+}\) biosensor is the insertion or graft of a Ca\(^{2+}\) sensitive protein such as calmodulin, or Ca\(^{2+}\) binding proteins like troponin C, into different mutated types of GFP to create various kinds of Ca\(^{2+}\) indicators. Therefore, fluorescence changes provide the information about [Ca\(^{2+}\)], which is produced by direct alteration in GFP or an increase in fluorescence resonance energy transfer (FRET). Cameleons, as one kind of the successfully synthesized Ca\(^{2+}\) indicators, consist of two GFP mutants with overlapping excitation/emission spectra and the Ca\(^{2+}\)-sensitive protein calmodulin.
with M13, which is a calmodulin-binding peptide with 26-residues from myosin light-chain kinase (Miyawaki et al., 1997). The calmodulin-M13 complex linked the two GFP mutants. When Ca\textsuperscript{2+} binds to calmodulin, it induces the conformational change of the complex, resulting in decreased distance between the two GPF mutants accompanied by an increase in FRET. Two combinations of donor and acceptor GFP mutants have been designed: blue fluorescent protein (BFP)-GFP and cyan fluorescent protein (CFP)-yellow fluorescent protein (YFP) (Tsien, 1998). Troponin C is another Ca\textsuperscript{2+} binding protein functioning only in muscle contraction (Heim et al., 2004). Troponin C is inserted between CFP and citrine that is a yellow fluorescent protein derivative, and this forms troponeons that is another type of FRET-based Ca\textsuperscript{2+} indicators. Compared to cameleons, troponeons exhibit better performance in targeting to a specific subcellular domains (Takahashi et al., 1999). GFP-based Ca\textsuperscript{2+} indicators have been widely used for its many advantages over the chemical fluorescent probes, such as the ratiometric measurement of [Ca\textsuperscript{2+}], brighter fluorescence, high sensitivity, precise expression in targeted intracellular compartments and so forth, although they demonstrated fewer changes than most chemical fluorescent Ca\textsuperscript{2+} indicators and pH-sensitivity seen in some GFP variants as the disadvantages (Takahashi et al., 1999, Kneen et al., 1998).

Yang’s lab has been devoted to exploring site-specific Ca\textsuperscript{2+} binding affinity of designed calcium sensors (Figure 1.15). They have successfully created a Ca\textsuperscript{2+} sensor by grafting an EF-hand motif with Ca\textsuperscript{2+} binding site into EGFP (enhanced green fluorescent protein) that contains F64L and S65T (Zou et al., 2007). The grafted Ca\textsuperscript{2+} sensor (G1) emitted a dual 510 nm fluorescence intensity ratio metric change when excited at 398 nm and 490 nm wavelengths. The fluorescent emission ratio (measured at 510 nm) for 385 nm to 480 nm excitation is indicative of [Ca\textsuperscript{2+}]. However, the dynamic range is small with only 10-15% change observed in mammalian cell imaging (Zou et al., 2007). The hypothesis for the design of EGFP-based Ca\textsuperscript{2+} sensor is that alteration of chromophore is associated with the conformational change induced binding of Ca\textsuperscript{2+}. It has been reported that the fluorescent sensitive sites are on the surface of
EGFP and site directed mutations at that area were designed to bind small molecules such as the GFP-based zinc sensors (Kikuchi et al., 2004). Therefore, a new Ca$^{2+}$ binding pocket was designed via site directed mutagenesis on the surface of EGFP. A series of different combinations of amino acid substitutions with negatively charged side chains that can coordinate with Ca$^{2+}$ was designed to (Tang et al., 2011). This de novo design of Ca$^{2+}$ sensors can provide theoretical support for developing GFP-based biosensors for diverse molecules by the means of site-directed mutagenesis.
Figure 1.16 Scheme of EGFP-based Ca$^{2+}$ biosensor. EGFP is used as a scaffold protein and EF-hand III of calmodulin was grafted into the fluorescent sensitive location of EGFP to produce Ca-G1; negatively charged amino acids were introduced on the surface of three antiparallel beta sheets of EGFP forming a pentagon to bind Ca$^{2+}$. 
1.2.4 Specific Aims and Brief Results of Studying GFP-based Calcium Sensor CatchER

Due to the need of efficient Ca\(^{2+}\) sensors to target cellular compartments with a high Ca\(^{2+}\) concentration, such as the ER/SR, a novel class of genetically encoded indicators was developed by designing a Ca\(^{2+}\) binding site in the enhanced green fluorescent protein (EGFP). CatchER (Calcium sensor for detecting high concentration in the ER) shows Ca\(^{2+}\) release kinetics with a fast off-rate estimated at around 700 s\(^{-1}\) and dissociation constant \(K_d\) around 1 mM of Ca\(^{2+}\) binding affinity (Tang et al., 2011). In my study of CatchER, crystallographic study was performed of CatchER in the absence of Ca\(^{2+}\) (apo form) and complexed with Ca\(^{2+}\) to identify where the calcium ion bound and to reveal the mechanisms of fluorescent change triggered by calcium (Zhang et al., 2013b). Analysis of these structures will help to further optimize the design of current calcium biosensors. In addition, to assist in locating the bound Ca\(^{2+}\) due to the weak Ca\(^{2+}\) binding affinity, high off-rate and difficulty in distinguishing calcium from water in the crystal structure, the heavy metal ion Gd\(^{3+}\) with similar metal binding coordination properties to calcium was soaked into apo CatchER crystals. The crystal structure of CatchER with Gd\(^{3+}\) ion demonstrated clearly the position of the bound metal ion. Analysis of these three structures shows that T203 has lost its direct hydrogen bond with the anionic form of chromophore and gained a water-mediated hydrogen bond with the chemically exchanged neutral chromophore in all three structures. Two partially occupied Ca\(^{2+}\) or Gd\(^{3+}\) ions were observed and refined within the designed calcium-binding site. In the crystal structures of CatchER with Ca\(^{2+}\) and Gd\(^{3+}\), the side chain of E222 has two alternative conformations and regains water-mediated hydrogen bonding interaction with Gln69 as in EGFP and main chain of residues from 202 to 206 has maximum shift of 0.8 Å. The binding of the metal ion induced structural changes of T203 and E222 around the chromophore corresponding to a concurrent increased intensity in absorbance around 490 nm (increase of deprotonated chromophore) and a decreased absorption intensity around 398 nm (decrease of protonated chromophore state) with the addition of Ca\(^{2+}\) or Gd\(^{3+}\). This structural information is of utility for optimizing the sensor. The detailed studies including experimental
methods and results on CatchER are described in Chapter 4 of this dissertation and Zhang et al., 2013 (Zhang et al., 2013b).
2 THE L76V DRUG RESISTANCE MUTATION DECREASES THE DIMER STABILITY AND RATE OF AUTOPROCESSING OF HIV-1 PROTEASE BY REDUCING INTERNAL HYDROPHOBIC CONTACTS


2.1 Abstract

The mature HIV-1 protease (PR) bearing drug-resistance mutation L76V (PR_{L76V}) is significantly less stable, with >7-fold higher dimer dissociation constant ($K_d$) of 71 ± 24 nM and twice the sensitivity to urea denaturation (UC$_{50}$ = 0.85 M) relative to PR. Differential scanning calorimetry showed a decrease in $T_m$ of 12 °C for PR$_{L76V}$ in the absence of inhibitors, and 5-7 °C in the presence of inhibitors darunavir (DRV), saquinavir (SQV) and lopinavir (LPV), relative to PR. Isothermal titration calorimetry gave a ligand dissociation constant of 0.8 nM for DRV, ~160-fold larger than that of PR, consistent with DRV resistance. Crystal structures of PR$_{L76V}$ complexed with DRV and SQV were determined at resolutions of 1.45-1.46 Å. Compared to the corresponding PR complexes, the mutated Val76 lacks hydrophobic interactions with Asp30, Lys45, Ile47, and Thr74, and exhibits closer interactions with Val32 and Val56. The bound DRV lacks a hydrogen bond with the main chain of Asp30 in PR$_{L76V}$ relative to PR, possibly accounting for resistance to DRV. SQV shows slightly improved polar interactions with PR$_{L76V}$ compared to PR. Although the L76V mutation significantly slows the N-terminal autoprocessing of the precursor TFR-PR$_{L76V}$ to give rise to the mature PR$_{L76V}$, the co-selected M46I mutation counteracts by enhancing this rate but renders the TFR-PR$_{M46I/L76V}$ precursor less responsive to inhibition by 6 µM LPV while retaining inhibition by SQV and DRV. The correlation of lowered stability, higher $K_d$ and impaired autoprocessing, with reduced internal hydrophobic contacts suggests a novel molecular mechanism for drug resistance.
2.2 Introduction

Human immunodeficiency HIV-1 protease (PR) is a major drug target for treatment of AIDS. PR functions in the last step of the HIV-1 life cycle by catalyzing the cleavage of viral polyproteins produced in the host cell (Turner et al., 1999). Hydrolysis of polyproteins into functional products is important for the maturation and production of infectious progeny virions. Protease inhibitors (PIs) act to prevent the maturation process. Currently, nine PIs are approved by the FDA as anti-viral drugs. However, the lack of a proofreading step of the viral reverse transcription leads to a high frequency of mutations, and under drug pressure, rapid selection of a combination of mutations confers drug resistance thus presenting a severe challenge in current anti-HIV treatment (Emini, 2002).

PR is a homodimer with 99 amino acids in each subunit (Figure 2.1) (Louis et al., 2000). The Asp25-Thr26-Gly27 triplets of both monomers form the catalytic site, where Asp25 and Asp25’ (the prime indicates the second subunit of the dimer) serve as general acid-base catalysts of polyprotein cleavage in the proposed reaction mechanism (Brik et al., 2003). Two flexible interacting flaps are formed by residues 44-57 of both subunits. The two characteristic triplets and two flaps make important contributions to the active site cavity for the binding of substrates or inhibitors (Weber et al., 2009a). The PR dimer is stabilized via noncovalent interactions of the residues at the dimer interface. Important intersubunit hydrogen bonds connect the catalytic triplets, flaps and beta sheet formed by the 4 terminal strands (Gustchina et al., 1990). The tertiary fold of each monomer is stabilized by hydrophobic interactions among the aliphatic residues in an internal hydrophobic core. Mutations occurring in over 30 of the 99 residues in each subunit are associated with drug resistance (Johnson et al., 2008). Two major types of PI-resistance mutations are proposed to influence the PR activity (Weber et al., 2009a). One type located near the active site can change the binding affinity and/or specificity of PR with the inhibitor by altering direct interactions (Tie et al., 2007). Other mutations do not change the binding cavity...
directly; however, they may affect PR stability and indirectly influence the binding of inhibitors via long-
range structural perturbations (Liu et al., 2005).

The L76V mutation has become more prevalent in datasets of HIV-1 mutants observed in pa-
tients (Johnson et al., 2008). Presence of L76V as a single mutation was shown to hamper viral replication severely (Nijhuis et al., 2009a). L76V, generally accompanied by other mutations, is considered a major mutation providing 2-6-fold decreased susceptibility to darunavir (DRV), fosamprenavir, indinavir (IDV), and lopinavir (LPV) (Young et al., 2010, Rhee et al., 2010). It is also selected as a drug resistance mutation in patients receiving LPV therapy (Nijhuis et al., 2009a). This mutation, however, is not associ-
ated with resistance to atazanavir (ATV), nelfinavir (NFV), saquinavir (SQV) or tipranavir (TPV). Instead, mutants containing L76V exhibit increased susceptibility to these drugs (Mueller et al., 2004, Vermeiren et al., 2007, Young et al., 2010). Hence, this mutation shows opposing roles in drug resistance, acting to increase susceptibility to some drugs and decrease susceptibility to others (Tartaglia et al., 2009).

We constructed PR with the single mutation of L76V in order to investigate its effect on the structure, stability and activity of the enzyme. Crystal structures of PR_{L76V} with DRV and SQV were de-
termined to investigate the molecular basis for the responses of this mutant to inhibitors. Leu76 is lo-
cated in the hydrophobic core of PR close to the active site cavity, although its side chain has no van der Waals contacts with most substrates or inhibitors. Additionally, the effect of the mutation on the auto-
catalytic processing of the precursor, which is required for onset of the catalytic activity characteristic of the mature, dimeric enzyme, was assessed \textit{in vitro}, and the effect of a second mutation, M46I, was ex-
amined in the precursor containing both mutations.
2.3 Experimental Procedures

2.3.1 Protein Expression and Purification

The HIV-1 PR (Genbank HIVHXB2CG) clone optimized for structural and biochemical studies contains mutations Q7K, L33I and L63I to minimize autoproteolysis, and C67A and C95A to prevent cysteine-thiol oxidation (Louis et al., 1999a). The L76V mutation was introduced in this PR template as well as the template encoding a PR precursor mimetic in which the full-length transframe region (TFR) is fused to the N-terminus of PR (TFR-PR) by use of the appropriate oligonucleotide primers using the QuikChange protocol (Stratagene, La Jolla, CA) and verified by DNA sequencing. A second construct, containing two mutations, M46I and L76V, in the PR domain of TFR-PR was also made using TFR-PR_{L76V} as the template. Proteins were expressed using pET11a vector and E. coli BL21(DE3), purified and refolded using established protocols (Wondrak et al., 1996, Ishima et al., 2007, Sayer et al., 2010), and the identities of the products verified by ESI-MS.

2.3.2 Enzyme Assays

The kinetic parameters were measured using a fluorescence assay as described (Liu et al., 2008) with the substrate Abz-Thr-Ile-Nle-pNO₂Phe-Gln-Arg-NH₂ (where pNO₂Phe is para-nitro-phenylalanine, Nle is norleucine and Abz is anthranilic acid) (Bachem Bioscience Inc., King of Prussia, PA), which is based on the p2/NC cleavage site of the viral polyprotein. 10 μl PR_{L76V} at a final concentration of 34 nM from active site titration with SQV was mixed with 100 μl of reaction buffer [100 mM MES (4-morpholineethanesulfonic acid), pH 5.6, 400 mM NaCl, 1 mM ethylenediaminetetraacetic acid, and 5% glycerol] at 26 °C. The reaction was initiated by adding substrate from a stock solution of 186 μM to give a final concentration of 12-84 μM. PR activity was measured by monitoring the increase in fluorescence over 5 min using an excitation wavelength of 340 nm and emission wavelength of 420 nm (pNO₂Phe as quencher-fluorescent acceptor; Abz as fluorescent donor) with a POLARstar OPTIMA 96-well microplate
inertial (BMC Labtech). Values for $k_{\text{cat}}$ and $K_m$ were obtained by fitting the curves to the Michaelis-Menten equation using the program SigmaPlot (SPSS Inc., Chicago, IL).

For dimer dissociation and urea denaturation studies, enzyme activity was measured by following the initial rates of hydrolysis of the chromogenic peptide substrate IV that mimics the CA/p2 cleavage site (Lys-Ala-Arg-Val-Nle-pNO$_2$Phe-Glu-Ala-Nle-NH$_2$, California Peptide Research, Napa, CA) at 310 nm ($\Delta \varepsilon = 1797 \, \text{M}^{-1}\text{cm}^{-1}$) in 50 mM sodium acetate buffer, pH 5.0, in 120 mL cuvettes at 28 °C. Substrate concentration (390 μM in the assay mixtures) was determined from the UV spectrum of substrate stock solutions ($\varepsilon_{280} = 12,000 \, \text{M}^{-1} \text{cm}^{-1}$). The dimer dissociation constant, $K_d$, was determined by fitting the curve of activity (initial rate/protease concentration) vs. protease concentration (40-1200 nM as monomers) to an equation described previously (Sayer et al., 2010). For urea denaturation studies, activity was measured at urea concentrations of 0-3.5 M (Mahalingam et al., 1999) and a enzyme concentration of 0.3-1.15 μM. The UC$_{50}$ value is the urea concentration at which the PR activity is half of the maximum activity in the absence of urea.

2.3.3 Calorimetry

Inhibitor concentrations were determined kinetically by active site titration against the wild-type PR. Samples for differential scanning calorimetry (DSC) were prepared by the quench protocol from stock solutions in HCl as previously described (Ishima et al., 2007) to give a final enzyme concentration of 14-15 μM (as dimer) in 50 mM sodium acetate buffer, pH 5.0. Final concentrations of inhibitors DRV, SQV and LPV were 28-29 μM (~2-fold molar excess relative to dimeric enzyme). DSC scans were run on a MicroCal VP-DSC microcalorimeter (GE Healthcare) at 90 °C/h from a starting temperature of 20 °C and terminated at 70-90 °C depending on the position of the transition. Data were processed using the instrument’s Origin software as described (Sayer et al., 2008). SDS-PAGE of the inhibitor-free enzyme on a 20% homogeneous PhastGel (GE Healthcare) before and after DSC showed no evidence for autoproteolysis during the course of the DSC experiment.
For isothermal titration calorimetry, PR\textsubscript{L76V} (1.9 mg/ml in 12 mM HCl) was folded by the quench protocol (Ishima \textit{et al.}, 2007) to give final concentrations of 8-13.5 μM in 50 mM sodium acetate buffer, pH 5, and titrated with 16 injections of DRV in the same buffer (60-140 μM depending on the protein concentration) by use of a MicroCal iTC\textsubscript{200} microcalorimeter (GE Healthcare) at 28 °C. Stock 160-200 μM inhibitor solutions were prepared in low ionic strength buffers at a final concentration of 0.5% (v/v) DMSO, diluted from a stock solution of inhibitor in 100% DMSO. All ITC experiments were performed at 0.5% or lower DMSO concentration. DMSO equivalent to that in the titrant was added to the protein solution in the cell to minimize thermal effects of changing DMSO concentration during the titration. As reduced autoproteolysis is intrinsic to the optimized PRs after folding at pH 5, concentrations of active PR\textsubscript{L76V} were verified from the known stoichiometry of inhibitor binding to the active site (N = 1) on titration. Data were processed using the instrument’s Origin software. Three separate experiments at different concentrations of PR\textsubscript{L76V} and DRV were consistent with an upper limit of ~10 nM for the dissociation constant ($K_d$). A displacement titration with 50 μM DRV in the presence of 16 μM RPB inhibitor (H-Arg-Val-Leu-(r)-Phe-Glu-Ala-Nle-NH\textsubscript{2}; Bachem Americas, Inc., Torrance, CA) added to the sample cell containing 4 μM PR\textsubscript{L76V} provided a more accurate value. Analogous displacement titrations under the same conditions with SQV or LPV did not give adequate thermal effect for reliable quantitation.

\textbf{2.3.4 Autoprocessing of TFR-PR\textsubscript{L76V} Precursor}

Isolation of PR precursors in significant quantities presents a substantial challenge because the intrinsic autoprocessing of the precursors to mature protease results in their depletion during expression. However, amounts of the full length precursors sufficient for small scale experiments can be recovered. Thus, remnants of the unprocessed precursor TFR-PR\textsubscript{L76V} and TFR-PR\textsubscript{M46I/L76V} were purified as described (Louis \textit{et al.}, 1999a, Sayer \textit{et al.}, 2010). The precursor was folded by addition of 5.66 volumes of 5 mM sodium acetate buffer, pH 5.3, with or without added DRV, to TFR-PR\textsubscript{L76V} in 12 mM HCl, to give a final pH of 4.5. Samples (5 µl) were removed at times from 0-26 hours, mixed with 3 µl SDS-PAGE sam-

ple buffer and frozen immediately. Samples were subjected to electrophoresis on 20% homogeneous PhastGels (GE Healthcare), and visualized by staining with PhastGel Blue R.

2.3.5 Crystallographic Analysis

PR_{L76V} at a concentration of 7 mg·ml$^{-1}$ was mixed with the inhibitors DRV or SQV at 5-fold molar excess. Crystals were grown by the hanging-drop vapor-diffusion method at room temperature using 24 well VDX plates (Hampton Research, Aliso Viejo, CA). The crystallization drops had equal volumes of the protein and reservoir solutions. Crystals of PR_{L76V} with DRV grew in solutions of 30 mM NaOAc buffer (pH 4.6-5.0), 1.4-1.8 M NaCl and 3% (v/v) DMSO. Crystals of PR_{L76V} with SQV grew in solutions of 100 mM Tris-HCl buffer (pH 6.0-9.0) and 1.3 M NaCl with 3% (v/v) DMSO. The crystals were frozen in liquid nitrogen with the cryoprotectant of 20-30% (v/v) glycerol. X-ray diffraction data for the crystals were collected on the SER-CAT beamline of the Advanced Photon Source, Argonne National Laboratory in Chicago.

X-ray data were processed with HKL2000 (Otwinowski et al., 1997). The structures were solved by molecular replacement using MolRep in the CPP4i suite of programs (Vagin et al., 1997) and the starting model was the wild type PR complex with DRV (PDB code 2IEN) (Tie et al., 2004) in the same space group as the new structures. The structures were refined with SHELXL (Sheldrick et al., 1997) and refitted using Coot 0.3.3 (Emsley et al., 2004). Alternate conformations were modeled for PR residues, inhibitors and solvent molecules based on the observed electron density maps. The solvent was modeled with over 200 water molecules, ions and other solvent present in the crystallization solutions, as described (Tie et al., 2004). Anisotropic $B$ factors were applied for all the structures. Hydrogen atom positions were calculated in the last stage of refinement, using all data once all other parameters, including disorder, had been modeled. The mutant crystal structures were compared with the wild type PR by superimposing their $C_\alpha$ atoms as described (Tie et al., 2004). Structural figures were made using PyMol (DeLano, 2002).
2.3.6  **Protein Data Bank Entries**

The structure coordinates and factors have been deposited in the RCSB Protein Data Bank with access number 3PWM for PR<sub>L76V-DRV</sub> and 3PWR for PR<sub>L76V-SQV</sub>.

2.4  **Results and Discussion**

2.4.1  **Properties and Stability of Mature PR<sub>L76V</sub>**

Purified mature PR<sub>L76V</sub> showed a $K_m$ of 37 ± 6 µM and $k_{cat}$ of 334 ± 23 min<sup>-1</sup> at pH 5.6 and 200 mM NaCl. The catalytic efficiency, $k_{cat}/K_m$, of 9.0 mM<sup>-1</sup> min<sup>-1</sup> for hydrolysis of the fluorogenic substrate based on the p2/NC cleavage site is essentially the same (1.2-fold) as the value of 7.2 mM<sup>-1</sup> min<sup>-1</sup> determined for PR under the same conditions (Liu et al., 2008). A plot of the dependence of activity on urea concentration (Figure 2.2A) shows a transition midpoint ($UC_{50}$) of 0.85 M urea for PR<sub>L76V</sub>, which is approximately half the value of 1.78 M for PR (Table 2.1). Increased susceptibility to urea denaturation is consistent with the lower stability of the PR<sub>L76V</sub> dimer relative to the wild type enzyme as reflected by the $K_d$ of 71 ± 24 nM for the PR<sub>L76V</sub> (Figure 2.2B), which is at least 7-fold higher than that of PR (Ishima et al., 2007).

2.4.2  **Inhibitor Binding**

Differential scanning calorimetry (Figure 2.3A) was used to assess the thermal denaturation of PR<sub>L76V</sub> in the absence and presence of inhibitors. The observed $T_m$ values for PR<sub>L76V</sub> in Table 2.1 are compared with those previously reported for PR under the same conditions (Sayer et al., 2009). In the absence of inhibitors PR<sub>L76V</sub> is markedly less stable to thermal denaturation than is PR, with a $T_m$ that is 12 °C lower than that of PR. Although the mechanisms of heat- and urea-induced denaturations are not necessarily the same, the low $T_m$ value for PR<sub>L76V</sub> is consistent with the observed effect of urea on its catalytic activity, as both heat- and urea- denaturations involve changes in the stability of the protein fold, which may precede and/or accompany dimer dissociation. Even when complexed with inhibitors,
which should stabilize the proteases as dimers, the $T_m$ value for PR$_{L76V}$ is about 7 °C lower than that for PR. The decreased thermal stability of PR$_{L76V}$ contrasts with results reported for several other PR mutants associated with drug resistance. For example, $T_m$ values for ATV resistant mutant I50L/A71V (Yanchunas Jr et al., 2005) and multi-drug resistant mutant V82F/I84V (Todd et al., 2000) were both higher than the values observed for PR by 2.2 and 4 °C, respectively. A multi-drug resistant mutant bearing 11 mutations also exhibited a slightly higher $T_m$ than the wild type enzyme (Muzammil et al., 2007).

An upper limit of ~10 nM for the dissociation constant ($K_L$) of the PR$_{L76V}$-DRV complex, with a range between 6.6 and 18 nM, was obtained by direct ITC titrations with the inhibitor. A substrate-analog inhibitor (RPB) with a reduced peptide bond exhibits a good negative thermal response ($\Delta H = -4.3$ kcal/mol) on binding to PR$_{L76V}$ and a $K_L$ of $61 \pm 13$ nM (Figure 2.3B). A displacement titration with DRV in the presence of RPB gave a value of $0.79 \pm 0.29$ nM (Figure 2.3C), approximately 160–fold larger than $K_L$ for the PR-DRV complex (King et al., 2004, Brower et al., 2008). Lack of an adequate thermal response under the same conditions, either on direct titration or displacement of RPB (10-15 mM), precluded $K_L$ measurements with SQV and LPV. Alternatively, a qualitative indication of the relative binding affinity of different inhibitors to the same protein can be obtained from the increase in thermal denaturation temperature ($\Delta T_m$) in the presence of inhibitor (Sayer et al., 2009, Sayer et al., 2008, Yanchunas Jr et al., 2005). The values of $\Delta T_m$ for DRV, SQV and LPV binding to PR$_{L76V}$ are significantly larger (by 5-7 °C) than for PR, reflecting both the low thermal stability of the free PR$_{L76V}$ and significant stabilization of the protein-inhibitor complexes. $\Delta T_m$ values for DRV, SQV and LPV binding to the PR$_{L76V}$ (Figure 2.3A) differ by less than 2 °C in the order DRV>SQV>LPV are consistent with large binding constants (small $K_L$ values), although the magnitude of their enthalpy values for thermal denaturation decrease significantly in the same order. The effects of these three inhibitors on autoprocessing of the TFR-PR$_{L76V}$ precursor (cf. Figure 2.4 and following section) suggest that $K_L$ for SQV and LPV will be larger than the 0.8 nM observed by ITC for DRV. By contrast, in vitro phenotype assays found the mutation L76V in PR derived from clini-
cal samples to be associated with enhanced susceptibility to SQV and resistance to DRV and LPV (Vermeiren et al., 2007, Young et al., 2010). However, in many, if not most, DRV resistant clinical isolates, a single resistance mutation does not occur alone but is associated with other mutations such as M46I and L90M (Mueller et al., 2004, Mitsuya et al., 2007), which are selected along with L76V in clinical settings. It is possible that the relative responses to these inhibitors as well as the dimer stability may be altered by the presence of such additional mutations.

2.4.3 Autoprocessing of PR_{L76V} and PR_{M46I/L76V} Precursors

During viral replication, the autocatalytic processing (autoprocessing) of PR from its Gag-Pol polyprotein precursor is essential to generate the active mature PR and requisite structural and functional proteins of the infective virion (Louis et al., 2000). Cleavage between the transframe region (TFR), encoded in the Pol open reading frame, and the N terminus of PR domain (TFR/PR site) is crucial for the formation of stable PR dimers with full catalytic activity from the monomeric TFR-PR. Consequently, compromising this process will adversely affect the viability of the virus. Previous studies with the wild type precursor, TFR-PR, have identified the pathways for its maturation (Louis et al., 2000, Louis et al., 2007). The time course of autoprocessing in the absence of inhibitors and in the presence of LPV, SQV and DRV is shown in Figure 2.4. At low pH an initial cleavage of TFR-PR (a) occurs between F8/L9 of the TFR to give an intermediate (b) with low catalytic activity similar to the full-length precursor. Subsequent cleavage at the TFR/PR site releases the fully active, mature PR (c) and the fragment TFR^{0-56} (d). Autoprocessing was assessed in vitro with a precursor comprising the 56-amino acid TFR fused to the N-terminus of the PR domain containing the L76V mutation (TFR-PR_{L76V}). Reactions conducted at ~6 μM protein concentration permit monitoring the autoprocessing of the precursor in small volumes suitable for SDS-PAGE on PhastGels. In the absence of inhibitors, only ~50% of the autoprocessing reaction is complete at 6 h (Figure 2.4A), in contrast to that of wild type TFR-PR, which is complete in < 1 h (Louis et al., 2000, Louis et al., 1999b). Thus, the slower rate of autoprocessing of TFR-PR_{L76V} seems to correlate
with the lower stability of the mature PR_{L76V}, namely the higher $K_d$ and increased susceptibility to denaturation by urea. Processing was only partially inhibited by LPV (Figure 2.4B), but completely inhibited by DRV (Figure 2.4D), using both inhibitors at a concentration of 6 µM. SQV is intermediate in its inhibition of TFR-PR_{L76V} autoprocessing being slightly better than LPV and poorer than DRV. In the presence of LPV and SQV some cleavage occurred at F8/L9 to give the intermediate TFR_{9-56}-PR_{L76V}, but very little or no subsequent cleavage was detected at the N-terminus of the protease domain to give mature PR_{L76V}.

The single mutation L76V severely compromises viral replication in cell cultures, whereas coexistence of another mutation, M46I, was found to increase resistance to LPV while also partially restoring the ability of the virus to replicate (Nijhuis et al., 2009a). This presumably occurs by increasing the stability and/or catalytic efficiency of the protease or its precursor. Unfortunately, it was not possible to isolate the mature PR_{M46I/L76V} because of its rapid autoproteolysis and very poor accumulation during its expression. However, limited quantities of the precursor TFR-PR_{M46I/L76V} could be obtained for comparison with TFR-PR_{L76V} under the same conditions. In the absence of inhibitors, TFR-PR_{M46I/L76V} undergoes autoprocessing significantly faster (Figure 2.4E) than TFR-PR_{L76V}. The additional M46I mutation has no effect on inhibition by 6 µM SQV or DRV (Figure 2.4G and 2.4H). Interestingly, however, TFR-PR_{M46I/L76V} almost completely evades inhibition (Figure 2.4F) under conditions where LPV inhibits processing of the single mutant TFR-PR_{L76V}. These results are consistent with the observations of Nijhuis et al. (Nijhuis et al., 2009a), and underscore the potential importance of secondary mutations for precursor processing, both by improving the intrinsic autoprocessing activity of an otherwise compromised precursor, and by decreasing its susceptibility to inhibition.

2.4.4 PR_{L76V}-Inhibitor Crystal Structures

Crystal structures of PR_{L76V} in complexes with DRV and SQV were determined to identify any structural changes caused by the mutation. The crystallographic data collection and refinement statistics
are displayed in Table 2.2. The crystal structures of PR<sub>L76V</sub>-DRV and PR<sub>L76V</sub>-SQV were refined to R-factors of 0.14 at the resolutions of 1.45 and 1.46 Å. The crystal structures had one PR dimer with residues labeled 1-99 and 1'-99' in each crystallographic asymmetric unit in the space group of P2<sub>1</sub>2<sub>1</sub>2<sub>1</sub>. DRV and SQV were observed in two alternate conformations in the active site cavity of the dimer of PR<sub>L76V</sub> with relative occupancies of 0.63/0.37 and 0.73/0.27, respectively. Alternate conformations were modeled for 10 and 17 residues in the PR<sub>L76V</sub>-SQV and PR<sub>L76V</sub>-DRV structures, respectively. The flap residues 46, 50 and 51 exhibited alternative conformations in both subunits of the two structures. The DRV complex also showed alternative conformations of the side chains for Val82, Pro81', Val82' and I84' in the inhibitor binding site; however, these residues showed a single conformation in the SQV complex. The structures were refined with more than 200 water molecules. The solvent included one sodium ion, two chloride and 3 acetate molecules in the DRV complex, while the SQV complex showed 8 glycerol molecules.

2.4.5 Structural Changes at the Site of Mutation

The PR<sub>L76V</sub> complexes with DRV and SQV were compared with the corresponding PR-inhibitor complexes. The PR-DRV (2IEN) (Tie et al., 2004) structure was solved in the same space group P2<sub>1</sub>2<sub>1</sub>2<sub>1</sub> at 1.30 Å resolution. The DRV complexes shared very similar backbone conformations with a low RMSD of 0.11 Å on Cα atoms. The PR-SQV (2NMW) (Tie et al., 2007) structure was refined at 1.16 Å resolution in the same space group, although the unit cell dimensions differed. Hence, the SQV complexes with PR<sub>L76V</sub> and PR superimposed with the larger RMSD value of 0.65 Å resulting from variations of up to 2.3 Å in residues 37-41 and 36'-45', as typically seen for PR complexes in different space groups.

Leu76 lies in the inner hydrophobic cluster in each subunit of the PR dimer (Ishima et al., 2001), and its side chain makes hydrophobic contacts with the side chains of Asp30, Val32, Lys45, Ile47, Val56, Gln58 and Thr74. The mutation of leucine 76 to valine gives a shorter side chain, which results in the loss of several van der Waals contacts. In the PR<sub>L76V</sub>-DRV complex, hydrophobic contacts with the side chains of Asp30, Lys45, and Thr74 and one contact with Ile47 are lacking in both subunits, as shown by an in-
teratomic distance of more than 4.2 Å (Figure 2.5A). Instead, the mutated Val76 forms more and closer hydrophobic interactions with the side chain of Val32, with interatomic distances of 3.8-4.0 Å rather than the longer 4.2 Å separation seen in the PR structure. Similar changes are observed in the PR<sub>L76V</sub>-SQV complex, with the exception of contacts with Val32 and Val56. The side chains of Val56 and residue 76 exhibit multiple hydrophobic contacts in the PR<sub>L76V</sub> complexes with both inhibitors and in the PR-DRV structure; however, the PR-SQV structure shows only a single van der Waals interaction between these residues (Figure 2.5B and C). Similarly, the side chains of Leu76 and Val32 show no hydrophobic contacts in PR-SQV, while multiple contacts are seen in the PR<sub>L76V</sub>-SQV complex and both DRV complexes. Thus, PR-SQV shows fewer internal contacts around Leu76 compared to those of Val76 in the mutant complex, whereas similar contacts of residues 32, 56 and 76 are maintained in both DRV complexes.

The decrease in internal van der Waals interactions is correlated with the lower stability of the mutant relative to the wild type enzyme described in the previous section. In contrast to other mutants exhibiting lower stability such as PR<sub>L24V</sub>, PR<sub>F53L</sub> and PR<sub>I50V</sub> (Liu et al., 2005, Liu et al., 2006), however, no significant changes were seen at the dimer interface of the mutant structures. Therefore, the major molecular change associated with the significantly lower stability of PR<sub>L76V</sub> is the loss of internal hydrophobic contacts characteristic of Leu76. This loss of hydrophobic contacts at the mutated residue stands in strong contrast to the minimal changes reported for the majority of single mutants (Liu et al., 2005, Shen et al., 2010). For example, our recent analysis of 6 mutants with single substitutions of residues in the hydrophobic clusters showed small structural adjustments that tended to conserve the hydrophobic interactions (Shen et al., 2010). Hence, the major loss of internal hydrophobic contacts for Val76 in PR<sub>L76V</sub>, coupled with reduced stability that is offset by secondary mutations, exemplifies a distinct mechanism for drug resistance.
2.4.6 Protease-Inhibitor Interactions

In the crystal structure of PR_{L76V}-DRV, the inhibitor is observed in two orientations bound within the active site cavity by a set of hydrogen bonds, which are similar to those in the PR-DRV structure (Tie et al., 2004). Both orientations of DRV show similar hydrogen bond interactions with PR_{L76V}. DRV formed 4 hydrogen bonds with the main chain atoms of Gly27’, Asp29’, and Asp30’, and 5 hydrogen bonds with the side chain atoms of Asp25/Asp25’ and Asp30 of PR_{L76V} (Figure 2.6A). The water-mediated hydrogen bonds between Ile50/50’ and DRV are conserved in the majority of crystal structures of PR and its mutants with inhibitors and substrates. Compared with the wild type complex, DRV shows weaker interactions with the main chain of Asp30, which is associated with the shift of the aniline group of DRV away from the Asp30 of PR_{L76V} (Figure 2.6B). The interatomic distance representing a hydrogen bond between the aniline amino group of DRV and the amide of Asp30 in PR-DRV is elongated from 3.2 to 3.6 Å in the A subunit of PR_{L76V}-DRV, while the interatomic distance between the DRV amino group and the carbonyl oxygen of Asp30 has lengthened from 3.3 to 4.4 Å showing loss of a hydrogen bond in the PR_{L76V} mutant. These changes may correlate with the loss of internal van der Waals contacts of Val76, especially those with Asp30, relative to those of the wild type Leu76.

SQV is bound at the active site in two pseudosymmetric orientations in PR_{L76V}-SQV. The major conformation of SQV forms 9 direct hydrogen bond interactions and four water-mediated hydrogen bonds with PR_{L76V} (Figure 2.6C). The mutant exhibits one improved direct and one more water-mediated interaction with SQV as compared with those in the wild type complex. Notably, a hydrogen bond interaction is formed between the carbonyl oxygen of Gly27 and the amide of SQV at 3.3 Å distance compared to the very weak 3.6 Å-long interaction observed in the wild type complex (Figure 2.6D). The conserved water-mediated hydrogen bonds are seen between Ile50/50’ and SQV, and another water-mediated interaction is conserved in reported high resolution crystal structures of PR and its mutants with SQV (Tie et al., 2007). Compared with the PR-SQV crystal structure, one more water-mediated hy-
drogen bond is observed between SQV and the amide of Asp29 in PR\textsubscript{L76V}. No water was visible at the equivalent position in the wild type structure; however, the observation of ordered solvent molecules partly depends on the crystallization conditions, quality and resolution of the diffraction data. Overall, SQV has slightly improved hydrogen bond interactions with PR\textsubscript{L76V} relative to the wild type PR suggesting that the mutant may retain high binding affinity for this inhibitor.

### 2.4.7 Implications for Drug Resistance

The structural changes in the PR\textsubscript{L76V}-inhibitor complexes suggest the molecular basis for the lowered stability of this mutant and slower autoprocessing of its precursor. The reduced hydrophobic interactions of Val76 can directly perturb the interactions of DRV with Asp30, although no direct effect on the SQV interactions was seen. Also, the mutation is likely to perturb the flaps through loss of contacts between residue 76 and flap residues Lys45 and Ile47. Movement of the flaps is required for binding of substrate and release of products (Liu et al., 2006). Inter-flap contacts also contribute to the dimer interface, so that destabilization of the flaps will likely contribute to the higher dimer dissociation constant of PR\textsubscript{L76V} and its increased susceptibility to thermal and urea-induced denaturations. Consequently, precursor processing, formation of the mature dimer and drug resistance depend on maintaining the correct flap conformations and their contacts with internal hydrophobic residues like Leu76.

The properties of this mutant are discussed in relation to the interactions of the nine clinical inhibitors in order to understand why the L76V mutation is associated with increased resistance to some drugs while retaining effective binding affinity for other clinical inhibitors. DRV forms no direct hydrogen bonds with flap residues suggesting that its binding may allow changes in flap conformation (Figure 2.6A). In contrast, SQV forms a hydrogen bond with the carbonyl oxygen of Gly48 in the flap, which suggests SQV binding may restrict such conformational changes (Figure 2.6C). This analysis of the PR\textsubscript{L76V} crystal structures and other PR complexes sheds light on the opposing clinical responses of this mutant to different drugs (Nijhuis et al., 2009a, Young et al., 2010, Mueller et al., 2004). L76V is associated with
resistance to the inhibitors LPV, DRV, APV, and IDV, which form hydrogen bond interactions with Gly27, Asp29 and/or Asp30 near the catalytic Asp25, as well as the conserved water-mediated interactions with Ile50/50’ (Tie et al., 2004, Mahalingam et al., 2004, Muzammil et al., 2007, Shen et al., 2010). On the other hand, inhibitors ATV, SQV and TPV, for which mutants containing L76V show increased susceptibility, all form direct or water-mediated hydrogen bond interactions with Gly48, while TPV is unique in forming direct, instead of water-mediated, hydrogen bonds with the amides of Ile50 and 50’ (Muzammil et al., 2007). NFV is the only exception with no direct flap interactions (Kaldor et al., 1997), however, susceptibility to NFV is less strongly associated with the L76V mutation compared to other drugs (Young et al., 2010, Rhee et al., 2010).

The observed structural changes provide insight into the effects of combining mutation L76V with M46I, which has been reported to contribute strongly to competence of the virus to replicate and its clinical resistance to LPV (Nijhuis et al., 2009b). Analysis of the wild type PR crystal structures suggests that the P2 group of LPV is unusual in forming multiple close hydrophobic contacts with Ile47, whereas DRV and SQV show only 1 or 2 hydrophobic contacts with Ile47. Thus, altered contacts with flap residues 45 and 47 associated with mutation L76V may partially diminish the binding affinity of LPV and contribute to resistance to this drug while having less effect on the other drugs. Residues 46 and 76 have no direct contacts in the dimer structure, since residue 76 is part of the internal hydrophobic cluster, while the side chain of residue 46 points away from the protein surface of the flexible flap. Instead, Leu76 contacts the side chains of Lys45 and Ile47 on either side of Met46. In PR_{L76V}, Lys45 and Ile47 show reduced interactions with Val76, suggesting that M46I may indirectly compensate for the destabilizing effects of L76V. Notably, although PR_{L76V} shows a much slower rate of autoprocessing of its precursor relative to the wild type, in agreement with reported defective viral replication (Nijhuis et al., 2009a), the rate is increased upon introduction of M46I (Figure 2.4D). Furthermore, the presence of both mutations significantly restores autoprocessing capability in the presence of LPV, although inhibi-
tion by DRV is retained. These results are consistent with previous data on viral replication in cell cultures (Nijhuis et al., 2009a) and suggest that investigating effects on precursor processing and its inhibition will provide a useful approach, which is complementary to studies of inhibitor interactions with the mature protease, for understanding the role and interactions of multiple mutations in drug resistance.

2.5 Acknowledgement

We thank Annie Aniana for technical assistance and the staff at SER-CAT beamline at the Advanced Photon Source, Argonne National Laboratory, for assistance during X-ray data collection. Use of the Advanced Photon Source was supported by the U. S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under Contract No. W-31-109-Eng-38. DRV, LPV and SQV were obtained through the NIH AIDS Research and Reference Reagent Program, Division of AIDS, NIAID, NIH.
Table 2.1 Dimer dissociation, urea and thermal denaturation of PR\textsubscript{L76V}

<table>
<thead>
<tr>
<th>Protease</th>
<th>UC\textsubscript{50} (M)</th>
<th>K\textsubscript{d} (nM)</th>
<th>T\textsubscript{m} (°C) no inhibitor</th>
<th>T\textsubscript{m} (ΔT\textsubscript{m}) (°C) DRV</th>
<th>T\textsubscript{m} (ΔT\textsubscript{m}) (°C) SQV</th>
<th>T\textsubscript{m} (ΔT\textsubscript{m}) (°C) LPV</th>
</tr>
</thead>
<tbody>
<tr>
<td>PR\textsubscript{L76V}</td>
<td>0.85</td>
<td>71 ± 24</td>
<td>53.7</td>
<td>80.9 (27.2)</td>
<td>79.5 (25.8)</td>
<td>79.1 (26.1)</td>
</tr>
<tr>
<td>PR</td>
<td>1.78</td>
<td>&lt;10\textsuperscript{5}</td>
<td>65.7\textsuperscript{a}</td>
<td>88.1\textsuperscript{b} (22.4)</td>
<td>85.0\textsuperscript{b} (19.3)</td>
<td>nd (20.4)\textsuperscript{c}</td>
</tr>
</tbody>
</table>

\textsuperscript{nd} – not determined by DSC
\textsuperscript{a}Data from (Louis et al., 1999a)
\textsuperscript{b}Data from (Sayer et al., 2009)
\textsuperscript{c}ThermoFluor data from (Sayer et al., 2009)
Table 2.2 Crystallographic data collection and refinement statistics

<table>
<thead>
<tr>
<th></th>
<th>PR&lt;sub&gt;L76V&lt;/sub&gt;</th>
<th>PR&lt;sub&gt;L76V&lt;/sub&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Protease</strong></td>
<td>PR&lt;sub&gt;L76V&lt;/sub&gt;</td>
<td>PR&lt;sub&gt;L76V&lt;/sub&gt;</td>
</tr>
<tr>
<td><strong>Inhibitor</strong></td>
<td>DRV</td>
<td>SQV</td>
</tr>
<tr>
<td><strong>Space group</strong></td>
<td>P&lt;sub&gt;2&lt;/sub&gt;₁&lt;sub&gt;2&lt;/sub&gt;₁&lt;sub&gt;2&lt;/sub&gt;₂</td>
<td>P&lt;sub&gt;2&lt;/sub&gt;₁&lt;sub&gt;2&lt;/sub&gt;₁&lt;sub&gt;2&lt;/sub&gt;₂</td>
</tr>
<tr>
<td>Unit cell dimensions (Å)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>58.32</td>
<td>58.84</td>
</tr>
<tr>
<td>b</td>
<td>86.33</td>
<td>86.17</td>
</tr>
<tr>
<td>c</td>
<td>45.98</td>
<td>46.24</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>50-1.46</td>
<td>50-1.45</td>
</tr>
<tr>
<td>Unique reflections</td>
<td>37895</td>
<td>38308</td>
</tr>
<tr>
<td>R&lt;sub&gt;merge&lt;/sub&gt; (%) overall&lt;sup&gt;a&lt;/sup&gt;</td>
<td>5.2% (49.2%)</td>
<td>6.4% (48.0%)</td>
</tr>
<tr>
<td>&lt;I/σ&gt; overall&lt;sup&gt;a&lt;/sup&gt;</td>
<td>34.9 (2.1)</td>
<td>27.3 (2.2)</td>
</tr>
<tr>
<td>Data range for refinement (Å)</td>
<td>10-1.46</td>
<td>10-1.45</td>
</tr>
<tr>
<td>Completeness (%) overall&lt;sup&gt;a&lt;/sup&gt;</td>
<td>91.9% (54.9%)</td>
<td>90.1% (56.4%)</td>
</tr>
<tr>
<td>R&lt;sub&gt;work&lt;/sub&gt;</td>
<td>0.1404</td>
<td>0.1449</td>
</tr>
<tr>
<td>R&lt;sub&gt;free&lt;/sub&gt;</td>
<td>0.1891</td>
<td>0.1971</td>
</tr>
<tr>
<td>No. of solvent (total occupancies)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>278 (209.9)</td>
<td>191 (151.7)</td>
</tr>
<tr>
<td>RMS Deviation from ideality</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bonds (Å)</td>
<td>0.011</td>
<td>0.011</td>
</tr>
<tr>
<td>Angle distance (Å)</td>
<td>0.032</td>
<td>0.029</td>
</tr>
<tr>
<td>Average B-factors (Å&lt;sup&gt;2&lt;/sup&gt;)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Main chain atoms</td>
<td>14.1</td>
<td>18.9</td>
</tr>
<tr>
<td>Side chain atoms</td>
<td>19.6</td>
<td>23.9</td>
</tr>
<tr>
<td>Inhibitor</td>
<td>10.9</td>
<td>15.4</td>
</tr>
<tr>
<td>Solvent</td>
<td>28.1</td>
<td>36.8</td>
</tr>
<tr>
<td>Relative occupancy of inhibitor</td>
<td>0.64/0.36</td>
<td>0.73/0.27</td>
</tr>
</tbody>
</table>

<sup>a</sup>Values in parentheses are given for the highest resolution shell  
<sup>b</sup>Values in parentheses are given for total occupancies
Figure 2.1 HIV-1 PR dimer structure (grey and yellow ribbons represent the two subunits) bound to darunavir (red sticks). The site of mutation is indicated by magenta sticks for Leu76 on one subunit. Positions of residues interacting with Leu76 are indicated and colored in green.
Figure 2.2 Urea denaturation of PR<sub>L76V</sub> and wild-type PR and kinetic determination of PR<sub>L76V</sub> dimer dissociation. A. UC<sub>50</sub> = 0.85 M and 1.78 M, for PR<sub>L76V</sub> and PR<sub>WT</sub>, respectively, dashed intersect on X-axis). B. A Dissociation constant, $K_d = 71 \pm 24$ nM for PR<sub>L76V</sub>, indicated by the dashed vertical line at 50% activity, was determined as described (see Materials and Methods). For the wild type PR, $K_d$ is < 10 nM [not shown, (Louis et al., 2007)].
Figure 2.3 Thermochemical data for interactions of inhibitors with PR₇₆V. A. DSC thermograms of PR₇₆V in the presence and absence of inhibitors in 50 mM sodium acetate buffer, pH 5.0. The thermal transition temperatures (curve maxima) are in parentheses. Inhibitors were in a twofold molar excess relative to dimeric PR₇₆V. B. ITC trace of 5 mM PR₇₆V titrated with 80 mM RPB (Kᵢ = 61 nM) in 50 mM sodium acetate buffer, pH 5. C. Displacement of RPB (total concentration of 16 mM) bound to 4.2 mM PR₇₆V by titration with DRV (50 mM; Kᵢ = 0.79 nM).
Figure 2.4 Autocatalytic maturation of precursors (TFR-PR<sub>L76V</sub> and TFR-PR<sub>M46I/L76V</sub>) at pH 4.5 and its inhibition. Inhibitors were present at a 1:1 molar ratio (6 mM) to precursors (as dimers). (A and E) Controls without inhibitor. (B and F) In the presence of LPV. (C and G) In the presence of SQV. (D and H) In the presence of DRV. Numbers indicated on top of the gels denote autoprocessing reaction times in hours. Lane M corresponds to molecular weight markers of 97, 66, 45, 30, 20 and 14 kDa from the top. Letter designations for the bands are: a, full length precursor; b, TFR<sup>9-56</sup>-PR; c, mature, active PR; d, the fragment TFR<sup>9-56</sup>. 
Figure 2.5 Hydrophobic interactions of residue 76 in the crystal structures of PR$_{L76V}$ (cyan), and wild type PR (grey). A. Interactions of residue 76 in subunit A of PR$_{L76V}$-DRV and PR-DRV. Leu76 forms van der Waals contacts (3.6-4.2 Å) with the side chains of Asp30, Val32 Lys45, Ile47, and Thr74 (black dashed lines), while Val76 in PR$_{L76V}$ only has hydrophobic contacts with Val32 and Ile47 (red dashed lines). Neighboring residue Gln58 is not shown since it forms similar interactions with residue 76 in PR and PR$_{L76V}$. Val56 is omitted for clarity. B. Interactions of residue 76 with Val32 and Val56 in PR$_{L76V}$-DRV and PR-DRV. C. Interactions of residue 76 with Val32 and Val56 in PR$_{L76V}$-SQV and PR-SQV. In B and C the number of hydrophobic contacts between the side chains is indicated by black (with Leu76) and red (with Val76) dashed arrows.
Figure 2.6 Protease-inhibitor hydrogen bond interactions. Hydrogen bonds are indicated by dotted lines. Water molecules are represented as red spheres. A. PR_{L76V} hydrogen bond interactions with the major conformation of DRV (in yellow). B. Superposition of PR_{L76V}-DRV (cyan bonds) and PR-DRV (grey bonds) showing PR_{L76V} has fewer hydrogen bond interactions with the aniline group of DRV. The side chain of Asp30 has two alternate conformations in the PR-DRV structure. Interatomic distances are shown in Å with red dotted lines indicating the hydrogen bond interactions in the wild type complex, and black broken lines show the larger interatomic separation in the mutant. C. PR_{L76V} hydrogen bond interactions with the major conformation of SQV (in green). The red dotted lines indicate the new hydrogen bond interactions formed by SQV in PR_{L76V} relative to PR-SQV. D. Superposition of PR_{L76V}-SQV (cyan bonds) and PR-SQV (grey bonds) showing the improved interaction of SQV with PR_{L76V} arising from only slight structural changes. Red dotted lines indicate the new hydrogen bond interaction in the mutant, and black broken lines show the larger interatomic separation in the wild type SQV complex.
3 STRUCTURES OF HIGHLY DARUNAVIR-RESISTANT HIV-1 PROTEASE MUTANT REVEAL WIDE OPEN FLAPS AND ATYPICAL BINDING SITE FOR DARUNAVIR


3.1 Abstract

The molecular basis for high resistance to clinical inhibitors of HIV-1 protease (PR) was examined for the variant designated PR$_{P51}$ that was selected for resistance to darunavir (DRV). High resolution crystal structures of PR$_{P51}$ with the active site D25N mutation revealed a ligand free form and an inhibitor-bound form showing a unique binding site and orientation for DRV. This inactivating mutation is known to increase the dimer dissociation constant and DRV affinity of PR. The PR$_{P51}$ dimers were in the open conformation with widely separated flaps, as reported for other highly resistant variants. PR$_{P51}$ dimer bound to two DRV molecules showed larger separation of 8.7 Å between the closest atoms of the two flaps compared with 4.4 Å for the ligand-free structure of PR$_{P51}$. Unlike the majority of PR structures, however, the ligand-free conformation lacked van der Waals contacts between Ile50 and Pro81’ from the other subunit. DRV binds inside the active site cavity, however, the inhibitor is oriented almost perpendicular to the typical position and exhibits mostly hydrophobic interactions with only 2 direct hydrogen bond and two water-mediated interactions with atoms of PR$_{P51}$ compared with 11 hydrogen bond interactions seen for DRV bound in the typical position in wild-type enzyme. However, the involvement of crystal lattice contacts cannot be ruled out due to DRV interactions with a symmetry-related dimer. The atypical location of darunavir may provide opportunities for design of novel inhibitors targeting the open conformation of PR drug resistant mutants.

3.2 Introduction

HIV-1 protease (PR) has been a successful target in AIDS therapy due to its critical role in viral maturation by hydrolyzing the Gag and Gag-Pol precursor polyproteins into mature structural and func-
tional proteins (Louis et al., 2000, Weber et al., 2009a). A series of clinical HIV-1 protease inhibitors (PIs) has improved the survival of AIDS patients. One such inhibitor, darunavir (DRV), which was designed to target drug resistance by introducing strong polar interactions with the main chain atoms of the PR (Ghosh et al., 2012, Koh et al., 2003, Tie et al., 2004), has been widely used for the treatment of drug-naïve patients and those infected with multidrug-resistant HIV-1 (de Meyer et al., 2008). DRV effectively inhibits PR enzymatic activity with picomolar binding affinity assessed by isothermal titration calorimetry (ITC) (King et al., 2004). However, HIV evolves resistance to DRV by selecting a combination of mutations (de Meyer et al., 2008).

Highly DRV-resistant HIV-1 variants were selected in the laboratory to elucidate the mechanism for resistance (Koh et al., 2010). A mixture of 8 highly DRV-susceptible HIV-1 clinical isolates (HIV-1\textsubscript{mix}) containing 9 to 14 PI-resistant mutations was propagated in the presence of DRV. The viral population at passage 51 (HIV-1\textsubscript{mix\_PS1}) replicated well at the concentration of 5 \(\mu\)M DRV and sequencing revealed 14 amino acid substitutions in the PR gene (Figure 3.1) (Koh et al., 2010). The viral strain HIV-1\textsubscript{mix\_PS1} was highly resistant, with half maximal effective concentration (EC\textsubscript{50}) for inhibition of viral replication increased to more than 1 \(\mu\)M for DRV and most other PIs, and showed moderate resistance to saquinavir (SQV) (0.3 \(\mu\)M EC\textsubscript{50}) (Koh et al., 2010).

We have investigated the physical and biochemical properties of several resistant variants, including the HIV-1\textsubscript{mix\_PS1} protease (PR\textsubscript{PS1}) (Louis et al., 2011a). PR\textsubscript{PS1} and another highly resistant variant with 20 mutations (PR20) showed several extreme properties contributing to resistance. The affinity of DRV and SQV for PR\textsubscript{PS1} as measured by isothermal titration calorimetry (ITC) gave \(K_i\) values of 37 and 54 nM, respectively, or about 7400-fold and 135-fold weaker than the corresponding values for wild-type PR. Autocatalytic cleavage (autoprocessing) of precursor Gag-Pol polyprotein is essential to produce mature and fully active PR (Louis et al., 1999a). Autoprocessing of the precursor comprising the 56-amino acid transframe region (TFR) fused to PR (TFR-PR) was inhibited the best by DRV and SQV as compared
to other clinical inhibitors, although in the low μM range. However, autoprocessing of the TFR-PR<sub>p51</sub> precursor was uninhibited by DRV and marginally inhibited by SQV, at 150 μM PI concentration (Louis et al., 2011a). These properties of PR<sub>p51</sub> are consistent with the high antiviral resistance to DRV measured for virus bearing this variant (>300-fold increased EC<sub>50</sub>) relative to wild type (Koh et al., 2010).

We have reported recently the structural analysis of another extremely resistant variant, PR20, which showed no inhibition of precursor autoprocessing, and K<sub>I</sub> values for mature PR20 increased by more than 8000- and 2000-fold, respectively, for DRV and SQV (Louis et al., 2011a, Agniswamy et al., 2012). Crystallographic analysis of PR20 showed fewer interactions with inhibitors and widely separated flaps in the absence of inhibitors with no contact of the flap with residues from the other subunit unlike interactions in the corresponding structures of wild-type PR (Spinelli et al., 1991, Heaslet et al., 2007). An increased separation of the flaps in the dimer may be typical of many resistant variants, as seen also for variant MDR769 (Logsdon et al., 2004) and in solution studies using Double Electron-Electron Resonance (DEER) spectroscopy (de Vera et al., 2012).

In order to further investigate the molecular mechanisms of high-level resistance we have determined the crystal structures of PR<sub>p51</sub> bearing the inactivating mutation D25N to abolish self-degradation (autoproteolysis) for sample handing during crystallization. When the D25N mutation was introduced into wild-type PR, the affinity for DRV was decreased by about 10<sup>6</sup> fold, while no substantial changes were observed in the crystal structures (Sayer et al., 2008). Two structures were obtained for PR<sub>p51</sub>: a DRV bound structure and a ligand free structure. Structural analysis revealed an unusual binding site for DRV and the widely separated flaps that characterize the ligand-free structures of many highly resistant variants (Arpino et al., 2012, Logsdon et al., 2004, Agniswamy et al., 2012).
3.3 Results and Discussion

3.3.1 Crystal Structures of Ligand-free and DRV Bound PRP51.

Repeated attempts with active PRP51 did not yield crystals likely due to its enhanced autoproteolysis. Consistent with this observation even storage of the protein in 12 mM HCl prior to folding showed degradation products unlike the optimized wild-type PR. The optimized wild-type PR bears the mutations L33I and L63I shown to significantly restrict autoproteolysis of wild-type PR in addition to Q7K, which exists in PRP51 (Louis et al., 1999a, Mildner et al., 1994). Also, as DRV interactions with PRD25N are nearly identical to those in wild-type PR/DRV complex (Sayer et al., 2008), we resorted to using PRP51 for our studies.

Crystal structures of mutant PRP51 in complex with DRV and SQV were determined to identify the structural changes associated with high level resistance. Electron density was observed within the ligand binding cavity only in the structure of PRP51/DRV. No inhibitor was visible in the PRP51 structure obtained from crystals grown in the presence of SQV. Similar results were obtained when the mutant was crystallized in the presence of amprenavir or tipranavir. The two crystal structures designated PRP51/DRV and PRP51 were refined with X-ray data at resolutions of 1.66 and 1.50 Å and R-factors of 18.9% and 15.9%, respectively. The crystallographic statistics are listed in Table 3.1. PRP51/DRV was refined in space group P4₁2₁2 with a monomer of residues numbered 1-99 in the asymmetric unit, while PRP51 was refined in space group P4₁ and contained one dimer of residues numbered 1-99 and 1’-99’ in the asymmetric unit. Alternate conformations were refined for 3 residues in PRP51/DRV and 7 residues in PRP51 structures. Residues 34 to 36 from each monomer of the two structures showed similar alternate conformations with 0.5 relative occupancy. The crystallographic dimer of PRP51/DRV was generated for structural analysis.
3.3.2 PR<sub>P51</sub>/DRV Exhibits Unusual Binding Conformations of DRV

DRV placed at the typical binding site (Figure 3.2A) (Tie et al., 2004, Kovalevsky et al., 2006), did not fit the electron density visible in the ligand binding cavity of PR<sub>P51</sub>/DRV and thus different locations were evaluated for the inhibitor. After manual adjustment, the final monomer structure was refined with full occupancy molecules of DRV oriented approximately perpendicular to DRV in the typical inhibitor-binding site of wild-type PR or most mutants (Figure 3.2B), as indicated by the clear electron density map (Figure 3.3A). The two molecules of DRV are related by 180-degree rotation and interact with each other in the binding cavity of the PR<sub>P51</sub> dimer (Figure 3.3B). The polar interactions between the two DRV molecules include a water mediated hydrogen bond connecting a sulfonyl oxygen and hydroxyl group with the amino group on the aniline of the other DRV, and C-H...water interactions with the phenyl group and the bis-THF of both DRVs. The bound DRV has relatively few hydrogen bonds and many hydrophobic contacts with the protein (Figure 3.3C and 3.3D). These DRV molecules have a different conformation and interactions from those for the regular binding mode of DRV, which has 11 direct hydrogen bond interactions and 4 water-mediated ones with PR in the wild-type PR/DRV complex (PDB ID 2IEN) (Figure 3.4A and 3.4B) (Tie et al., 2004). Many van der Waals contacts were observed between PR<sub>P51</sub> and DRV with distances ranging from 3.6-4.2 Å. In the crystal structure of PR<sub>P51</sub>, DRV forms two direct hydrogen bonds with the main chain amides of Gly49 and Gly50, two hydrogen bonds with the side chain of Gln61 from a symmetry-related dimer and three water-mediated hydrogen bonds with protein atoms. Van der Waals interactions occurred between DRV and PR<sub>P51</sub> residues Asn25, Ala28, Ile47, Gly48, Gly49, Ile50, Gly52, Phe53, Met54, Thr80, Pro81′, Ile82 and Val84. These interacting residues include mutations of V32I, I54M, V82I and I84V from the selected P51 isolate and the D25N mutation. It is probable that these mutations facilitate binding of DRV to this atypical site. In addition, DRV forms interactions with residues from symmetry-related dimers of PR<sub>P51</sub> as shown in Figure 3.4B and Figure 3.9. The two DRV molecules have almost identical interactions with Ile72 and Gln61 from two
different symmetry-related PR_{ps1} dimmers (Figure 3.9). The interactions include two direct hydrogen bonds between the two oxygens of the bis-THF and the amino group on the side chain of Gln61 and van der Waals contacts with Ile72 and Gln61. The existence of DRV interactions with the symmetry related PR_{ps1} raises the possibility that crystal lattice contacts influence the atypical binding of inhibitor.

The question of whether DRV might bind at this atypical site in wild-type enzyme can be addressed by structural comparisons. The PR_{ps1}/DRV monomer shares an almost identical wide open conformation (RMSD of 0.33 Å on equivalent Cα atoms) with the wild-type PR crystallized with Mg^{2+} coordinated at the active site (PDB ID 2PC0) (Heaslet et al., 2007). Superposition of the two monomers reveals that rotations of three side chains in each subunit would be required for DRV to bind in this atypical site in wild-type enzyme (Figure 3.5). The side chain of Ile54 in wild type PR would hinder binding of the second THF of the bis-THF group of DRV due to short interatomic distances of 2.0-3.1 Å (Figure 3.5). Also, the side chains of Val82 and Ile84 in wild type PR extend closer to the aniline group of DRV compared to the good hydrophobic contacts formed by Ile82 and Val84 in PR_{ps1}/DRV. However, rotated conformers of the side chains of Ile54, Val82 and Ile84 could allow the possible binding of DRV in the atypical site. The affinity of DRV for binding in the typical location is expected to be decreased drastically by the mutations. The PR_{ps1} variant showed 7400-fold lower affinity for DRV (Louis et al., 2011a) while the D25N mutation alone produces about 10^6 decreased affinity for DRV as measured by ITC (Sayer et al., 2008). Therefore, the combination of D25N plus the 14 mutations in PR_{ps1} is expected to compromise the affinity to low µM levels (Sayer et al., 2008). Consequently, DRV may favor the weaker atypical binding location observed in PR_{ps1}.

It is not the first time that DRV has been observed to bind at an unusual site in HIV-1 protease variants. A second binding site was observed in an ultra-high resolution (0.84 Å) structure of PR_{V32}/DRV in addition to the typical active site binding. The second DRV bound in a groove on one flap surface where the residues Glu35′, Lys45′, Lys55′, Val56′, and Arg57′ participated in the major interactions
(Kovalevsky et al., 2006). A similar DRV binding site occurred in the crystal structures of PR46/DRV and PR20/DRV (Kovalevsky et al., 2006, Agniswamy et al., 2012). Also for SQV, a second molecule was found in a location adjacent to the usual active site location in PR20/SQV and PRV32/I47/V82/SQV structures (Agniswamy et al., 2012, Tie et al., 2012). Another unique binding mode was found for GB-18, specifically [3-cobalt bis(1,2-dicarbollide)]-ion, which belongs to a novel class of inorganic cobaltacarborane inhibitors, in the active site cavity of the wild-type PR (PDB ID 1ZTZ) (Kozisek et al., 2008). This binding mode involves two molecules of GB-18 positioned asymmetrically inside the pseudosymmetric active site cavity (Kozisek et al., 2008). These binding pockets for GB-18 are formed by the residues Pro81, Ile84, and Val82 and covered by the flap residues Ile47, Gly48, and Ile54 forming a semi-open conformation of the PR dimer (Kozisek et al., 2008).

### 3.3.3 Flaps of PR$_{PS1}$/DRV and PR$_{PS1}$ Display Different Intersubunit Interactions

Both DRV bound and ligand-free PR$_{PS1}$ dimers have flaps separated by a large distance between their tips. The flaps of PR$_{PS1}$/DRV had a larger separation of 8.7 Å between the closest atoms, while the flaps were separated by 4.4 Å for PR$_{PS1}$. Superposition of the monomer of PR$_{PS1}$/DRV with each subunit of PR$_{PS1}$ gave the overall RMSD value of 0.97 Å and 0.41 Å on Cα atoms, respectively, with large differences in the conformation of the two flaps and $80^\prime$s loops (residues 79/79' to 83/83') (Figure 3.6A). One of the two flaps (residues 47/47' to 54/54') of the PR$_{PS1}$/DRV dimer is further away from the catalytic site than seen for the equivalent flap of PR$_{PS1}$ as indicated by the distance of 5.6 Å between the equivalent Cα atoms of Ile50 in these two structures, while the other flap conformation is more similar in the two structures with only 1.5 Å distance between the Cα atoms of Ile50' residues (Figure 3.6B).

The two PR$_{PS1}$ structures were compared with open conformation structures of PR and PR20 as well as their DRV-bound complexes. The open conformations of ligand-free PR20 have two unusual features relative to other reported dimer structures: widely separated flaps and no intersubunit van der Waals contacts between the flap tip and residues from the other subunit (Agniswamy et al., 2012).
Therefore, the conformations were assessed by measuring the closest distance between atoms at the tip of the two flaps within one dimer structure and the closest intersubunit contact of Ile50/50’ at the flap tip with Pro81’/81. The shortest interatomic distances between the flap tips were 4.0 Å in ligand-free PR\textsubscript{P51}, compared to 3.0 Å in a typical open conformation of wild-type PR (PDB ID 1HHP), 7.7 Å in another wild-type PR with Mg\textsuperscript{2+} coordinated at the active site (PDB ID 2PC0), and 6.0 Å in ligand-free PR20 (PDB ID 3UF3) (Agniswamy \textit{et al.}, 2012, Heaslet \textit{et al.}, 2007, Spinelli \textit{et al.}, 1991). The various flap conformations in wild-type PR (PDB ID 1HHP), ligand-free PR\textsubscript{P51} and ligand-free PR20 are compared in Figure 3.10A. For the DRV complexes, the shortest interflap distance was 8.7 Å in PR\textsubscript{P51}/DRV compared with about 3.3 Å for typical closed conformation dimers of PR/DRV (PDB ID 2IEN) and PR20/DRV (PDB ID 3UCB) (Figure 3.10B) (Agniswamy \textit{et al.}, 2012, Tie \textit{et al.}, 2004). Ile50/50’ at the flap tips has intersubunit van der Waals contacts of about 4.0 Å with Pro81’/81 in PR\textsubscript{P51}/DRV as observed for closed conformation inhibitor-bound dimers. The majority of dimers show intersubunit van der Waals contacts between Ile50/50’ and Pro81’/81, except for the ligand-free PR20 where these two side chains are separated by about 7 Å. Moreover, this intersubunit contact was lost in ligand-free PR\textsubscript{P51} since the closest atoms of Ile50 and Pro81’ were separated by 8.5 Å. The ligand-free PR\textsubscript{P51} has asymmetric flap conformations like another open conformation dimer of PR20 (PDB ID 3UHL), which had van der Waals contact between the flap tips, and asymmetric flaps with 12.2 and 5.4 Å intersubunit separation between Ile50/50’ and Pro81’/81 (Figure 3.7). Therefore, the large separation (~7-12 Å) between side chain atoms of Ile50 and Pro81’ from the other subunit is conserved in the ligand-free dimers of the two highly resistant mutants PR20 and PR\textsubscript{P51}.

\subsection*{3.3.4 Multiple Mutations Contribute to the Structural Changes in PR\textsubscript{P51}}

PR\textsubscript{P51} bears 14 mutations (L10I, I15V, K20R, L24I, V32I, L33F, M36I, M46L, I54M, L63P, K70Q, V82I, I84V and L89M) relative to the standard “wild-type” PR sequence. Four mutations, V32I, M46I, V82I and I84V, alter residues in the active site cavity where substrates and inhibitors usually bind. Addi-
tionally, “second shell” mutations L10I, L24I, L33F, I54M alter residues that form direct interactions with residues in the active site cavity. Our crystallographic and biochemical analysis has demonstrated the changes due to the individual single mutations of L24I, V32I, M46L, I54M, and I84V, as reviewed in (Weber et al., 2009a). Several of the other substitutions in PR<sub>PS1</sub> are shared by the highly resistant multiple mutant PR20 and their coordinated effects have been described previously (Agniswamy et al., 2012).

The mutations of residues V32I, V82I and I84V in the active site cavity are assumed to contribute to the poor affinity of PR<sub>PS1</sub> for inhibitors and the observed unusual binding site for DRV (Figure 3.3). The side chains of these residues form hydrophobic interactions with each other and with Val47 in all structures. Mutations V32I and I47V have altered interactions with inhibitors DRV and SQV (Agniswamy et al., 2012, Weber et al., 2009a), however, V82I in the triple mutant V32I/I47V/V82I bearing the active site residues of HIV-2 PR does not significantly alter direct contacts with inhibitor (Tie et al., 2012). Further comparison of inhibitor interactions is limited since the crystal structures of the PR<sub>PS1</sub> mutant are in the open conformation without inhibitors bound in the standard active site location.

Mutations M46L and I54M alter residues in the flaps and are proposed to have small indirect effects on inhibitor binding and may alter the flap dynamics (Shen et al., 2010, Zhang et al., 2013a, Kovalevsky et al., 2006). In PR<sub>PS1</sub>/DRV, the I54M mutation introduced new van der Waals interactions with DRV (Figures 3.3C and 3.5), while M46L had no contacts with the ligand. The changes in interactions of mutated residue L24I are similar in PR<sub>PS1</sub> and the single mutant PR<sub>L24I</sub> (Liu et al., 2005). Ile24 gains two good van der Waals contacts with Leu90 and loses or elongates contacts (4.7 and 4.2 Å in the two subunits compared to 3.8 Å distance in wild-type PR) with Phe99' from the other subunit. Unlike the mutated side chain of L10F in PR20/DRV, in which a new hydrophobic contact was formed between the side chains of Phe10 and mutated Ile82 (Agniswamy et al., 2012), mutated residue L10I in PR<sub>PS1</sub> yields no new interactions with nearby residues. Mutation L33F introduces the large bulky Phe side chain, which maintains hydrophobic contacts of the wild-type enzyme, including contacts with mutated residues
I15V, M36I substituting shorter side chains as reported for L33F in PR20 (Agniswamy et al., 2012). Also, the flap hinge region comprising residues 34 to 43 shares a very similar conformation in PR$_{PS1}$ and in the ligand free conformation of PR20, which is likely due to the presence of mutations M36I and I33F in both highly drug-resistant variants. The mutations in the flap hinge and flaps are likely to contribute to the extended flap conformations observed in PR$_{PS1}$ and in inhibitor-free PR20 structures.

Mutation L89M has not been analyzed previously in structures. The side chain of Leu89 in wild-type PR forms hydrophobic contacts with the side chains of Ile64, Ile66, Ala71, Gly73, Ile85 and Asn88, as well as van der Waals and C-H...O interactions with the side chain of Thr31 (Figure 3.8). The mutated Met89 forms very similar contacts in PR$_{PS1}$, except for an additional van der Waals contact with the side chain of Val75. Mutation K20R alters a residue near the protein surface showing varied interactions with other surface side chains. Mutations L63P and K70Q also alter surface side chains that form a hydrophobic contact in the wild-type PR, which is eliminated in the mutant PR$_{PS1}$.

Two crystal structures were analyzed for the PR$_{PS1}$ variant that was selected for high levels of resistance to DRV. These structures confirm the increased separation of the two flaps in the dimer and/or loss of intersubunit contacts between the flap tip and Pro81 in comparison to the open conformations seen for the ligand free wild-type PR, as previously described for the MDR769 (Logsdon et al., 2004) and PR20 (Agniswamy et al., 2012) highly resistant variants. Other highly resistant PR mutants have shown greater flap mobility in studies employing various techniques (Agniswamy et al., 2012, Logsdon et al., 2004, de Vera et al., 2012, Cai et al., 2012) and decreased interactions with inhibitors or substrate analogs in crystal structures (Liu et al., 2013, Agniswamy et al., 2012, Saskova et al., 2009). In PR$_{PS1}$, however, DRV showed a unique mode of binding within the open conformation flaps and lying almost perpendicular to the typical active site position. Importantly, this new binding site for DRV may hint at designs for novel types of antiviral inhibitors that capture the open, inactive conformation of the protease.
3.4 Methods

3.4.1 Construction, Expression and Purification

The HIV-1 PR from Group M (Genbank HIVXB2CG) is designated as PR. The PR$_{p51}$ construct contains 14 mutations (L10I, I15V, K20R, L24I, V32I, L33F, M36I, M46L, I54M, L63P, K70Q, V82I, I84V, and L89M) plus three other mutations Q7K to minimize autoproteolysis and C67A and C95A to prevent cysteine-induced aggregation (Wlodawer et al., 1993, Louis et al., 2011a, Louis et al., 1999a). The mutant DNA was constructed by oligonucleotide synthesis and cloned into the pET11a vector between Nde1 and BamH1 restriction sites. To eliminate autoproteolysis, the inactivating mutation of D25N was introduced using the QuikChange II Site-Directed Mutagenesis Kit and confirmed by DNA sequencing. The protein was expressed in *Escherichia coli* BL21(DE3), purified and folded using the protocol described previously (Sayer et al., 2010, Louis et al., 2009).

3.4.2 Crystallization and Data Collection

Crystals of PR$_{p51}$ complexed with clinical inhibitors DRV and SQV were obtained by the hanging-drop vapor-diffusion method at room temperature using 24 well VDX plates (Hampton Research, Aliso Viejo, CA, USA). PR$_{p51}$ with a monomer concentration of 1.29 mg·ml$^{-1}$ was mixed with the inhibitors at 5-10 fold molar excess. Screening Kit I solutions (Hampton Research, Aliso Viejo, CA, USA) gave good crystals of PR$_{p51}$ complexed with DRV (0.1 M HEPES sodium pH 7.5, 0.8 M Potassium sodium tartrate tetrahydrate), and crystals of PR$_{p51}$ grown in the presence of SQV (0.1 M imidazole pH 6.5, 1.0 M sodium acetate trihydrate). The crystals were frozen in liquid nitrogen using 25% (v/v) glycerol as a cryoprotectant. X-ray diffraction data were collected at 100K by remote access on the beamline BM-22 of the Southeast Regional Collaborative Access Team (SER-CAT), the Advanced Photon Source, Argonne National Laboratory, Chicago.
3.4.3 Data Processing and Structure Determination

The X-ray data were indexed, integrated and scaled with HKL2000 (Otwinowski et al., 1997). The structures were solved by molecular replacement with MOLREP in the CPP4i suite of programs (Murshudov et al., 1997) using the PR20 complex with Yb⁺ (PDB ID 3UF3) as the starting model (Agniswamy et al., 2012). The structures were refined by REFMAC 5.2 in the CCP4 program suite 6.1.13 (Murshudov et al., 1997) and refitted using COOT 0.6.1 (Emsley et al., 2004). Alternate conformations were modeled for PR residues, inhibitors and solvent molecules based on the observed electron density maps. Anisotropic B factor refinement was applied for the higher resolution structure, and TLS restrained refinement was used for the lower resolution structure. Structural figures were prepared with PyMol (DeLano, 2002).

3.5 Acknowledgments

This work was supported in part by the Georgia State University Center for Diagnostics and Therapeutics Doctoral Fellowship (Y.Z.). Research reported in this publication was supported by the National Institute of General Medical Sciences of the National Institutes of Health under Award Number U01GM062920. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. X-ray data were collected at the Southeast Regional Collaborative Access Team (SER-CAT) beamline 22BM at the Advanced Photon Source, Argonne National Laboratory. Supporting institutions may be found at www.ser-cat.org/members.html. Use of the Advanced Photon Source was supported by the US Department of Energy, Basic Energy Sciences, Office of Science, under Contract No. W-31-109-Eng-38. Darunavir was obtained through the NIH AIDS Research and Reference Reagent Program, Division of AIDS, NIAID, NIH.
Table 3.1 Crystallographic data collection and refinement statistics

<table>
<thead>
<tr>
<th></th>
<th>PR&lt;sub&gt;P51&lt;/sub&gt;/DRV</th>
<th>PR&lt;sub&gt;P51&lt;/sub&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Complex name</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>X-ray source</td>
<td>APS_22BM</td>
<td>APS_22BM</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Space group</td>
<td>P4₁₂₂</td>
<td>P4₁</td>
</tr>
<tr>
<td>Unit cell dimensions (Å)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a = b</td>
<td>46.80</td>
<td>46.69</td>
</tr>
<tr>
<td>c</td>
<td>100.63</td>
<td>101.62</td>
</tr>
<tr>
<td>α = β = γ</td>
<td>90°</td>
<td>90°</td>
</tr>
<tr>
<td>Unique reflections</td>
<td>13867</td>
<td>34414</td>
</tr>
<tr>
<td>Redundancy&lt;sup&gt;a&lt;/sup&gt;</td>
<td>8.3 (4.6)</td>
<td>8.0 (3.4)</td>
</tr>
<tr>
<td>R&lt;sub&gt;merge&lt;/sub&gt; (%)</td>
<td>8.1% (42.2%)</td>
<td>6.7% (27.7%)</td>
</tr>
<tr>
<td>&lt;I/sigma&gt; overall&lt;sup&gt;b&lt;/sup&gt;</td>
<td>18.5 (3.3)</td>
<td>23.2 (4.6)</td>
</tr>
<tr>
<td>Completeness (%) overall&lt;sup&gt;a&lt;/sup&gt;</td>
<td>99.4% (98.3%)</td>
<td>99.4% (95.5%)</td>
</tr>
<tr>
<td>Refinement resolution range</td>
<td>22.16-1.66</td>
<td>23.64-1.50</td>
</tr>
<tr>
<td>R&lt;sub&gt;work&lt;/sub&gt;</td>
<td>18.9%</td>
<td>15.9%</td>
</tr>
<tr>
<td>R&lt;sub&gt;free&lt;/sub&gt;</td>
<td>22.6%</td>
<td>19.7%</td>
</tr>
<tr>
<td>No. of water molecules atoms (total occupancies)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>48 (45.5)</td>
<td>144 (119.5)</td>
</tr>
<tr>
<td>RMSD from ideality</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bonds (Å)</td>
<td>0.021</td>
<td>0.023</td>
</tr>
<tr>
<td>Angle distance (degrees)</td>
<td>2.326</td>
<td>2.070</td>
</tr>
<tr>
<td>Average B-factors (Å²)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Main chain atoms</td>
<td>26.33</td>
<td>23.17</td>
</tr>
<tr>
<td>Side chain atoms</td>
<td>31.10</td>
<td>27.88</td>
</tr>
<tr>
<td>Ligand</td>
<td>33.10</td>
<td>-</td>
</tr>
<tr>
<td>Solvent</td>
<td>35.35</td>
<td>32.43</td>
</tr>
</tbody>
</table>

<sup>a</sup>Values in parentheses are given for the highest resolution shell

<sup>b</sup>Values in parentheses are given for total occupancies
Figure 3.1 PR<sub>PS1</sub> mutations. (A) Sites of the 14 drug resistant mutations are mapped onto the PR<sub>PS1</sub> dimer (cyan cartoon representation). The mutations located in the active site cavity are shown as red spheres, while the flap mutations are shown as blue spheres and the mutations distal from the active site are indicated as green spheres. (B) Amino acid sequence of HIV-1 PR (upper line) and PR<sub>PS1</sub> (lower line). The amino acids are colored as in (A). Note that the wild type PR sequence used for structural comparison includes mutations Q7K, L33I, and L63I to prevent autoproteolysis, and both proteins include C67A and C95A to eliminate potential cysteine-thiol oxidation.

1PQITLWKRPL VTIKGGQLK EAALDTGADD TVIEEMSLPG RWKPKMIGGI
1PQITLWKRPI VTIKVGQLR EALIDTADD TIFEEISLPG RWKPKLIGGI

51GGFIKVRQYD QIIIEIAGHK AIGTVLVGPT PVNIIGRNL TQIGATLNF
51GGFMKVRQYD QIIIEIAGHQ AIGTVLVGPT PINVIGRNL TQIGATLNF
Figure 3.2 DRV binds to different sites in PR and PR\textsubscript{P51}. (A) The structure of wild type PR (grey color) has two conformations of DRV (red sticks) bound perpendicular to the dimer interface. (B) A different position and conformation was seen for the two molecules of DRV (red sticks and pink sticks) bound symmetrically in the PR\textsubscript{P51} dimer (green color). The black two-sided arrows indicate approximately the dimer interface in PR-DRV and PR\textsubscript{P51}. On the right side, the chemical structure of DRV is indicated in approximately the same orientation for PR and PR\textsubscript{P51}. 
Figure 3.3 Unique binding site for DRV in PRₚₛ₁. (A) \( F_o-F_c \) omit map contoured at 2.0σ showing one molecule of DRV (yellow sticks) in the crystal structure of PRₚₛ₁/DRV. (B) Interaction between two DRV molecules. Water molecules are shown as red spheres. The dotted lines represent the hydrogen bonds and the dashed lines represent potential CH…O interactions with the interatomic distances ranging from 3.0-3.8 Å. The numbers beside the lines indicate the distances. (C-D) Interactions of one DRV with PRₚₛ₁ (green sticks). The interactions with two parts of DRV are shown in (C) and (D) for clarity. The dotted lines show the hydrogen bond interactions and the double-sided arrows represent the van der Waals contacts. The numbers of van der Waals contacts are indicated. The residues labeling underlined show the mutations.
Figure 3.4 DRV has distinct hydrogen bonding interactions with wild type PR (A) and PR\textsubscript{P51} (B). DRV is represented in grey sticks and yellow sticks in PR and PR\textsubscript{P51}, respectively. The dotted lines indicate hydrogen bond interactions. Water molecules are represented by W. NH is the main chain amide group and C=O is the main chain carbonyl group. Gln61 from a symmetry-related dimer is shown in blue with parentheses.
Figure 3.5 Comparison of dimers of PR$_{PS1}$/DRV and wild type PR (PDB ID 2PC0). Superposition of the overall structures of PR$_{PS1}$/DRV (green) and wild type PR (grey). The double-sided arrows represent the close contacts between DRV and side chains of Ile54, Val82 and Ile84 in wild-type PR. The close distance ranges from 2.0-3.1 Å.
Figure 3.6 Comparison of two structures of PR_{PS1}. (A) Superposition of the overall structures of PR_{PS1} (pink) and PR_{PS1}/DRV (green). (B) The interactions of flap residues 47 to 54 are shown in the blue box; the numbers beside the black arrows show the distance between corresponding Cα atoms of Ile50/50’in the two structures.
Figure 3.7 Comparison of dimers of PR$_{PS1}$ and PR20/p2-NC$_{open}$. (A) Superposition of the overall structures of PR$_{PS1}$ (pink) and PR20/p2-NC$_{open}$ (PDB ID 3UHL, blue). (B) The flap residues 47 to 53 and Pro81 of PR$_{PS1}$ and PR20/p2-NC$_{open}$ are shown below. The numbers beside the black arrows show the distance in Å between the flap tips on Ile50 and Pro81’ of PR20/p2-NC$_{open}$ and PR$_{PS1}$. 
Figure 3.8 Interactions of Leu89 in PR/DRV (grey) and Met89 in PR$_{P51}$/DRV (green) with neighboring residues in stick representation. The double-sided arrows and the dashed lines represent the van der Waals contacts and putative C-H...O interactions, respectively, and are colored grey and green to match the structures when differences occur, while black lines indicate identical numbers of contacts in both structures. The number of van der Waals contacts is indicated by the double-sided arrows.
Figure 3.9 DRV interactions with symmetry-related monomer of PR$_{51}$. The dotted lines show the hydrogen bond interactions and the double-sided arrows represent the van der Waals contacts. The numbers of van der Waals contacts are indicated.
Figure 3.10 Comparison of PR$_{P51}$, PR$_{20}$ and wild type PR structures. (A) Superposition of ligand-free PR (PDB ID: 1HHP) (grey), PR$_{20}$ (PDB ID: 3UF3) (blue) and PR$_{P51}$ (pink) in ribbon representation with perpendicular view of flap residues 46-54. (B) Superposition of PR/DRV (PDB ID: 2IEN) (grey), PR$_{20}$/DRV (PDB ID: 3UCB) (blue) and PR$_{P51}$/DRV (green) in ribbon representation and perpendicular view of residues 46-54. DRV is omitted for clarity.
4 STRUCTURAL BASIS FOR HAND-LIKE SITE IN CALCIUM SENSOR CATCHER WITH FAST KINETICS


4.1 Abstract

Calcium ions, which are important signalling molecules, can be detected in the endoplasmic reticulum by an engineered mutant of green fluorescent protein (GFP) designated CatchER with fast off rate. Crystal structures at high resolution (1.78 to 1.20 Å) were analyzed for CatchER in the apo form and in the complexes with calcium or gadolinium to probe the binding site for metal ions. While CatchER exhibits 1:1 binding stoichiometry in solution, two positions were observed for each of the metal ions bound within the hand-like site formed by the carboxylate side chains of mutated residues S147E, S202D, Q204E, F223E and T225E that may be responsible for its fast kinetic properties. Comparison of the structures of CatchER, wild type GFP and enhanced GFP confirmed that different conformations of T203 and E222 are associated with the two forms of Y66 of the chromophore, which are responsible for the absorbance wavelengths of the different proteins. Calcium binding to CatchER may shift the equilibrium for conformational population of the E222 side chain, and lead to further changes in its optical properties.

4.2 Introduction

The calcium ion Ca$^{2+}$ acts as a ubiquitous signalling molecule in the regulation of numerous biological functions including heart beat, muscle contraction, cell development and proliferation (Berridge et al., 1998, Bers et al., 2005). Ca$^{2+}$ signals exhibit different amplitude and duration as the ions flow between subcellular compartments. Ca$^{2+}$ functions as first messenger in the central nervous system and works as an extracellular ion source for postsynaptic ligand gated channels (Berridge, 1998). The endoplasmic reticulum (ER) functions as an intracellular Ca$^{2+}$ store and the release of ER Ca$^{2+}$ triggers a series
of biological processes via binding to intracellular Ca\(^{2+}\)-sensing proteins like calmodulin (CaM) and troponin C (TnC) (Zhang et al., 2001). The Ca\(^{2+}\) signalling events are controlled by the basal ER/SR Ca\(^{2+}\) level, as well as the amplitude and the kinetics of Ca\(^{2+}\) release from the calcium stores. Hence, the determination of the concentration of free Ca\(^{2+}\) in the ER is of extensive interest, which has stimulated the development of tractable intracellular Ca\(^{2+}\) sensors.

Many efforts have been devoted to green fluorescent protein (GFP)-based Ca\(^{2+}\) fluorescent indicators, such as the cameleons (Miyawaki et al., 1997, Miyawaki et al., 1999), pericams (Persechini et al., 1997), and TN-XL (Mank et al., 2006), and TN-XXL (Mank et al., 2008). Their detection is based on either fluorescence resonance energy transfer (FRET) between two different GFP variants or the pH-dependent change in ionization state of the chromophore in circularly permutated GFP (Miyawaki et al., 1997, Persechini et al., 1997, Nagai et al., 2001, Baird et al., 1999). One common property of these sensors is that they involve the insertion of naturally occurring Ca\(^{2+}\)-sensing proteins like CaM and its target binding peptide and are capable of sensing cytosolic calcium responses in the nM to μM range (Takahashi et al., 1999, Solovyova et al., 2002, Tsien, 1998). Several ER/SR sensors with lower metal binding affinities have been developed by modifying the Ca\(^{2+}\) binding loops or the peptide interaction surface of CaM (Ohkura et al., 2005, Miyawaki et al., 1997, Palmer et al., 2004, Evanko et al., 2005, Palmer et al., 2006). These sensors exhibit some limitations such as off rates, which are not fast enough to detect the calcium release during action potentials. In addition only ~50% of the skeletal muscle cells show a response to calcium stimulation. For the FRET pair involved sensors, their highly variable basal CFP (cyan fluorescent protein)/YFP (yellow fluorescent protein) ratio and poor signal-to-noise ratio also limit quantitative determination of calcium concentration and calcium release (Rudolf et al., 2006, Jimenez-Moreno et al., 2010). Therefore, there is pressing need for new Ca\(^{2+}\) sensors targeted to cellular compartments with putative high Ca\(^{2+}\) concentration as in the ER/SR to overcome these limitations. In a previous attempt to meet this urgent need, our laboratory has engineered a Ca\(^{2+}\) sensor, “G1”, by grafting an EF-hand motif
into enhanced green fluorescent protein (EGFP) (Zou et al., 2007). Unlike GFP that can be excited at 395 nm and 475 nm, EGFP contains two mutations F64L and S65T and has one absorption maximum at 488 nm (Tsien, 1998, Ormo et al., 1996). Mutation F64L is responsible for improved folding efficiency at 37 °C, while S65T is a critical mutation for suppressing the 395 nm absorbance peak (Tsien, 1998, Ormo et al., 1996, Arpino et al., 2012). This G1 sensor has an apparent $K_\text{d}$ of 0.8 mM and responds to $\text{Ca}^{2+}$ with a ratiometric fluorescence change, but with a slow kinetic response.

Recently, we reported a new strategy for creating $\text{Ca}^{2+}$ indicators by introducing a $\text{Ca}^{2+}$-binding site into EGFP via site-directed mutagenesis of selected residues in the fluorescent sensitive location (Tang et al., 2011). The single EGFP-based $\text{Ca}^{2+}$ biosensor termed CatchER was generated by the substitutions of S147E, S202D, Q204E, F223E and T225E of EGFP in the designed $\text{Ca}^{2+}$ binding site (Figure 4.1). CatchER provides multiple advantages for reliably monitoring $\text{Ca}^{2+}$ signalling in high [$\text{Ca}^{2+}$] environments: (1) it exhibits a unique calcium-induced change in optical property. Calcium binding results in ratiometric changes in absorption while fluorescence emission at 510 nm is increased when excited at either 398 or 490 nm (Figure 4.2(a)); the high signal to noise ratio for fluorescent change in response to $\text{Ca}^{2+}$ as well as avoidance of cooperativity associated with multiple binding sites allows an accurate detection of calcium both in vitro and in vivo; (2) CatchER exhibits unprecedented dissociation kinetics with an off-rate $>100$ s$^{-1}$ and a fast kinetic response to $\text{Ca}^{2+}$ changes within milliseconds; recent work has also shown that CatchER is able to detect multiple calcium spikes during muscle contraction and relaxation (Wang et al., 2012); (3) the $K_\text{d}$ of CatchER (around 1 mM) allows the accurate calibration of SR $\text{Ca}^{2+}$ signalling; CatchER is able to report considerable differences in SR/ER $\text{Ca}^{2+}$ concentration between epithelial HeLa, kidney HEK293, and muscle C2C12 cells; (4) no invasive methods are required for CatchER detection in living organelles compared with current $\text{Ca}^{2+}$ dyes. Such cumulative advantages, especially fast kinetic properties, allowed us to monitor SR luminal $\text{Ca}^{2+}$ in flexor digitorum brevis (FDB) muscle fibers to understand the mechanism of diminished SR $\text{Ca}^{2+}$ release in aging mice (Tang et al., 2011, Wang et al., 2012).
In this report, we describe crystallographic analysis of CatchER to understand the structural basis for the calcium induced fluorescent and absorption changes and fast response. Crystal structures were determined of CatchER in the absence of Ca\(^{2+}\) (CatchER(apo)), in the presence of Ca\(^{2+}\) (CatchER-Ca\(^{2+}\)), and from crystals soaked with Gd\(^{3+}\) (CatchER-Gd\(^{3+}\)). To overcome the challenges for visualizing Ca\(^{2+}\) binding sites in the proteins due to weak Ca\(^{2+}\) binding affinity and high off rate and difficulty in distinguishing calcium from water in the crystal structure, we used the heavier Gd\(^{3+}\) ions with similar metal binding coordination properties to calcium to identify the position of the metal ion. These X-ray crystal structures of CatchER and its complexes can assist the future development of protein-ligand interaction based biosensors for the detection of various physiological molecules.

### 4.3 Experiments and Methods

#### 4.3.1 Expression and Purification

Protein was expressed by modification of the procedure described in (Zou et al., 2007). BL21DE3 cells transformed with pET28a vector containing the CatchER DNA were pre-cultured in 10 mL of LB media containing 6 \(\mu\)L of 50 mg/mL kanamycin (30 mg/mL) and shaken overnight at 37°C. The pre-culture was transferred to 1 L of Luria Bertani media containing 30 mg/mL of kanamycin and allowed to shake at 220 rpm at 37°C until the O.D. reached \(~0.6\) followed by the induction of protein expression with the addition of 200 \(\mu\)L of 1 M IPTG (0.2 mM) and a reduction in temperature to 25°C. Cells were collected by centrifugation at 7,000 rpm for 30 min at 4°C. Cell pellets were re-dissolved in extraction buffer (20 mM Tris pH 8.0, 100 mM NaCl, 0.1% Triton X-100) and sonicated. Cell lysate mixture was centrifuged for 30 minutes at 17,000 rpm and 4°C. Supernatant was filtered using a 0.45 \(\mu\)m Whattman filter and protein purified by using a 5 mL Hitrap chelating column (Amersham Biosciences, Sweden) loaded with Ni\(^{2+}\). The protein has a 6 histidine tag at the N-terminal for Immobilized Metal Ion Affinity Chromatography. High purity fractions were then concentrated to 1-2 mL and purified further using size exclusion chromatog-
raphy with a Superdex 75 100 mL column (Pharmacia Biotech) at a flow rate of 1 mL/min with 10 mM N-[2-hydroxyethyl]piperazine-N’-2-ethanesulfonic acid (HEPES) buffer pH 7.4 to ensure high purity for crystallization. The 6XHis-tag was not removed. The protein without the tag contains 244 amino acids and has ~30 kDa molecular mass.

4.3.2 Crystallization, X-ray Data Collection and Structure Determination

Crystals of Ca$^{2+}$ free and Ca$^{2+}$ loaded CatchER were obtained via the hanging drop method of vapor diffusion using 2 μL protein: 2 μL reservoir solutions at room temperature in 24-well VDX plates (Hampton Research, Aliso Viejo, CA). Ca$^{2+}$ free CatchER crystals (0.9 mM protein, 5 μM EGTA) grew in solutions containing 51 mM HEPES pH 7.0, 1 mM β-mercaptoethanol, 50 mM NaOAc, and 17% PEG 4000. The Ca$^{2+}$ loaded CatchER complex was created by adding 50 mM CaCl$_2$ to a 0.9 mM protein solution (final concentration of 0.45 mM). Crystals of Ca$^{2+}$ loaded CatchER grew in mother liquors containing 53 mM HEPES pH 7.0, 1 mM β-mercaptoethanol, 50 mM NaOAc, and 16% PEG 3350. Crystals of CatchER-Gd$^{3+}$ were obtained via the soaking technique. Crystals of apo CatchER were soaked for 1 to 2 days in a solution of mother liquor with the final concentration of 2 mM GdCl$_3$. The crystals were mounted in liquid nitrogen with 20% (v/v) glycerol as cryoprotectant in the mother liquor without added metal ions. X-ray diffraction data for the crystals were collected on the SER-CAT beamline of the Advanced Photon Source, Argonne National Laboratory, Argonne, IL. No anomalous scattering was detected for Gd$^{3+}$ or Ca$^{2+}$ likely due to the short X-ray wavelengths used.

X-ray diffraction data were processed and scaled with HKL2000 (Otwinowski & Minor, 1997), and the structures were solved by molecular replacement using MOLREP CPP4i suite of programs (Vagin & Teplyakov, 1997, Potterton et al., 2003) with the chain A of EGFP [Protein Data Bank (PDB) entry 2OKW] (Chapleau et al., 2008) as the starting model. The structures of CatchER(apo) and CatchER-Gd3+ were refined with REFMAC5 (Murshudov et al., 1997) and the near-atomic resolution structure of CatchER-Ca$^{2+}$ was refined with SHELX (Sheldrick & Schneider, 1997, Sheldrick, 2008). Manual adjustment
of the models used Coot 0.5.2 (Emsley & Cowtan, 2004). The full length protein has 238 amino acids, however, an N-terminal methionine was observed in these structures and labelled residue 0 for consistency with other published GFP structures, and several C-terminal residues were not visible in the electron density maps. In the three structures, a single protein molecule were refined with the residues labelled 0-231 in CatchER(apo) and CatchER-Gd\(^{3+}\) and 0-229 in CatchER-Ca\(^{2+}\), respectively, while the chromophore is labelled CRO66 as in other published GFP structures. Alternate conformations were modelled for CatchER residues where observed. The solvent was modelled with waters, metal ions and other solvent molecules according to the observed electron density maps. Anisotropic B factors were applied for the whole structures of CatchER-Ca\(^{2+}\) and only the Gd\(^{3+}\) atoms in CatchER-Gd\(^{3+}\). In the CatchER-Ca\(^{2+}\) and CatchER(apo) structures, residues R73 and E225 were refined with the same free variable number in SHELX or the same reasonable occupancy assignment in REFMACS due to the short distance between two alternative conformations. Residues 155-159 that show two alternate conformations in all three structures were refined in a similar manner. The mutant crystal structures were compared with each other and also with the wild type EGFP (PDB entry 4EUL), a recently published high resolution structure (Arpino et al., 2012), by superimposing their Cα atoms using SUPERPOSE in the CPP4i suite of programs (Krissinel & Henrick, 2004, Potterton et al., 2003). Structural figures were made using PyMol (DeLano, 2002). The atomic coordinates and structure factors have been deposited in the Protein Databank (PDB) with PDB codes of 4L13, 4L1I and 4L12 for CatchER(apo), CatchER-Ca\(^{2+}\) and CatchER-Gd\(^{3+}\), respectively.

4.3.3 Absorbance Spectrum of CatchER and Metal Binding Affinity via Fluorescence 

Spectroscopy

The absorbance spectra of CatchER and EGFP were obtained using a Shimadzu UV-1601 spectrophotometer. Samples were prepared using 20 µM protein in 10 mM Tris pH 7.4. To the CatchER sample, 5 µM of ethyleneglycoltetraacetic acid (EGTA) was added to get the Ca\(^{2+}\) free spectrum. To get the Ca\(^{2+}\) loaded CatchER spectrum, 5 mM CaCl\(_2\) was added to the sample. The absorbance spectrum of Gd\(^{3+}\)
loaded CatchER was obtained in the same manner with the addition of 200 μM GdCl₃ to the CatchER sample prepared in 20 mM PIPES pH 6.8. The fluorescence response of CatchER to Gd³⁺ and Ca²⁺ was analyzed using a Photon Technology International (PTI) (Canada) spectrofluorimeter. The spectra were collected with Felix32 fluorescence analysis software. Slit widths were set at 0.35 mm for excitation and emission. Samples of 10 μM CatchER with 2 μM EGTA were prepared in triplicate in 20 mM PIPES pH 6.8 for Gd³⁺ titrations and in 10 mM Tris pH 7.4 for Ca²⁺ titrations. Samples were excited at 395 and 488 nm with emission recorded from 500-600 nm. To determine the $K_d$, the following 1:1 binding equation was used,

$$\frac{[PM]}{[PT]} = \frac{[M_T]}{K_d + [M_T]}$$

where $K_d$ is the dissociation constant, $P_M/PT$ is the fractional change of complex formation, and $M_T$ is the total metal concentration. The equation was derived as follows:

$$[P_T] = [P_T] + [PM]$$

$$[M_T] = [M_T] + [PM]$$

$$[P_T] = [P_T] - [PM]$$

$$K_d = [P_T][M_T]/[PM]$$

$$K_d = (([P_T]-[PM])([M_T]))/([PM])$$

$$([PM])(K_d + [M_T]) = ([M_T][P_T])$$

$$([PM]/[P_T]) = [M_T]/K_d + [M_T]$$

Dividing by $[P_T]$ gives the fractional saturation of the protein. Because $[M_T] = [M_T]$, we can replace $[M_T]$ with $[M_T]$. 
4.4 Results and Discussion

4.4.1 Metal Binding Properties of CatchER

The absorption and fluorescence response of CatchER to Ca\textsuperscript{2+} and Gd\textsuperscript{3+} is shown in Figures 4.2(a)-(f) with fluorescence monitored at 510 nm upon excitation at 488 nm. The normalized response shows an excellent fit with the 1:1 binding equation. Figure 4.8(a)-(d) shows the fluorescence response and normalization for Gd\textsuperscript{3+} and Ca\textsuperscript{2+} at 395 nm excitation. From these results, the $K_d$ of CatchER for Gd\textsuperscript{3+} is $53.0 \pm 4.0 \mu$M and $177.0 \pm 13.6 \mu$M excited at 395 and 488 nm, respectively. The $K_d$ of CatchER for Ca\textsuperscript{2+} was determined to be $315.4 \pm 40.0 \mu$M and $227.0 \pm 3.3 \mu$M excited at 395 and 488 nm, respectively. The addition of Ca\textsuperscript{2+} or Gd\textsuperscript{3+} greatly enhances fluorescence emission of CatchER at 510 nm while excited at 395 or 488 nm.

The stoichiometric interaction of CatchER with Gd\textsuperscript{3+} was investigated and determined using the Job Plot method (Figure 4.2(f)). Figure 4.8(c) lists the calculated relative amounts of Gd\textsuperscript{3+} loaded CatchER from the plot in Figure 4.2(f). The largest amount of Gd\textsuperscript{3+} bound CatchER was obtained at equi-molar amounts of CatchER and Gd\textsuperscript{3+} (25 $\mu$M). Thus, CatchER forms 1:1 complex with Gd\textsuperscript{3+} (Table 4.1). In (Tang et al., 2011), CatchER was shown to exhibit 1:1 binding stoichiometry with Ca\textsuperscript{2+} as well. Overall, CatchER exhibits a 1:1 stoichiometric interaction with Gd\textsuperscript{3+} and Ca\textsuperscript{2+}.

4.4.2 Crystallographic Analysis of CatchER Structures

Crystal structures of CatchER in the apo form, Ca\textsuperscript{2+} form and Gd\textsuperscript{3+} loaded form were determined to identify and analyze the Ca\textsuperscript{2+} binding site in the designed sensor. The crystallographic data collection and refinement statistics are summarized in Table 4.2 and the data statistics versus resolution for each CatchER are listed in Table 4.4. The crystal structures of CatchER(apo), CatchER-Ca\textsuperscript{2+}, and CatchER-Gd\textsuperscript{3+} were refined to R factors of 18.1%, 15.0% and 19.6% at resolutions of 1.66, 1.20 and 1.78 Å, respectively. These three structures are in two different space groups: CatchER(apo) and CatchER-Gd\textsuperscript{3+} in C222\textsubscript{1}
and CatchER-Ca$^{2+}$ in $P_{2_1}2_12_1$. Structure validation was performed and the Ramachandran plots are shown in Figure 4.10. The p-hydroxylbenzilideneimidazolidinone chromophore (CRO66) is clearly visible in the electron density for all structures, as shown in Figure 4.3(a) for CatchER-Ca$^{2+}$. The three crystal structures have very similar backbone conformation, as demonstrated by low RMSD values of 0.09-0.20 Å for each pair of structures. Slightly more variation is seen relative to the EGFP structure (PDB code: 4EUL) with RMSD values of 0.39-0.41 Å.

Because of the high resolution of the diffraction data, the solvent was fitted with 167 waters for CatchER(apo), 197 water molecules for CatchER-Ca$^{2+}$, and 138 for CatchER-Gd$^{3+}$. One acetate molecule was refined with 1.0 occupancy in CatchER(apo), two Ca$^{2+}$ ions with 0.5 relative occupancy each in CatchER-Ca$^{2+}$, and two Gd$^{3+}$ ions with the relative occupancy of 0.7 and 0.3 in CatchER-Gd$^{3+}$. These molecules were identified by the shape and peak height in the electron density maps, B-factors and potential interactions with other molecules, as described in the next section. The occupancies were calculated with SHELX for CatchER-Ca$^{2+}$, and estimated with REFMAC5 for the other structures.

Alternative conformations were modelled for a total of 19, 23, and 13 residues in CatchER(apo), CatchER-Ca$^{2+}$ and CatchER-Gd$^{3+}$ structures, respectively. The surface loop of residues 155-159 shows two alternative conformations with about 0.5/0.5 relative occupancy in all three structures (Figure 4.9), while most other reported structures have a single conformation of these residues. This disordered loop is located on the opposite side of the protein to the designed metal binding site. Notably, E222 consistently shows two alternate conformations in CatchER-Ca$^{2+}$ and CatchER-Gd$^{3+}$ structures (Figure 4.3(b)) and a single conformation in CatchER(apo). Among the five designed mutations located on three neighboring β-strands, the side chain of E225 has two alternate conformations in CatchER(apo) and CatchER-Ca$^{2+}$ that interact with the two alternative conformations of R73 side chain. Due to the surface location of the five mutated residues and the potential for radiation damage to the carboxylate side chains, E204, E223 and E225 showed relatively poor electron density in the different CatchER complexes.
4.4.3 Identification of Metal Ions in Designed Binding Site of CatchER

CatchER was designed with five mutated residues S147E, S202D, Q204E, F223E, and T225E compared to EGFP. The mutations are located on three β-strands, pointing out of the protein β-barrel to form a penta-carboxylate ionic environment suitable for binding Ca$^{2+}$ (Figure 4.1). In CatchER(apo), one of the carboxyl oxygen atoms of E147 forms close interactions with the carboxylate of D202, suggesting possible protonation of the carboxylates (Figure 4.4(a)). Protonated carboxylates have been reported in other protein crystal structures (Tie et al., 2005). The E147 and D202 carboxylates are further apart with no direct interaction in the CatchER structures with Ca$^{2+}$ and Gd$^{3+}$ and additional solvent peaks were observed in this region of the electron density maps.

Identification of Ca$^{2+}$ in the designed binding site of CatchER was not trivial partly because of the fast off rate related to its weak calcium binding affinity. Crystals of CatchER were grown in high concentrations of 50 mM Ca$^{2+}$ to ensure saturation of CatchER molecules; however, the non-protein peaks in the electron density maps near the mutated residues were indistinguishable from those assigned to water elsewhere (Figure 4.4(b)). Therefore, the presence of Ca$^{2+}$ was deduced from the interactions with nearby protein residues and water molecules. Two possible locations for Ca$^{2+}$ were identified, mostly by the presence of shorter distances of 1.8-2.5 Å to interacting oxygen atoms of E147 and water molecules and further interactions with D202 and other water molecules in the designed site (Figure 4.4(b); Table 4.3). These interatomic distances are within the range observed in high resolution crystal structures of proteins (Harding, 2001). The two locations were fitted with Ca$^{2+}$ ions refined at partial occupancy (0.5/0.5). However, this deduction for Ca$^{2+}$ cannot exclude the possible binding of Na$^{+}$ or water molecules from the crystallization solution. No significant electron density was present in the apo structure at the positions assigned to Ca$^{2+}$ near E147 and D202.

In order to pinpoint the metal binding site more definitively, the structure of CatchER-Gd$^{3+}$ was obtained from apo crystals soaked in high concentrations of GdCl$_3$. Ca$^{2+}$ has 18 electrons orbiting the
nucleus, while the Gd$^{3+}$ has 61 orbital electrons. Therefore, it is easier to locate Gd$^{3+}$ with increased diffraction over Ca$^{2+}$ since X-ray atomic scattering factor increases with atomic number. The major 0.7 occupancy Gd$^{3+}$ ion was identified unambiguously from the very high peak at 22 $\sigma$ in the electron density indicative of a heavy metal ion (Figure 4.4(c)). This Gd$^{3+}$ ion is located between the side chains of E147, D202 and E204, forming four ionic interactions with these three residues at 2.2 Å, 2.5 Å, 2.6 Å and 2.6 Å distance and one with a nearby water molecule at 2.1 Å (Figure 4.4(c)) and Table 4.3. The second Gd$^{3+}$ ion with 0.3 occupancy was deduced from positive difference density observed in ($F_o-F_c$) maps when a water molecule or a partial occupancy Na$^+$ ion was refined at this site. Overall, the Gd$^{3+}$ ions coordinate with the side chains of residues E147, D202 and E204 of the designed Ca$^{2+}$ binding site as well as the water molecules.

Superposition of CatchER-Ca$^{2+}$ and CatchER-Gd$^{3+}$ structures revealed that the major occupancy site for the Gd$^{3+}$ ion is identical to one of the sites deduced for the Ca$^{2+}$ ion (Figure 4.4(d)). This Ca$^{2+}$ ion coordinates with the side chains of E147 and D202 and 3 water molecules. No Gd$^{3+}$ ion was visible at the other site where the Ca$^{2+}$ ion coordinates with the carboxylate side chains of E147 and D202 and 4 water molecules (Figure 4.4(b)). It is possible that the presence of the high occupancy Gd$^{3+}$ ion at the adjacent site precludes binding to the inner site occupied by a Ca$^{2+}$ ion in the CatchER-Ca$^{2+}$ structure.

The extended binding site formed by the carboxylate side chains of mutated residues E147, D202, E204, E223 and E225 traps metal ions at three possible positions, as shown by superposition of CatchER-Ca$^{2+}$ and CatchER-Gd$^{3+}$ (Figure 4.4(d)). The metal ions mainly interact with side chains of E147, D202 and E204. No direct interactions of designed metal ligand residues E223 and E225 with the metal ions are visible, and their side chains are not well defined in the electron density maps, possibly due to radiation damage. Nevertheless, both Ca$^{2+}$ and Gd$^{3+}$ ions are well situated at the designed Ca$^{2+}$ binding site in CatchER, which suggests that these X-ray structures provide snapshots for steps in the likely dynamic metal binding process.
4.4.4 Structural Changes around the Chromophore

The chromophore interactions were compared for the CatchER structures, EGFP (PDB code: 4EUL) (Arpino et al., 2012) and GFP (PDB code: 1EMB) (Brejc et al., 1997). The chromophore is buried centrally in the protein molecule and well protected from solvent. It can exist as neutral and anionic forms, which are responsible for the absorbance at 395 nm and 475 nm, respectively (Yang et al., 1996) (Figure 4.5). The spectroscopic characterization of CatchER and its response to Ca$^{2+}$ shows two absorption maxima with a major peak at 398 nm and a smaller peak at 490 nm, and thus, it resembles GFP with two similar excitation wavelengths unlike EGFP with the one single excitation peak at 488 nm (Tsien, 1998) (Figure 4.2(e)). Comparisons of the chromophore environment in the three CatchER structures and currently solved EGFP and GFP structures have shed light on the relationship between the spectroscopic properties and structures. The intricate hydrogen bond networks around the chromophores of CatchER(a apo), CatchER-Ca$^{2+}$ or Gd$^{3+}$, EGFP and GFP are shown in Figures 4.6(a)-4.6(d) (Arpino et al., 2012, Brejc et al., 1997). Most hydrogen bond interactions are conserved in the vicinity of the carbonyl group of the imidazolidinone ring of the chromophore (Arpino et al., 2012, Brejc et al., 1997, Ormo et al., 1996). Two residues, T203 and E222, show a critical role in the chemical environment for chromophore. T203 has been observed with two different conformations: in one, the side chain of T203 can have direct contact with the tyrosyl group of the chromophore; in the other, the side chain of T203 rotates away from the tyrosyl group and the main chain moves towards the chromophore resulting in the elimination of direct hydrogen bond interactions but permitting a water-mediated interaction between the main chain carbonyl group and the chromophore tyrosyl group. In addition, the side chain of E222 has shown two alternate conformations in some EGFP structures (Royant et al., 2011, Arpino et al., 2012).

In our structural analysis and comparisons, CatchER-Ca$^{2+}$ and CatchER-Gd$^{3+}$ have similar interactions around the chromophore as shown schematically in Figure 4.6(b), while CatchER(a apo) has different
interactions for E222 (Figure 4.6(a)). In all the three CatchER structures, T203 formed a water-mediated hydrogen bond with the chromophore tyrosyl via the second type of conformation mentioned above. The representative omit map of D202-D206 adjacent to CRO66 and the mediating water for hydrogen bond in CatchER(apo) is shown in Figure 4.3(c). This type of interaction between the carbonyl group of T203 and chromophore was also found in the GFP structure (PDB code: 1EMB) (Figure 4.6(d)), even though it also has an alternate side chain conformation with 0.15 occupancy which can form the direct hydrogen bonds with chromophore (Brejc et al., 1997). In contrast, in the EGFP structure (PDB code: 4EUL), T203 only forms the first type of interaction (Figure 4.6(c)).

Regarding the side chain of E222, the conformational population differs in the three CatchER structures: only one conformation of E222 was observed in CatchER(apo) and its side chain is considered to be deprotonated and forms a hydrogen bond with S205 and another one with the hydroxyl group of chromophore (Figure 4.6(a)); in CatchER-Ca<sup>2+</sup> and CatchER-Gd<sup>3+</sup>, one additional alternative conformation of E222 was determined which lacks interactions with chromophore; instead it participates in interactions with a network of water molecules linking to Q69 (Figure 4.6(b)). Recent crystallographic study of EGFP has revealed two alternate conformations for E222 not only in the EGFP structure (PDB code: 4EUL) used here for comparison but in another reported structure (PDB code of 2YOG) (Royant et al., 2011). In both the published structures, E222 shows similar interactions with chromophore and surrounding environment as in Figure 4.6(c), which is quite similar to the arrangement in CatchER-Ca<sup>2+</sup> and CatchER-Gd<sup>3+</sup>. However, in the GFP structure (PDB code: 1EMB), one conformation of E222 was defined forming hydrogen bond interactions like in CatchER(apo).

Previously, the profound but opposite effects of residues T203 and E222 were reported from mutagenesis and analysis of crystal structures (Heim et al., 1994, Ehrig et al., 1995, Brejc et al., 1997). Introduction of T203I in GFP retains the 395 nm peak but eliminates the 475 nm peak, whereas GFP with E222G retains the 475 nm peak but lacks the 395 nm peak (Heim et al., 1994, Ehrig et al., 1995). Crystal
structures of GFP and EGFP revealed that the side chain of T203 can stabilize a negative charge on the chromophore (anionic form chromophore) as a hydrogen bond donor through a direct hydrogen bond with the chromophore tyrosyl, but the carboxylate of charged E222 can maintain the neutral form of chromophore through electrostatic repulsion and the hydrogen-bonding network via water and S205 (Brejc et al., 1997). In the CatchER structures, the conformation of T203 preferred the proposed protonated form of the chromophore. The hydrogen-bond network via water and S205 is achieved by one conformation of E222 in deprotonated or negatively charged states and helps maintain the neutral form of chromophore. In CatchER-Ca\(^{2+}\) and CatchER-Gd\(^{3+}\), even though the other conformation of E222 does not interact with the chromophore threonine as in EGFP (Ormo et al., 1996, Royant et al., 2011, Arpino et al., 2012), this alternate conformation with almost half occupancy is proposed to be protonated and can no longer maintain the neutral form of chromophore. Therefore, although there is only one conformation of E222 in CatchER(apo), we suggest that the chromophore of CatchER has a mixture of neutral and negatively charged states, as observed in wild type GFP, which has two conformations of T203 (Brejc et al., 1997).

The ratio of neutral and negatively charged chromophore can differ in the CatchER structures based on the spectroscopic properties. In the absorbance spectra, addition of Ca\(^{2+}\) results in a concurrent increased intensity around 490 nm (increase of deprotonated chromophore) and a decreased absorption intensity around 398 nm (decrease of protonated chromophore state), exhibiting an optical spectral feature more closely resembling that of EGFP, as shown in Figure 4.2(a). Tang et al. has also reported that the presence of calcium results in a decrease in the pKa value of the chromophore of CatchER (Tang et al., 2011), suggesting that the chromophore in CatchER is more deprotonated upon calcium binding. This change could be related to the observed equilibrium shift in conformational populations of E222 in CatchER: the increase of deprotonated chromophore on calcium binding is likely due to the stabilizing capability of E222 as a hydrogen bond donor with the hydrogen bond acceptor of deprotonated T65 in the chromophore (Figure 4.6(b)). Such stabilizing effect is manifested by the
protonated T65 in the chromophore (Figure 4.6(b)). Such stabilizing effect is manifested by the changes in the proton wire network between apo- and calcium-loaded forms of CatchER and EGFP as shown in Figures 4.6(a)-(c). Thus, binding of calcium induces an equilibrium shift in conformational populations of the E222 side chain and its network of interactions through two water molecules to Q69. Overall, the crystal structures of CatchER have reinforced support for the proposed excited-state photon transfer pathway for the photoisomerization of GFP, which was based on structural and spectroscopic studies (Chattoraj et al., 1996, Brejc et al., 1997). The similar observation and proposed interpretation apply to Gd$^{3+}$ induced change (Figure 4.2(b) and Figure 4.6(b)).

No significant difference was observed for L42, T43, Y143 and T153 in the structural comparison of CatchER(apo) with CatchER-Ca$^{2+}$, despite the chemical shift changes related to Ca$^{2+}$ binding (Tang et al., 2011) shown in dynamic NMR. The addition of Ca$^{2+}$ leads to splitting of one resonance gradually into two for Q69, which is buried inside the protein (Tang et al., 2011). In both CatchER-Ca$^{2+}$ and CatchER-Gd$^{3+}$, Q69 forms a hydrogen bond network through two water molecules with E222 and this network was also found in EGFP (PDB code: 4EUL) (Figure 4.6(b)). However, this network was interrupted due to a missing water molecule in CatchER(apo) (Figure 4.6(a)). Therefore, here we propose that the calcium induced change in optical properties could be also associated with the Q69 hydrogen bond network.

### 4.4.5 Relationship between Mutations of the Novel Metal Binding Site and Optical Properties

CatchER was selected from a series of Ca$^{2+}$ sensors designed by introducing different mutations around the desired Ca$^{2+}$ binding site, designated D8, D9, D10, D11 (CatchER) and D12 (Tang et al., 2011). These mutants all show an increase in the peak at 398 nm and a decrease at 490 nm to different extents. D8 has only two mutations of S202D and F223E, while the other variants have additional mutations, which indicated that S202D and/or F223E might play an important role in the conserved absorbance changes for the designed proteins. Based on the structural changes around the chromophore,
T203 and E222 are two key residues that account for the altered chemical environment of the chromophore. The effects of these two residues might be modified by the mutations of adjacent residues S202D and F223E. However, F223E has no direct interaction with Ca$^{2+}$ in CatchER-Ca$^{2+}$, although it forms a water-mediated hydrogen bond with one Gd$^{3+}$ in CatchER-Gd$^{3+}$ structure. Instead, mutation S147E appears to function as the anion for interacting with metal cations, while S202D is also involved in metal coordination based on our crystal structures. The main chain of residues 202 to 206 has shifted by 0.7 to 0.9 Å in association with the rotation of T203 in CatchER relative to EGFP, which leads to the ionization change of the chromophore (Figure 4.7).

4.5 Conclusions

The binding of metal ions Ca$^{2+}$ and Gd$^{3+}$ to the designed calcium sensor CatchER has been investigated by spectroscopic methods and X-ray crystallography. Both the kinetic assays and structures demonstrated binding of the two types of metal ions to CatchER, however, there were unexpected differences. Crystal structures of CatchER in the apo form and in its complexes with Ca$^{2+}$ and Gd$^{3+}$ reveal snapshots of the dynamic binding of metal ions to the designed site comprising five carboxylate side chains. Both Ca$^{2+}$ and Gd$^{3+}$ ions were observed in two locations within the designed binding site. The high (mM) concentrations of Ca$^{2+}$ and Gd$^{3+}$ used to obtain the crystal structures of their CatchER complexes resulted in two alternative binding sites for each metal ion with one central common binding site. In solution, however, these two metal ions bind CatchER with 1:1 stoichiometry and micromolar affinity. These structures suggest that the ability of Ca$^{2+}$ ions to jump between two possible binding sites may be partly responsible for the fast kinetics of metal ion binding to CatchER.

4.6 Acknowledgements

We thank Drs. Johnson Agniswamy, Yanyi Chen, and Huiming Li and Malcolm Degrado for valuable discussions. Data were collected at the Southeast Regional Collaborative Access Team (SER-CAT)
beamline at the Advanced Photon Source, Argonne National Laboratory. Supporting institutions may be found at www.ser-cat.org/members.html. Use of the Advanced Photon Source was supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under Contract No. W-31-109-Eng-38. This work was supported, in part, by NIH grants GM081749 NIH and EB007268 to JJY, a seed grant from Georgia State University Brains & Behavior Area of Focus, and a University Doctoral Fellowship in Diagnostics and Therapy fellowship (Y. Z.), and Brains & Behavior graduate fellowship to S.T and Y.Z.
Table 4.1 Relative amounts of CatchER-Gd complex formation with corresponding ratios

<table>
<thead>
<tr>
<th>[CatchER] (μM)</th>
<th>[Gd$^{3+}$] (μM)</th>
<th>[CatchER-Gd] Abs 493 nm</th>
<th>[CatchER-Gd] Fluo 395 nm</th>
<th>[CatchER-Gd] Fluo 488 nm</th>
<th>[CatchER]/[Gd$^{3+}$]</th>
</tr>
</thead>
<tbody>
<tr>
<td>40</td>
<td>10</td>
<td>11.1±2.3</td>
<td>6.7 ± 4.4</td>
<td>6.5±3.7</td>
<td>4.0</td>
</tr>
<tr>
<td>35</td>
<td>15</td>
<td>18.0±1.4</td>
<td>17.5±1.2</td>
<td>16.4±1.0</td>
<td>2.3</td>
</tr>
<tr>
<td>30</td>
<td>20</td>
<td>23.0±0.3</td>
<td>22.8±1.0</td>
<td>20.5±0.7</td>
<td>1.5</td>
</tr>
<tr>
<td>25</td>
<td>25</td>
<td>30.8±3.5</td>
<td>32.7±0.5</td>
<td>20.0±1.7</td>
<td>1.0</td>
</tr>
<tr>
<td>20</td>
<td>30</td>
<td>15.7±2.2</td>
<td>23.6±0.0</td>
<td>14.4±0.5</td>
<td>0.6</td>
</tr>
<tr>
<td>15</td>
<td>35</td>
<td>10.4±0.3</td>
<td>19.5±0.8</td>
<td>11.4±0.5</td>
<td>0.4</td>
</tr>
<tr>
<td>10</td>
<td>40</td>
<td>7.2±0.0</td>
<td>13.7±0.6</td>
<td>7.8±0.3</td>
<td>0.2</td>
</tr>
</tbody>
</table>
Table 4.2 Crystallographic data and refinement statistics

<table>
<thead>
<tr>
<th>Complex Name</th>
<th>CatchER (apo)</th>
<th>CatchER-Ca^{2+}</th>
<th>CatchER-Gd^{3+}</th>
</tr>
</thead>
<tbody>
<tr>
<td>X-ray source</td>
<td>APS_22ID</td>
<td>APS_22ID</td>
<td>APS_22BM</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>0.8</td>
<td>0.8</td>
<td>1.0</td>
</tr>
<tr>
<td>Temperature (K)</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Space group</td>
<td>C222,1</td>
<td>P2₁,2,1,2</td>
<td>C222,1</td>
</tr>
<tr>
<td>a (Å)</td>
<td>61.40</td>
<td>54.24</td>
<td>61.06</td>
</tr>
<tr>
<td>b (Å)</td>
<td>88.53</td>
<td>61.06</td>
<td>88.33</td>
</tr>
<tr>
<td>c (Å)</td>
<td>118.43</td>
<td>67.40</td>
<td>118.1</td>
</tr>
<tr>
<td>Protein molecules per asymmetric unit</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Unique reflections</td>
<td>38048</td>
<td>70,349</td>
<td>30,260</td>
</tr>
<tr>
<td>R&lt;sub&gt;merge&lt;/sub&gt; (%) overall&lt;sup&gt;a&lt;/sup&gt;</td>
<td>6.8% (22.8%)</td>
<td>8.1% (39.2%)</td>
<td>8.5% (18.5%)</td>
</tr>
<tr>
<td>&lt;I/σ(I) &gt; overall&lt;sup&gt;b&lt;/sup&gt;</td>
<td>22.6 (7.7)</td>
<td>19.9 (6.1)</td>
<td>14.2 (8.7)</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>31.08 – 1.66</td>
<td>50 – 1.20</td>
<td>27.13 – 1.78</td>
</tr>
<tr>
<td>Completeness (%) overall&lt;sup&gt;c&lt;/sup&gt;</td>
<td>99.1% (93.6%)</td>
<td>99.2% (100%)</td>
<td>98.3% (93.8%)</td>
</tr>
<tr>
<td>R&lt;sub&gt;work&lt;/sub&gt;</td>
<td>0.181</td>
<td>0.150</td>
<td>0.196</td>
</tr>
<tr>
<td>R&lt;sub&gt;free&lt;/sub&gt;</td>
<td>0.203</td>
<td>0.191</td>
<td>0.219</td>
</tr>
<tr>
<td>No. of protein atoms (includes alternative conformations)</td>
<td>1956</td>
<td>1967</td>
<td>1922</td>
</tr>
<tr>
<td>No. of H₂O molecules (total occupancies&lt;sup&gt;b&lt;/sup&gt;)</td>
<td>167 (151)</td>
<td>197 (191.5)</td>
<td>138 (128.5)</td>
</tr>
<tr>
<td>No. of ions (occupancy)</td>
<td>1 (1.0)</td>
<td>2 (0.5/0.5)</td>
<td>2 (0.7/0.3)</td>
</tr>
<tr>
<td>RMS deviation from ideality</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bonds (Å)</td>
<td>0.012</td>
<td>0.012</td>
<td>0.014</td>
</tr>
<tr>
<td>Average B-factors (Å²)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Main chain atoms</td>
<td>17.46</td>
<td>16.60</td>
<td>20.40</td>
</tr>
<tr>
<td>Side chain atoms</td>
<td>19.34</td>
<td>20.52</td>
<td>22.01</td>
</tr>
<tr>
<td>H₂O</td>
<td>25.47</td>
<td>28.80</td>
<td>25.79</td>
</tr>
<tr>
<td>Ions</td>
<td>43.74</td>
<td>35.67</td>
<td>37.56</td>
</tr>
<tr>
<td>Ramachandran plots results</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of residues in favored region</td>
<td>221 (98.2%)</td>
<td>220 (98.7%)</td>
<td>221 (98.2%)</td>
</tr>
<tr>
<td>Number of residues in allowed region</td>
<td>4 (1.8%)</td>
<td>3 (1.3%)</td>
<td>4 (1.8%)</td>
</tr>
</tbody>
</table>

<sup>a</sup>Values in parentheses are given for the highest resolution shell
<sup>b</sup>Total occupancies are the sum of calculated occupancies of all the atoms or ions.
<sup>c</sup>The angle rmsd in SHELX97 is indicated by distance in Å
<sup>d</sup>The angle rmsd in REFMAC 5.2 is indicated by angle in degrees
Table 4.3 CatchER-metal interactions within proposed coordination site

<table>
<thead>
<tr>
<th>CatchER atoms</th>
<th>B factors(Å²) /Occupancy</th>
<th>Distance (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Ca²⁺(1)</td>
</tr>
<tr>
<td>E147/OE2</td>
<td>30.76/1.0</td>
<td>1.9</td>
</tr>
<tr>
<td>D202/OD1</td>
<td>18.84/1.0</td>
<td>3.0</td>
</tr>
<tr>
<td>D202/OD2</td>
<td>27.29/1.0</td>
<td>2.6</td>
</tr>
<tr>
<td>H₂O1</td>
<td>35.36/1.0</td>
<td>3.1</td>
</tr>
<tr>
<td>H₂O2</td>
<td>31.95/1.0</td>
<td>2.5</td>
</tr>
<tr>
<td>H₂O3</td>
<td>25.38/1.0</td>
<td>1.8</td>
</tr>
<tr>
<td>H₂O4</td>
<td>44.32/0.5</td>
<td>2.5</td>
</tr>
<tr>
<td>H₂O5</td>
<td>25.02/1.0</td>
<td>2.1</td>
</tr>
<tr>
<td>H₂O6</td>
<td>42.44/1.0</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CatchER atoms</th>
<th>B factors(Å²) /Occupancy</th>
<th>Distance (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Gd³⁺(1)</td>
</tr>
<tr>
<td>E147/OE2</td>
<td>30.51/1.0</td>
<td>2.2</td>
</tr>
<tr>
<td>D202/OD1</td>
<td>22.52/1.0</td>
<td>2.6</td>
</tr>
<tr>
<td>D202/OD2</td>
<td>30.51/1.0</td>
<td>2.5</td>
</tr>
<tr>
<td>E204/OE1</td>
<td>32.69/1.0</td>
<td>2.6</td>
</tr>
<tr>
<td>E204/OE2</td>
<td>29.53/1.0</td>
<td></td>
</tr>
<tr>
<td>H₂O1</td>
<td>33.89/1.0</td>
<td>2.1</td>
</tr>
<tr>
<td>H₂O2</td>
<td>24.74/1.0</td>
<td></td>
</tr>
<tr>
<td>H₂O3</td>
<td>33.62/1.0</td>
<td></td>
</tr>
</tbody>
</table>
### Table 4.4 Data statistics vs resolution

**CatchER(apo)**

<table>
<thead>
<tr>
<th>Resolution Å</th>
<th>Lower limit</th>
<th>Upper limit</th>
<th>Average I</th>
<th>Average error</th>
<th>Norm. stat</th>
<th>Linear Chi**2</th>
<th>Square R-fac</th>
<th>R-fac</th>
</tr>
</thead>
<tbody>
<tr>
<td>50.00</td>
<td>3.58</td>
<td>5088.0</td>
<td>187.5</td>
<td>37.2</td>
<td>0.971</td>
<td>0.052</td>
<td>0.064</td>
<td></td>
</tr>
<tr>
<td>3.58</td>
<td>2.84</td>
<td>2732.9</td>
<td>116.4</td>
<td>21.1</td>
<td>0.968</td>
<td>0.066</td>
<td>0.076</td>
<td></td>
</tr>
<tr>
<td>2.84</td>
<td>2.48</td>
<td>1118.7</td>
<td>50.3</td>
<td>13.5</td>
<td>0.974</td>
<td>0.072</td>
<td>0.081</td>
<td></td>
</tr>
<tr>
<td>2.48</td>
<td>2.25</td>
<td>754.5</td>
<td>33.6</td>
<td>12.6</td>
<td>1.050</td>
<td>0.076</td>
<td>0.081</td>
<td></td>
</tr>
<tr>
<td>2.25</td>
<td>2.09</td>
<td>584.7</td>
<td>32.0</td>
<td>12.6</td>
<td>0.931</td>
<td>0.088</td>
<td>0.091</td>
<td></td>
</tr>
<tr>
<td>2.09</td>
<td>1.97</td>
<td>427.2</td>
<td>25.7</td>
<td>12.3</td>
<td>0.960</td>
<td>0.103</td>
<td>0.107</td>
<td></td>
</tr>
<tr>
<td>1.97</td>
<td>1.87</td>
<td>277.3</td>
<td>19.7</td>
<td>11.7</td>
<td>0.956</td>
<td>0.123</td>
<td>0.119</td>
<td></td>
</tr>
<tr>
<td>1.87</td>
<td>1.79</td>
<td>176.9</td>
<td>14.8</td>
<td>11.3</td>
<td>1.041</td>
<td>0.156</td>
<td>0.147</td>
<td></td>
</tr>
<tr>
<td>1.79</td>
<td>1.72</td>
<td>120.3</td>
<td>13.2</td>
<td>11.2</td>
<td>1.015</td>
<td>0.197</td>
<td>0.183</td>
<td></td>
</tr>
<tr>
<td>1.72</td>
<td>1.66</td>
<td>95.3</td>
<td>12.4</td>
<td>10.9</td>
<td>0.987</td>
<td>0.228</td>
<td>0.224</td>
<td></td>
</tr>
<tr>
<td>All reflections</td>
<td></td>
<td></td>
<td>1171.9</td>
<td>51.8</td>
<td>15.6</td>
<td>0.985</td>
<td>0.068</td>
<td>0.068</td>
</tr>
</tbody>
</table>

**CatchER-Ca**

<table>
<thead>
<tr>
<th>Resolution Å</th>
<th>Lower limit</th>
<th>Upper limit</th>
<th>Average I</th>
<th>Average error</th>
<th>Norm. stat</th>
<th>Linear Chi**2</th>
<th>Square R-fac</th>
<th>R-fac</th>
</tr>
</thead>
<tbody>
<tr>
<td>50.00</td>
<td>2.59</td>
<td>15117.6</td>
<td>803.1</td>
<td>245.3</td>
<td>0.901</td>
<td>0.060</td>
<td>0.071</td>
<td></td>
</tr>
<tr>
<td>2.59</td>
<td>2.05</td>
<td>4838.3</td>
<td>172.0</td>
<td>34.5</td>
<td>1.098</td>
<td>0.074</td>
<td>0.084</td>
<td></td>
</tr>
<tr>
<td>2.05</td>
<td>1.79</td>
<td>2084.5</td>
<td>106.1</td>
<td>23.6</td>
<td>0.955</td>
<td>0.098</td>
<td>0.110</td>
<td></td>
</tr>
<tr>
<td>1.79</td>
<td>1.63</td>
<td>811.5</td>
<td>39.3</td>
<td>11.3</td>
<td>1.094</td>
<td>0.128</td>
<td>0.147</td>
<td></td>
</tr>
<tr>
<td>1.63</td>
<td>1.51</td>
<td>448.4</td>
<td>23.4</td>
<td>9.4</td>
<td>1.124</td>
<td>0.157</td>
<td>0.175</td>
<td></td>
</tr>
<tr>
<td>1.51</td>
<td>1.42</td>
<td>262.5</td>
<td>16.1</td>
<td>8.6</td>
<td>1.122</td>
<td>0.210</td>
<td>0.224</td>
<td></td>
</tr>
<tr>
<td>1.42</td>
<td>1.35</td>
<td>163.4</td>
<td>12.6</td>
<td>8.5</td>
<td>1.077</td>
<td>0.281</td>
<td>0.295</td>
<td></td>
</tr>
<tr>
<td>1.35</td>
<td>1.29</td>
<td>120.5</td>
<td>11.3</td>
<td>8.8</td>
<td>1.108</td>
<td>0.317</td>
<td>0.331</td>
<td></td>
</tr>
<tr>
<td>1.29</td>
<td>1.24</td>
<td>87.4</td>
<td>10.7</td>
<td>9.2</td>
<td>1.063</td>
<td>0.351</td>
<td>0.370</td>
<td></td>
</tr>
<tr>
<td>1.24</td>
<td>1.20</td>
<td>64.1</td>
<td>10.5</td>
<td>9.5</td>
<td>1.013</td>
<td>0.392</td>
<td>0.408</td>
<td></td>
</tr>
<tr>
<td>All reflections</td>
<td></td>
<td></td>
<td>2389.8</td>
<td>119.8</td>
<td>36.6</td>
<td>1.061</td>
<td>0.081</td>
<td>0.076</td>
</tr>
</tbody>
</table>

**CatchER-Gd**

<table>
<thead>
<tr>
<th>Resolution Å</th>
<th>Lower limit</th>
<th>Upper limit</th>
<th>Average I</th>
<th>Average error</th>
<th>Norm. stat</th>
<th>Linear Chi**2</th>
<th>Square R-fac</th>
<th>R-fac</th>
</tr>
</thead>
<tbody>
<tr>
<td>50.00</td>
<td>3.83</td>
<td>2142.8</td>
<td>130.3</td>
<td>21.5</td>
<td>0.951</td>
<td>0.065</td>
<td>0.071</td>
<td></td>
</tr>
<tr>
<td>3.83</td>
<td>3.04</td>
<td>1484.5</td>
<td>109.5</td>
<td>15.9</td>
<td>0.980</td>
<td>0.086</td>
<td>0.088</td>
<td></td>
</tr>
<tr>
<td>3.04</td>
<td>2.66</td>
<td>611.6</td>
<td>47.1</td>
<td>8.3</td>
<td>0.951</td>
<td>0.091</td>
<td>0.094</td>
<td></td>
</tr>
<tr>
<td>2.66</td>
<td>2.42</td>
<td>379.2</td>
<td>29.5</td>
<td>6.3</td>
<td>0.912</td>
<td>0.093</td>
<td>0.092</td>
<td></td>
</tr>
<tr>
<td>2.42</td>
<td>2.24</td>
<td>283.8</td>
<td>22.6</td>
<td>5.9</td>
<td>0.930</td>
<td>0.102</td>
<td>0.101</td>
<td></td>
</tr>
<tr>
<td>2.24</td>
<td>2.11</td>
<td>217.9</td>
<td>16.8</td>
<td>5.3</td>
<td>1.048</td>
<td>0.115</td>
<td>0.107</td>
<td></td>
</tr>
<tr>
<td>2.11</td>
<td>2.00</td>
<td>171.4</td>
<td>13.7</td>
<td>5.0</td>
<td>0.973</td>
<td>0.118</td>
<td>0.108</td>
<td></td>
</tr>
<tr>
<td>2.00</td>
<td>1.92</td>
<td>122.6</td>
<td>10.2</td>
<td>4.7</td>
<td>0.989</td>
<td>0.131</td>
<td>0.114</td>
<td></td>
</tr>
<tr>
<td>1.92</td>
<td>1.84</td>
<td>78.8</td>
<td>7.1</td>
<td>4.4</td>
<td>1.004</td>
<td>0.162</td>
<td>0.149</td>
<td></td>
</tr>
</tbody>
</table>
This table is from SCALEPACK

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1.84</td>
<td>1.78</td>
<td>52.8</td>
<td>6.1</td>
<td>4.7</td>
<td>1.008</td>
<td>0.185</td>
</tr>
<tr>
<td>All reflections</td>
<td>572.3</td>
<td>40.4</td>
<td>8.4</td>
<td>0.974</td>
<td>0.085</td>
<td>0.079</td>
</tr>
</tbody>
</table>

This table is from SCALEPACK
Figure 4.1 Structure of CatchER (green cartoon) indicating the locations of mutated residues (red sticks) S147E, S202D, Q204E, F223E and T225E, and chromophore CRO66 in green sticks colored by CPK.
Figure 4.2 Absorbance spectra of CatchER and metal binding analysis via fluorescence spectroscopy.

(a) Absorbance response of CatchER to 5 mM Ca\(^{2+}\). 20 µM CatchER with 5 µM EGTA (dashed line) in 10 mM Tris pH 7.4 and the same CatchER sample with 5 mM Ca\(^{2+}\) added (solid line). Addition of Ca\(^{2+}\) produces more of the anionic chromophore (increase in 488 nm absorbance peak) and less of the neutral chromophore (decrease in the 395 nm absorbance peak). (b) Absorbance response of CatchER to 200 µM Gd\(^{3+}\). A 20 µM sample of CatchER was prepared in 20 mM PIPES pH 6.8. The dashed line represents the sample with 5 µM EGTA and the solid line is the same sample with 200 µM Gd\(^{3+}\) added. (c and d) Fluorescence response of CatchER to Ca\(^{2+}\) and Gd\(^{3+}\) excited at 488 nm with emission at 510 nm with inset binding curves. Normalized fluorescence response for Ca\(^{2+}\) and Gd\(^{3+}\) was fitted with a 1:1 binding equation producing $K_D$ values of $227.0 \pm 3.3$ µM and $177.0 \pm 13.6$ µM for Ca\(^{2+}\) and Gd\(^{3+}\), respectively. (e) Overlay of the absorbance spectrum of 20 µM CatchER with 5 µM EGTA (dashed line) and 20 µM EGFP (solid line) in 10 mM Tris pH 7.4. (f) Binding stoichiometry of CatchER to Gd\(^{3+}\) via Job Plot. The relative amount of CatchER bound to Gd\(^{3+}\) was determined using fluorescence and absorbance intensity changes in the absence and presence of Gd\(^{3+}\). Concentrations of CatchER were 40, 35, 30, 25, 20, 15, and 10 µM (actual concentrations via 280 nm absorbance peak were 38, 35, 30, 24, 20, 15, and 10.5 µM). The total molar ratio was held constant at 50 µM. Plot represents fluorescence data of complex formation from 395 nm excitation.
Figure 4.3 Selected $F_o-F_c$ omit map in CatchER-Ca$^{2+}$ CatchER-Gd$^{3+}$ and CatchER(apo). (a) $F_o-F_c$ omit map showing the chromophore from CatchER-Ca$^{2+}$ contoured at 4.0σ. (b) $F_o-F_c$ omit maps showing two conformations of E222 from CatchER-Gd$^{3+}$ and two water molecules contoured at 3.0σ. (c) $F_o-F_c$ omit map showing residues from D202 to A206 from CatchER(apo) and one water molecule contoured at 3.0σ with chromophore (CRO66). The dotted lines represent hydrogen bond interactions.
Figure 4.4 Hydrogen bond interactions at the designed Ca$^{2+}$ binding site in CatchER structures and superposition of CatchER-Ca$^{2+}$ and CatchER-Gd$^{3+}$. (a)-(c) The hydrogen bond interactions at the designed Ca$^{2+}$ binding site in CatchER structures. The protein is represented as yellow, light blue and cyan sticks in CatchER(apo), CatchER-Ca$^{2+}$, and CatchER-Gd$^{3+}$, respectively. Ca$^{2+}$, Gd$^{3+}$ and water molecules are represented as spheres. The numbers (0.5/0.5, 0.7/0.3) give the relative occupancy of the alternate positions of the Ca$^{2+}$, Gd$^{3+}$ ions, respectively. The interatomic (non-hydrogen) distance range of 2.6-3.2 Å was used for hydrogen bonds (black dotted lines). Shorter distances in the range of 2.0-2.4 Å (red dotted lines) suggest coordination with a metal ion. The coordinating waters are numbered as in Table 3. (d) Superposition of CatchER-Ca$^{2+}$ and CatchER-Gd$^{3+}$; the protein backbones are shown as grey cartoons and the chromophore (CRO66) is represented as green sticks. The five mutated residues and ions are shown in sticks and spheres in lightblue for CatchER-Ca$^{2+}$ and cyan for CatchER-Gd$^{3+}$. The cyan arrow points to the major site for Gd$^{3+}$. 
Figure 4.5 Protonation states of the GFP chromophore with the corresponding absorbance wavelength.
Figure 4.6 Scheme of the hydrogen bond interactions around chromophore in CatchER structures, EGFP and GFP. (a)-(d) Scheme of the hydrogen bond interactions between chromophore and surrounding residues and water molecules (W) in CatchER(apo), CatchER-Ca\(^{2+}\) or Gd\(^{3+}\), EGFP and GFP. Hydrogen bonds are shown as dashed lines. The interatomic (non-hydrogen) distance range of 2.6-3.2 Å was used for hydrogen bonds.
Figure 4.7 Comparison of EGFP and CatchER(apo) at the chromophore and nearby residues. Residues 202 to 206 of EGFP and CatchER(apo) are shown as sticks in grey and yellow, respectively. S202/D202 and Q204/E204 label mutated residues in EGFP/CatchER(apo). The main chain from residues 202 to A206 shifts with a maximum value of 0.8 Å as indicated by the arrow.
Figure 4.8 Fluorescence response of CatchER to Ca$^{2+}$ and Gd$^{3+}$ excited at 395 nm (a and c) and fitted normalized intensity (b and d). CatchER experiences a large fluorescence increase when titrated with Ca$^{2+}$ and with Gd$^{3+}$. The normalized fluorescence was fitted with the 1:1 binding equation to give $K_d$ values of 315.4 ± 40.0 µM for Ca$^{2+}$ and 53.0 ± 4.0 µM for Gd$^{3+}$. 
Figure 4.9 (a-b) Flexible loop region comprising residues 155-159. CatchER-Gd$^{3+}$ is shown in a cyan cartoon representation with the mutated residues of the designed Ca$^{2+}$ binding site in red sticks and chromophore colored by CPK. The alternative conformations of residues 155-159 are indicated in cyan and orange by CPK. In the circle, $F_o-F_c$ omit map (in green) for the two conformations of residues 155-159 contoured at 2.0$\sigma$. In EGFP (PDB code: 4EUL) and GFP structure (PDB code: 2WUR), only the conformation in orange was observed.
Figure 4.10 Structure validation by Ramachandran plots for the three CatchER structures, as performed by PROCHECK CCP4i suite program. (Potterton et al., 2003)
REFERENCES


APPENDICES

Appendix A: List of Crystal Structures

<table>
<thead>
<tr>
<th>Protein</th>
<th>ligand</th>
<th>PDB accession code</th>
<th>Resolution (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PR&lt;sub&gt;L76V&lt;/sub&gt;</td>
<td>darunavir</td>
<td>3PWM</td>
<td>1.46</td>
</tr>
<tr>
<td>PR&lt;sub&gt;L76V&lt;/sub&gt;</td>
<td>saquinavir</td>
<td>3PWR</td>
<td>1.45</td>
</tr>
<tr>
<td>PR&lt;sub&gt;P51&lt;/sub&gt;</td>
<td>darunavir</td>
<td>4NPT</td>
<td>1.66</td>
</tr>
<tr>
<td>PR&lt;sub&gt;P51&lt;/sub&gt;</td>
<td></td>
<td>4NPU</td>
<td>1.50</td>
</tr>
<tr>
<td>CatchER</td>
<td>Gd&lt;sup&gt;3+&lt;/sup&gt;</td>
<td>4L12</td>
<td>1.78</td>
</tr>
<tr>
<td>CatchER</td>
<td></td>
<td>4L13</td>
<td>1.66</td>
</tr>
<tr>
<td>CatchER</td>
<td>Ca&lt;sup&gt;2+&lt;/sup&gt;</td>
<td>4L11</td>
<td>1.20</td>
</tr>
</tbody>
</table>

Appendix B: List of Publications


