

**X-ray Crystallographic Analysis of HIV-1
Protease: substrate/inhibitor complexes**

By

Vishal Prashar

LIFE01200604006

BHABHA ATOMIC REASERCH CENTRE

*A thesis submitted to the
Board of Studies in Life Sciences Discipline*

In partial fulfillment of requirements

For the Degree of

DOCTOR OF PHILOSOPHY

of

HOMI BHABHA NATIONAL INSTITUTE



February, 2012

Homi Bhabha National Institute

Recommendations of the Viva Voce Board

As members of the Viva Voce Board, we certify that we have read the dissertation prepared by **Vishal Prashar** entitled “**X-ray Crystallographic Analysis of HIV-1 Protease: substrate/inhibitor complexes** ” and recommend that it may be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy.

Date: 05/09/2012

Chairman- Dr. S.F. D'Souza

Date: 05/09/2012

External Examiner- Dr. Rajan Sankaranarayanan (CCMB, Hyderabad)

Date: 05/09/2012

Guide / Convener- Dr. M.V. Hosur

Date: 05/09/2012

Member 1- Dr. M. Seshadri

Date: 05/09/2012

Member 2- Dr. Vinay Kumar

Final approval and acceptance of this dissertation is contingent upon the candidate's submission of the final copies of the dissertation to HBNI.

I hereby certify that I have read this dissertation prepared under my direction and recommend that it may be accepted as fulfilling the dissertation requirement.

Date: 05/09/2012

Guide – Dr. M.V. Hosur

Place: Mumbai

STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at Homi Bhabha National Institute (HBNI) and is deposited in the library to be made available to borrowers under rules of the HBNI.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgement of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the Competent Authority of HBNI when in his or her judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

(Vishal Prashar)

DECLARATION

I, hereby declare that the investigations presented in the thesis have been carried out by me. The work is original and has not been submitted earlier as a whole or in part for a degree/diploma at this or any other Institution/University.

(Vishal Prashar)

ACKNOWLEDGEMENTS

It gives me immense pleasure to express my deep sense of gratitude in the form of words for the guidance and help received in the conduct of this piece of research from:

Dr. M.V. Hosur, for deft guidance, valuable suggestions, constructive criticism and academic support. He introduced me to the theoretical and practical aspects of crystallography and guided me at every stage of my research work and gave constant encouragement.

Dr. K.K. kannan, whose thought provoking queries and useful comments helped a lot.

Dr. J.K. Dattagupta, Dr. S.F. D'souza, Dr. Vinay Kumar and Dr. M. Seshadri as members of my HBNI doctoral committee for providing valuable suggestions for the completion of this work.

Dr. Jean-Luc Ferrer, for help in data collection.

Dr. D.R. Rao and Dr. Smita Mahale who were magnanimous in providing saquinavir and peptide substrates, respectively.

Dr. Bindu Pillai, Dr. Mukesh Kumar, Amit Das and Subhash Bihani for cooperation extended whenever I approached them. I am thankful to S.R. Jadhav as well for all the help.

National facility for Macromolecular crystallography, BARC, for data collection and other biochemical facilities.

I owe a lot more than I can express in words to my parents and other family members for all their sacrifices and affectionate blessings during the course of this study. I would also like to thank my wife "Mona" for her constant encouragement for doing the good work.

Though, all cannot be mentioned, but none is forgotten.

February, 2012

Vishal Prashar

SSPD, BARC

Contents

		Page No.
Synopsis		i-vii
List of Figures		viii-x
List of Tables		xi
Chapter 1	Introduction	1-24
Chapter 2	Protein crystallographic methods	25-54
Chapter 3	Catalytic water co-existing with a product peptide in the active site of HIV-1 protease revealed by X-ray structure analysis	55-77
Chapter 4	Insights into the mechanism of drug resistance: X-ray structure analysis of G48V/C95F tethered HIV-1 protease dimer/saquinavir complex	78-95
Chapter 5	X-ray structure analysis of L90M/C95F and C95F mutants of tethered HIV-1 protease dimer in complex with saquinavir	96-119
Chapter 6	X-ray structure analysis of D30N tethered HIV-1 protease dimer/saquinavir complex	120-134
Chapter 7	Summary and future directions	135-142
References		143-174

Synopsis

X-ray Crystallographic Analysis of HIV-1 protease:

substrate/inhibitor complexes

As the pandemic spread of Acquired Immunodeficiency Syndrome (AIDS) continues into its third decade, a complete cure from it still eludes the scientific community. In the absence of a cure, suppressing viral replication and maintaining it at low to undetectable levels have become critical goals in the field of HIV research. To this end, highly active antiretroviral therapy (HAART) which includes simultaneous use of inhibitors of viral enzymes, HIV protease and HIV- reverse transcriptase, has become a successful strategy in providing long, quality life for infected individuals. Since the catalytic activity of mature protease and ordered processing of the viral polyproteins have been shown to be critical for the liberation of infective progeny virus, protease inhibitors are essential components of HAART therapies. All the Food and Drug Administration (FDA) approved drugs that target HIV protease resulted from the structure based drug design efforts of the academia and the pharmaceutical industry. However, the clinical emergence of drug resistant variants of HIV in response to HIV protease inhibitors has often limited the prolonged effectiveness of most promising of these drugs. In these variants, HIV protease is found to be mutated in both the active site and non-active site regions of the enzyme. This necessitates continuous improvement of existing drugs and design of new HIV protease inhibitors based on structural understanding of: (1) chemical interactions between HIV protease and its substrates/products (2) mechanism of action of HIV protease, and (3) mechanism of drug resistance. X-ray crystallography is a powerful tool to get such information at atomic level. Keeping these objectives in mind, the present work aims

at crystallographic studies on complexes between substrate-peptides/ inhibitors and native/drug resistant mutants of HIV protease.

Chapter 1 gives general introduction about Acquired Immunodeficiency Syndrome (AIDS), Human Immunodeficiency Virus (HIV), HIV-protease and drug resistant mutations. HIV protease is an aspartyl protease formed by non-covalent association of two polypeptide chains, each 99 residues long. In the 3-D structure, the two subunits are related by an almost perfect two-fold rotation axis. There is a single active site in this homodimer and this active site is located at the interface of the two subunits. The carboxyl terminus of subunit 1 is about 5 Å away from the amino terminus of subunit 2, and functional single chain HIV protease have been produced by covalently linking these two ends through oligopeptide linkers. A water molecule is implicated as a nucleophile in the various proposed mechanisms of reaction. The enzyme is functional over a wide pH range of 3-7.

Chapter 2 gives the general methodology followed in the present work. HIV-1 protease tethered dimer used in the present study contains a five residue linker, GGSSG, covalently linking the two monomers. The cloned insert contains 57 extra codons in the beginning, which is a part of N-terminal polyprotein of polymerase gene. Therefore the inserted gene product is a 29 kDa precursor protein, containing natural cleavage site for HIV-1 protease, which after self cleavage results in a mature protein of 22 kDa. Mutant clones were prepared by me in our laboratory. Mid transition temperatures (T_m) of native and mutant proteins were determined using Circular Dichroism (CD) spectrometry. Kinetic constants of native and mutant enzymes were determined by spectrophotometric enzyme assay. Single crystals of tethered HIV-1 protease and mutant-drug complexes were obtained by hanging drop vapour diffusion method. The crystals were equilibrated in a cryoprotectant solution

before flash freezing, for exposure to X-rays on beamlines at European synchrotron radiation facility (ESRF) and Swiss Light Source (SLS). Diffraction data were collected as a number of oscillation frames. The diffraction data were processed and scaled by using the computer programs XDS, MOSFLM and SCALA. Crystal structures were solved by difference Fourier method using native coordinates. Graphics Programmes O and COOT were used for electron density interpretation and model building. The structures were refined using computer programmes CNS, REFMAC and PHENIX. The quality of models were checked using programme PROCHECK. A brief description of the method of X-ray crystallography for structure determination is also given in this chapter.

Understanding the catalytic mechanism of the enzyme is of crucial importance as transition state structure directs inhibitor design. Most mechanistic proposals invoke nucleophilic attack on the scissile peptide bond by a water molecule. But such a water molecule coexisting with any ligand in the active site has not been found so far in the crystal structures. **In chapter 3**, we report the first observation of the coexistence, in the active site, of a water molecule WAT1, along with the carboxyl terminal product (Q product) peptide. Crystals of the complex were prepared by soaking in an undecapeptide substrate. The structure was determined to a resolution of 1.69 Å with $R_{\text{work}}/R_{\text{free}}$ of 21.9/25.6%. The product peptide has been generated in situ through cleavage of the full-length substrate of sequence which matches the CA-p2 junction of the viral polyprotein. The N-terminal product (P product) has diffused out and is replaced by a set of water molecules while the Q product is still held in the active site through hydrogen bonds. The position of WAT1, which hydrogen bonds to both the catalytic aspartates, is different from when there is no substrate bound in the active site. We propose WAT1 to be the position from where catalytic water attacks the

scissile peptide bond. The structure is suggestive of the repositioning, during substrate binding, of the catalytic water for activation and subsequent nucleophilic attack. The structure could be a snap shot of the enzyme active site primed for the next round of catalysis. This structure further suggests that to achieve the goal of designing inhibitors mimicking the transition-state, the hydrogen-bonding pattern between WAT1 and the enzyme should be replicated.

Saquinavir (SQV) is one of the important protease inhibitor which is currently in the HIV-treatment-regimen. Mutations at select positions both in the active site and non-active site regions of HIV protease confer resistance to the virus against this drug. The mutation G48V in HIV-1 protease is a major active site mutation against the drug saquinavir. Recently, G48V mutation is found to co-exist with the mutation C95F in AIDS patients treated with saquinavir. **In chapter 4**, we report the three-dimensional crystal structure of G48V/C95F tethered HIV-1 protease/saquinavir complex. Crystals of the complex were grown by co-crystallizing the protein and the drug. Crystal structure was solved and refined to a resolution of 2.5Å and $R_{\text{work}}/R_{\text{free}}$ of 21.6/25.3%. While the overall protein structure is similar to that of wild type-SQV complex, the side chain of flap residue PHE-53 has assumed a different rotamer conformation and the quinoline ring in the P3 position of SQV has moved away from the flap due to the mutation G48V. The structure indicates following as the possible causes of drug resistance: (1) loss of direct van der Waals interactions between saquinavir and enzyme residues, (2) loss of water-mediated hydrogen bonds between the carbonyl oxygen atoms in saquinavir and amide nitrogen atoms of flap residues 50 and 1050, (3) changes in inter-monomer interactions, which could affect the energetics of domain movements associated with inhibitor-binding, and (4) significant reduction in the stability of the mutant dimer. The T_m was indeed measured to be reduced by 6 °C.

L90M is a major non-active site mutation against the drug SQV which occurs in combination with the mutation C95F. **In chapter 5**, we describe the structures of L90M/C95F and C95F mutants of tethered HIV-1 protease complexed with the drug SQV. The crystal structures of the L90M/C95F_SQV and C95F_SQV complexes were refined to $R_{\text{work}}/R_{\text{free}}$ values of 20.4/24.3% and 20.6/24.5% at resolutions of 2.04 and 1.75 Å, respectively. Comparisons of SQV bound L90M/C95F mutant structure with that of native enzyme-SQV complex have shown reshaping of the active site cavity in the mutant structures. This is the result of altered packing in the core of the enzyme. SQV and in fact most of the other anti HIV-1 protease drugs are designed as competitive inhibitors so that they bind to the enzyme more tightly than its natural substrate. Hence, any change in the shape of active site cavity may affect the binding of these drugs more than that of natural substrates. Small changes in the conformation of SQV were also seen in response to changes of enzyme conformation in the active site. Conversely, flexible substrates could adapt to these changes comparatively easily. Presence of these mutations may limit the structural flexibility of the active site loop which is required to accommodate the incoming inhibitor.

Introduction of bulky phenylalanine residue at position 95 in lieu of cysteine leads to repacking in the inter subunit four stranded beta sheet, with the result that the inter-subunit hydrogen bond between PRO-1001N and PHE-99 CO is lost in the mutant. There were changes in inter-monomer interactions, which could affect the energetics of domain movements associated with inhibitor-binding. All these mutants have lower dimer stability as compared to wild type enzyme as shown by reduction in the mid transition temperatures (T_m) by about 6-8°C that were determined by using circular dichroism spectroscopy. Major contribution to this reduction seems to come from C95F mutation.

Nelfinavir (NFV) was developed by truncating the N-terminal moiety in SQV and replacing the P2 asparagine with 3-hydroxy-2-methylbenzamide fragment. These changes significantly reduced molecular weight and improved the bioavailability, though NFV is less potent than SQV. D30N is a major mutation against the drug nelfinavir but does not provide resistance against the drug saquinavir. The structure of D30N-saquinavir complex was solved to 1.8 Å and refined to $R_{\text{work}}/R_{\text{free}}$ of 20.2/23.4% and this result forms **chapter 6** of the thesis. Structural analysis showed that P2' Asn of saquinavir forms an additional hydrogen bond to the mutated side chain of residue 30 apart from a main chain hydrogen bond and this could be the reason why D30N is not a drug resistant mutation against saquinavir.

In summary, structure of HIV-1 protease-product peptide complex gives an insight into the molecular mechanism of HIV-1 protease. The structure suggests repositioning of catalytic water, WAT1, during substrate binding for activation and subsequent nucleophilic attack. Structures of mutants in complex with saquinavir give insight into the mechanisms of drug resistance. Residues that undergo conformational changes leading to loss of van der Waals interactions, reshaping of active site cavity and reduction in dimer stability through repacking at the inter domain regions have been identified. The reason why D30N mutation is not effective against saquinavir has been found out. For future efforts of designing inhibitors mimicking the transition-state, the hydrogen-bonding pattern between WAT1 and the enzyme should be replicated. Main chain interactions between inhibitors and the active site residues of the enzyme should be maximized, and inhibitors should contain flexible elements that permit accommodation to the change in shape of the active site. Introduction of moieties which can mimic flap water should be considered.

Publications from Thesis work:

Vishal Prashar, Amit Das, Subhash Bihani, J.-L.Ferrer and M.V. Hosur (2009)
Catalytic Water Co-Existing with a Product Peptide in the Active Site of HIV-1
Protease revealed by X-Ray Structure Analysis. PLoS ONE 4: e7860.
doi:10.1371/journal.pone.0007860.

Vishal Prashar, Subhash C. Bihani, Amit Das, D.R. Rao and M.V. Hosur (2010)
Insights into the mechanism of drug resistance: X-ray structure analysis of
G48V/C95F tethered HIV-1 protease dimer/saquinavir complex. Biochem.Biophys.
Res. Commun. 396: 1018–1023.

Vishal Prashar, Subhash Bihani, Amit Das, J.-L. Ferrer and M.V. Hosur (2011)
Saquinavir resistance mechanism of non-active site mutants of HIV-1 protease. Acta
Cryst A 67: C292-C293.

Vishal Prashar, Subhash Bihani, Amit Das, J.-L.Ferrer and M.V.Hosur; X-ray
structure analysis of L90M/C95F and C95F mutants of tethered HIV-1 protease dimer
in complex with saquinavir. (To be published)

Vishal Prashar, Subhash Bihani, Amit Das, J.-L.Ferrer and M.V. Hosur; X-ray
structure analysis of D30N tethered HIV-1 protease dimer/saquinavir complex. (To be
published)

List of Figures

Figure	Page No.
Figure 1.1: Diagrammatic sketch of Human Immunodeficiency Virus (HIV)	4
Figure 1.2: HIV genome	4
Figure 1.3: HIV life cycle	7
Figure 1.4: Chain tracing of the dimer of HIV-1 protease	9
Figure 1.5: Fireman's grip	9
Figure 1.6: Schechter and Berger nomenclature of amino acid residues of peptide substrates and corresponding sub-sites of the enzyme	12
Figure 1.7: Cartoon of the inter domain beta sheet of Eukaryotic aspartic protease (Rhizopepsin) and inter subunit beta sheet of HIV-1 Protease	16
Figure 1.8: FDA approved HIV protease inhibitors	21
Figure 2.1: Velocity vs. Substrate concentration curve for native tethered dimer of HIV-1 protease	54
Figure 3.1: HIV-1 protease precursor coded by tethered dimer insert	58
Figure 3.2: SDS PAGE at various stages of protein purification	59
Figure 3.3: Fo-Fc difference density contoured at 2.0σ level in the active site of HIV-1 protease/product complex	64
Figure 3.4: Fit of carboxyl terminal (Q) product peptide and active site water molecules into $2F_o-F_c$ electron density	65
Figure 3.5: Hydrogen bonding interactions at the active site	67
Figure 3.6: Relative positions of WAT1 and the modelled substrate in the active site	68
Figure 3.7: Structural comparison of present complex with tetrahedral intermediate complex and product peptide complex	70
Figure 3.8: Position of the nucleophilic water molecule	72

Figure 3.9: Proposed sequence of steps (a-f) in the cleavage of peptide bond by HIV-1 protease	73
Figure 4.1: Determination of kinetic constants	83
Figure 4.2: Structural effects around G48V mutation site	88
Figure 4.3: Structural comparison of present complex (green), wild-type complex (yellow) and G48V/L90M complex (purple) with the same drug SQV	89
Figure 4.4: Structural effects around C95F mutation site	91
Figure 4.5: Hydrogen bonds from flap-water molecule	92
Figure 5.1: Determination of kinetic constants	98
Figure 5.2: F_O-F_C omit maps for Methionine 90 and Phenylalanine 95	102
Figure 5.3: Structural comparison of L90M/C95F_SQV (green carbons) complex with C95F_SQV complex (yellow carbons)	106
Figure 5.4: Structural effects on the active site floor	107
Figure 5.5: Rotation of Ile1084 side chain around χ_1 torsion angle	108
Figure 5.6: Shift of residues in the helix due to L90M mutation	109
Figure 5.7: Shift in the 80's loop	110
Figure 5.8: Additional interactions of Cϵ methyl group of Methionine90	111
Figure 5.9: Loss of interactions of SQV in the active site	113
Figure 5.10: Local effects of C95F mutation in the second subunit	115
Figure 5.11: Loss of hydrogen bond in the four stranded inter subunit beta sheet	116
Figure 5.12: Repacking at the interface of core and terminal domains	118
Figure 6.1: Chemical structures of (a) Saquinavir (SQV) (b) Nelfinavir (NFV)	122
Figure 6.2: Fit of SQV in the F_O-F_C omit map	126

Figure 6.3: Alternate conformations of SQV and Ile84	127
Figure 6.4 Structural comparison of the present complex with the D30N-NFV complex in the active site	130
Figure 6.5: Comparison of hydrogen bonding interactions of P2 residues in SQV and NFV in D30N mutants	131
Figure 6.6: Structural effects around D30N mutation site	133

List of Tables

Table	Page No.
Table 3.1: Data collection and refinement statistics	62
Table 3.2: Comparison of conformation of catalytic aspartates in the structures at pH 2.0 and pH7.0	75
Table 3.3: Comparison of interactions of catalytic aspartates in the structures at pH 2.0 and pH7.0	75
Table 4.1: Data collection and refinement statistics	85
Table 4.2: Kinetics data	86
Table 4.3: Comparison of main chain and side chain torsion angles of Phe99 (Phe1099) in the wild-type complex (PDB Id 1HXB) and present structures	90
Table 4.4: T_m values for native and G48V/C95F mutant proteins in the presence and absence of saquinavir (SQV)	94
Table 5.1: Kinetics data	100
Table 5.2: Data collection and refinement statistics	103
Table 5.3: Comparison of conformation of Asp25/1025 in L90M/C95F_SQV and C95F_SQV complex structures	105
Table 5.4: Comparison of active site volumes and T_m's	118
Table 6.1: Data collection and refinement statistics	125
Table 6.2: Comparison of conformation of Arg8 in D30N-NFV and D30N-SQV Complexes	129

Chapter1

Introduction

Acquired Immunodeficiency Syndrome (AIDS) is a pathogenic state in which ability of immune system to control infections or malignant disorders is destroyed. AIDS was first reported in a Centre for disease control (CDC), USA, newsletter in 1981. Human Immunodeficiency Virus (HIV), the causative agent of AIDS [1, 2], was discovered by several independent groups who called it by different names, such as lymphadenopathy virus (LAV), human T-cell lymphotropic virus type III (HTLV III), and AIDS related virus (ARV). Then, in 1986, an international committee renamed the virus as Human Immunodeficiency Virus (HIV). Following the discovery of an antigenic variant in 1986, the original virus was designated HIV-1 and variant was designated HIV-2. However, HIV-2 is less widespread than HIV-1. Both HIV-1 and HIV-2 are genetically related to simian immunodeficiency viruses (SIVs), which are found in African primates. AIDS has assumed pandemic proportions since its first report by CDC, USA, in the early 1980's. According to a recent report by Joint United Nations Programme on HIV/AIDS (UNAIDS); about 33 million people worldwide are infected with HIV [3].

1.1 HIV groups and subtypes:

HIV has a wide range of genetic variation and classified into distinct types, groups and clades. The two major types of HIV, HIV-1 and HIV-2, are distinguished by their genomic organization and phylogenetic relationship. The origin of HIV in humans is thought to be due to multiple zoonotic infections. HIV-1 strains are related to SIV which naturally infects chimpanzee (SIV_{cpz}), while HIV-2 is closer to SIV infecting

sooty mangabeys (SIV_{sm}). HIV-1 is responsible for the AIDS pandemic and can be divided into four groups (M, N, O and P), each one derived from a distinct introduction of SIV_{cpz} in the human population [4]. HIV-1 group O (Outlier) is the most divergent group, and it has been suggested that it could have originated from SIV infecting wild gorillas (SIV_{gor}) [4]. SIV_{gor} is related to SIV_{cpz} and thus it is possible that gorillas are an intermediate reservoir of HIV-1 group O [5]. The HIV-1 group O is restricted to the West and Central Africa, and around 15,000 people are estimated to be infected with this variant in Cameroon [6, 7]. Group N (New) was identified only in 1998 [8], and originated from a recombination event between the ancestor of group M and SIV_{cpz} [9]. HIV-1 group N is very rare epidemiologically and less than 50 cases have been identified to date, all in Cameroon [10-12]. Recently, a new HIV-1 group (P) was described in a Cameroonian woman [13]. The origin of this group was correlated with SIV_{gor} , without evidence of recombination with other HIV-1 groups. Currently, this is the only identified case of this variant, but other cases are expected to be hidden in Cameroon. HIV-1 group M (Major) alone is responsible for more than 95% of the AIDS pandemic and virtually all studies in HIV research have been conducted with representatives of this group.

Based on sequencing and alignment of viral *env* and *gag* genes from different strains of HIV-1, group M was divided into subtypes or clades A-K. Nucleotide divergence in the *pol* gene (the most conserved in retroviruses) between different HIV-1 subtypes is around 9–11%. However, studies of the HIV-1 full-length genome sequences (14–17) revealed that HIV-1 can generate complex pattern of inter-subtype recombinant strains. Therefore, according to the current classification, HIV-1 group M is divided into nine different “pure” subtypes or non-recombinant forms (A-D, F-H, J and K) [18]. Inter-subtypic mosaic isolates are classified into a circulating recombinant form

(CRF), when it is found in a population (in at least three people with no epidemiologic link and with the same inter-subtypic breakpoints) and unique recombinant form (URF), when it is found only in a single patient. Currently, over 40 CRFs have been characterized. “Pure” subtypes E and I have been reclassified to CRF01_AE and CRF06_cpx, respectively. Subtypes A and F are further divided into sub-subtypes A1 through A5 and F1 and F2 [19-21]. Genetic differences between sub-subtypes are around 7% in *pol*.

1.2 HIV structure:

HIV is a member of the *Lentivirus* genus of the retrovirus family *Retroviridae*, whose genome is encoded in RNA. The mature HIV virion is almost spherical and is about 100 nm in diameter [22]. Outer viral envelope is a lipid bilayer that is derived from the infected host cell by a process called budding. The viral envelope is studded with the proteins from the host cell, as well as about 72 copies of heterodimer complex composed trimers of external surface glycoprotein 120 (gp120), and transmembrane glycoprotein 41 (gp41) that anchor the structure in the viral envelope. Matrix (MA) protein p17 lines the inner surface of the viral membrane. Virus core is composed of bullet shaped capsid made of 2000 copies of a protein called p24 (CA). Enclosed within the capsid are two copies of positive single-stranded RNA genome. The single-stranded RNA is tightly bound to nucleocapsid proteins, p7, and enzymes needed for the development of the virion such as reverse transcriptase, protease and integrase.

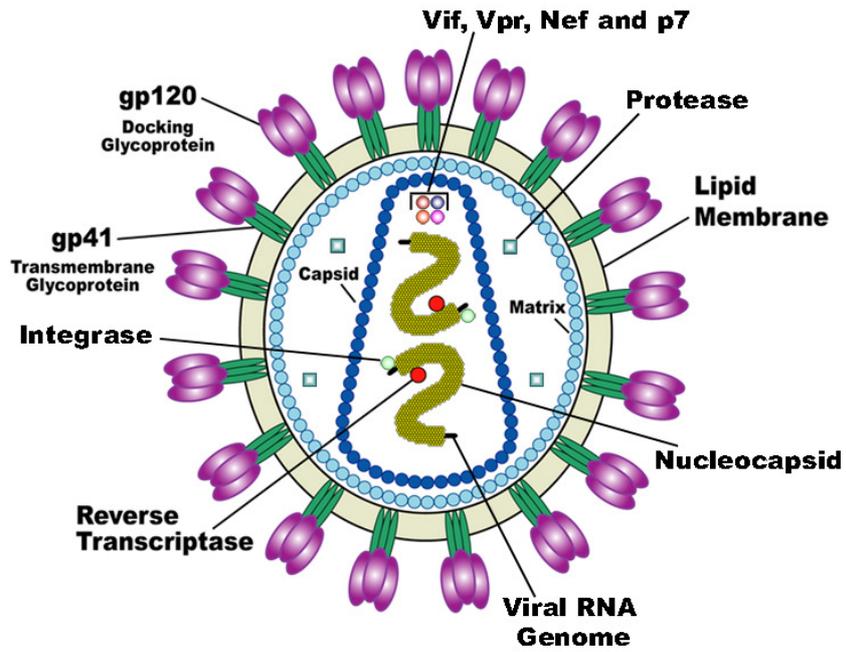


Figure 1.1: Diagrammatic sketch of Human Immunodeficiency Virus (HIV) (Credit: PD-US Gov-HHS-NIH/Wikimedia Commons/Public Domain)

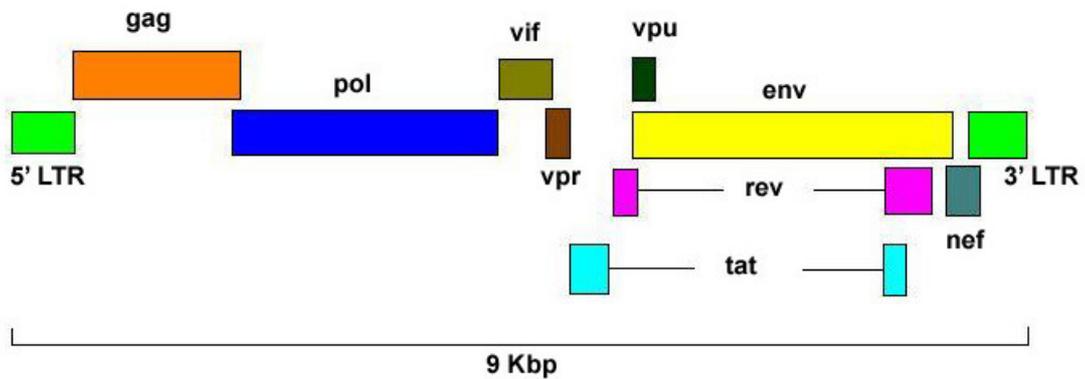


Figure 1.2: HIV genome

1.3 HIV genome:

HIV genome is about 9 Kbp in size and contains a total of nine genes. Organisation of HIV genome is shown in the figure 1.2. Like all other retroviruses, HIV provirus is flanked on either side by repetitive sequences known as long terminal repeats (LTR's). 5' LTR contains promoter and enhancer elements essential for transcription while 3' LTR is required for polyadenylation of transcripts. HIV genome contains three major genes *gag*, *env* and *pol* which encode for viral core proteins, surface envelope proteins and viral enzymes respectively. All these genes encode for polyprotein precursors which are then cleaved to give final gene products. Polyprotein encoded by *pol* is cleaved to give three viral enzymes: reverse transcriptase (RT), protease (PR) and integrase (IN). The *gag* gene codes for 55kDa precursor (p55) which is cleaved by *pol* encoded protease to p17, p24, p7 and p9. The *env* gene codes for glycosylated precursor (p160) which is cleaved by host protease to give gp120 and gp41. Of the remaining six genes, three (*tat*, *rev* and *nef*) encode regulatory proteins that play a major role in controlling expression, two (*vif*, *vpu*) encode proteins required for viral maturation and one (*vpr*) encodes a weak transcriptional activator.

1.4 HIV life cycle:

Major cellular receptor for HIV is CD4. Because the T-helper lymphocyte cells express the highest level of CD4, HIV is said to be lymphotropic. Other cells that bind HIV include macrophages, dendritic cells, langerhans cells, hematopoietic stem cells, certain rectal lining cells, and microglial cells. The infection of the viral and cellular membranes is a process that is mediated by the viral envelope glycoproteins (gp120, gp41) and receptors (CD4 and co-receptors, such as CCR5 or CXCR4) on the target cell. Membrane fusion is followed by a poorly understood uncoating event of the

capsid that allows the release of the viral content into the host-cell cytosol. As the virus enters a cell, its single stranded RNA is reverse-transcribed to yield a double-stranded DNA molecule by a virally encoded enzyme, the reverse transcriptase (RT). The viral DNA enters the cell nucleus, where it is permanently integrated into the genetic material of the host cell, forming a provirus, by the catalytic activity of a second virally encoded enzyme, the integrase (IN). Activation of host cell results in the transcription of the viral DNA into messenger RNA, which is then translated into viral proteins (e.g. tat, rev, and nef regulatory proteins and gag and gag-pol polyproteins). HIV protease, the third virally encoded enzyme, is required in this step to cleave a viral polyprotein precursor (gag and gag-pol) into mature functional enzymes and structural proteins. As the viral proteins begin to assemble within the host cell, the host cell membrane is modified by insertion of gp41 and associated gp120. The viral RNA and core proteins then assemble beneath the modified host plasma membrane as its envelope in a process called budding. Viral gp120 expressed on the host cell plasma membrane binds to membrane CD4 as budding occurs. In cells expressing high levels of CD4, this autofusion disrupts the membrane integrity and leads to cell lysis. The gag-pol precursor results from a ribosomal frame shift and read-through during translation of the gag gene. The gag and gag-pol gene products in assembled immature virions are found in the ratio of 20:1, which represents the frequency of ribosomal frameshifting of about 5%.

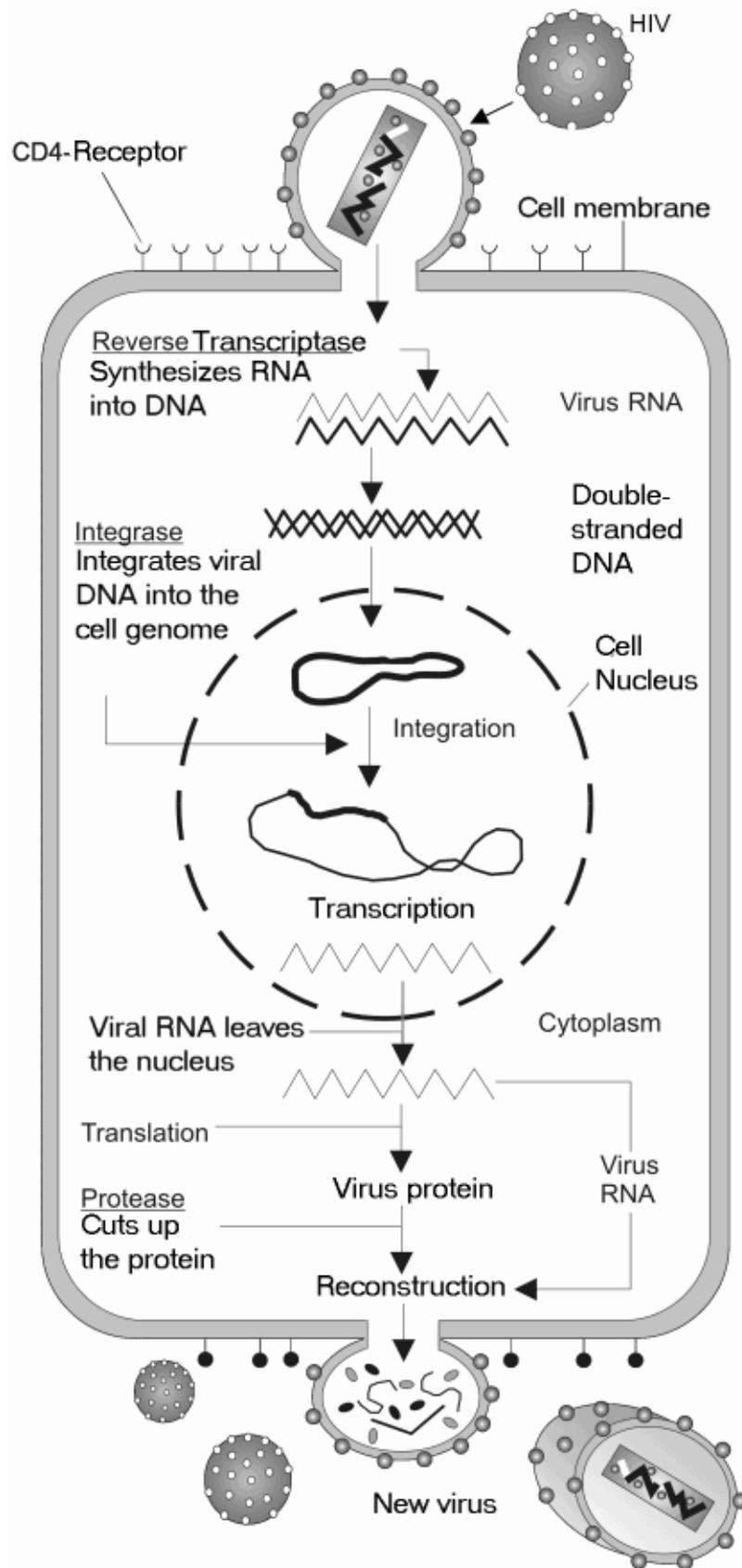


Figure 1.3: HIV life cycle (© Optigan 13/ Wikimedia Commons /CC-BY-SA-3.0)

1.5 HIV protease structure:

Navia and colleagues from Merck Laboratories were the first group to obtain a crystal structure of HIV-1 protease in 1989 [23] and since then, protease structure has been extensively characterized in terms of function, substrate specificity and inhibitor binding. This enzyme functions as a homodimer (two identical polypeptide chains) with only one active site. Each monomer consists of 99 amino acid residues. Dimer interface is composed of a four stranded beta sheet with beta strands from both the amino and carboxyl termini (Figure 1.4). The amino terminal beta strand 'a' (residues 1-4) forms the outer part of dimer interface beta sheet. Strand 'a' continues through a loop to strand 'b' (residues 9-15). Strand 'b' is followed by beta strand 'c' ending in active site triad (residues 25-27). The active site loop is followed by beta strand 'd' (residues 30-35). This is followed by a broad loop (residues 36-42). The topology of second half of the monomer is related to the first half by a pseudo-dyad and corresponding substructures are indicated by primed labels. Residues 43-49 form beta strand a' which forms a part of the flap. Other part of the flap is formed by residues 52-58 which are part of a long beta strand b' (residues 52-66). Residues 68-78 form part of the beta chain c' which continues through a loop (residues 79-82) to beta chain d'(residues 83-85). This is followed by a short helix h' (residues 86-94). The hydrogen bonding pattern of this helix is intermediate between an alpha helix and 3_{10} helix. Helix h' is followed by carboxyl terminus beta strand q (residues 95-99) which form a part of the inner core of dimer interface four stranded beta sheet.

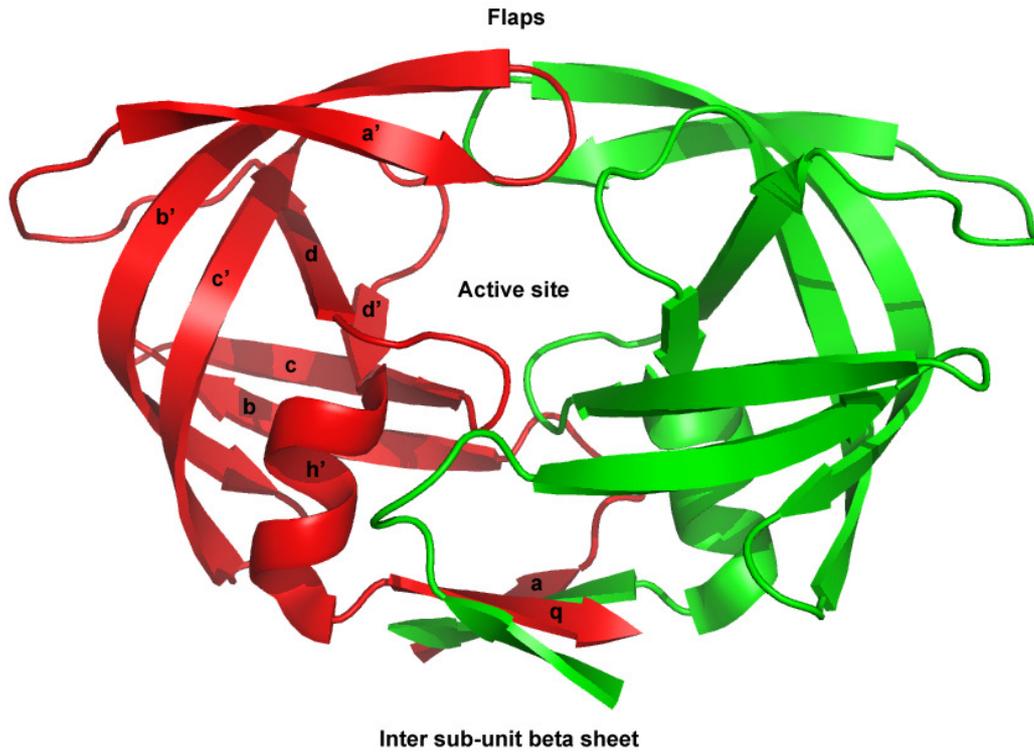


Figure 1.4: Chain tracing of the dimer of HIV-1 protease

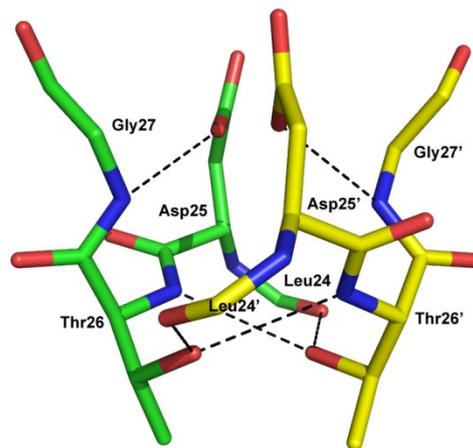


Figure 1.5: Fireman's grip

Four strands of the molecular core are organised as psi shaped sheet which is characteristic of aspartic proteases. One of the letters of psi is composed of strands c (residues 23-25), d and d' while other is composed of strands c' (residues 76-78), d' and d. The active site triad (Asp25-Thr26-Gly27) is located in a wide loop which is stabilized by a network of hydrogen bonds including the so called 'Fireman's grip' between the two monomers (Figure 1.5). In Fireman's grip, each Thr of the active site triad accepts a hydrogen bond from the Thr amide group in the other loop and donates a hydrogen bond to the carbonyl oxygen atom of Leu24 from the other strand. The structure of each individual loop is reinforced by a hydrogen bond between O δ 1 of the Asp25 and the amide group of the Gly27, the last residue in the triad. The amino and carboxyl termini of the two monomers are interdigitated and linked by a network of hydrogen bonds to form the dimer interface. Also, the distance between the carboxyl terminal of one monomer and amino terminal of the other is around 5 Å. This observation led several groups to express HIV-1 protease as a tethered dimer which is a single polypeptide chain in which two copies of monomeric coding sequences are genetically linked by a short linker. Several research groups have produced the protease in a single chain form with various linker chain lengths and sequences [24]. Biochemical studies have shown that there are no major differences in the activities of tethered HIV-1 protease as compared to the two chain HIV-1 protease [25]. Using tethered dimer, it has been shown that Asp25 from both the subunits are required for activity.

1.6 HIV protease active site:

The enzyme active site is situated at the dimer interface. While the two catalytic aspartates are located on the floor of the active site cavity, two double stranded beta-

ribbon structures, one coming from each subunit, form the roof of the cavity. These beta-ribbons, described as flaps contain many glycine residues and are therefore conformationally very flexible segments of the protein. Comparison of unliganded HIV protease crystal structures with its complexes with inhibitors shows a 7 -15 Å movement of the flap tips (around residues 50/50') [26, 27]. The flaps are held in a closed position by hydrogen bonding from the flap residues Ile50/50' to a conserved water molecule, which in turn is hydrogen bonded to two carbonyl groups in the inhibitor/substrate peptides. The substrate binds in its extended conformation, in which its interactions with the different amino acid side chains determine the specificity of the enzyme [28]. According to the Schechter and Berger nomenclature, amino acids proceeding toward the amino terminus of scissile peptide bond of substrate are designated P1, P2, P3...Pn and those going toward the carboxyl terminus are referred to as the P1', P2', P3'...Pn' (Figure 1.6) [29]. Corresponding sub-sites of the enzyme interacting non-covalently with the respective side chains of the peptide (substrate or inhibitor) are termed S1, S2, S3...Sn and S1', S2', S3'...Sn', respectively. Using the above standard nomenclature, the S1 and S1' (similarly S2 and S2' etc.) sub-sites are located at equivalent positions. The two S1 sub-sites are highly hydrophobic with the exception of the active site aspartates. The residues which line up the S1/S1' sub-sites include Arg8, Leu23, Asp25, Gly27, Gly48, Gly49, Ile50, Thr80, Pro81 and Val82. The S2/S2' sub-sites are mostly hydrophobic (Ala28/28', Leu23/23', Ile47/47', Gly49/49', Ile50/50', Leu76/76' and Ile84/84') except Asp29, Asp29', Asp30 and Asp30'. These sub-sites are smaller than the S1/S1' or the S3/S3'

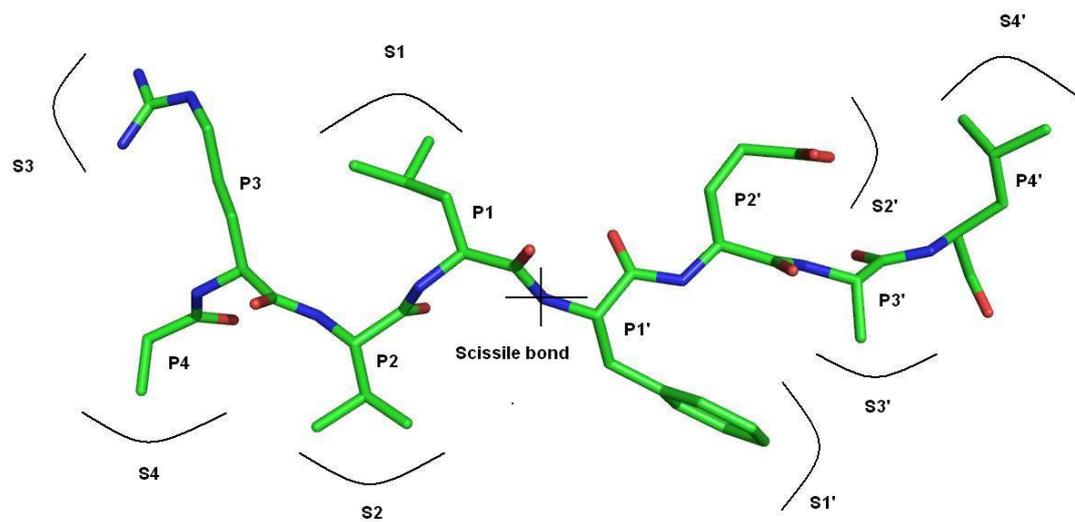


Figure 1.6: Schechter and Berger nomenclature of amino acid residues of peptide substrates and corresponding sub-sites of the enzyme.

binding sites and have been shown to be more specific, restricting the type and size of residues at P2/P2' in substrates or inhibitors relative to other binding pockets in the protease molecule [30, 31]. The S3 sub-sites adjacent to S1 sub-sites are also mostly hydrophobic and are known to have a rather broad specificity due to their ability to accept residues of different types and sizes [32]. Residues forming the S3/S3' sub-sites include Arg8/8', Leu23/23', Asp29/29', Gly48/48', Gly49/49', Pro81/81' and Val82/82' which form parts of the S1/S1' pockets. On the other hand, due to limited number of available structures of retroviral proteases in complex with ligands that extend beyond P3/P3', sub-sites S4/S4' and S5/S5' have not been described in comparable detail [33, 34]. It has also been shown that the HIV protease most efficiently cleaves peptide substrates seven or eight amino acids long (P4-P3') with the major processing subsites (S4-S3').

The HIV-1 protease processes the gag and gag-pol polyproteins proteolytically at specific sites. It is specific for cleavage of these sites *in vivo*, although the general sequence homologies among these are small [35, 36]. Based on the sites of processing, it is not possible to give a consensus substrate sequence and this is a general characteristic of the retroviral proteases. Nevertheless, based on similarities in amino acid sequences, primate lentiviral (HIV-1, HIV-2, SIV) cleavage sites were grouped into three classes [37], while analysis of a broader range of retroviral protease cleavage site sequences suggested two types of cleavage sites [38]. Later systematic specificity studies on HIV-1 protease also verified the existence of two types of cleavage sites, type 1 - having an aromatic residue and Pro, and type 2 having hydrophobic residues (excluding Pro) at the site of cleavage [39, 40], defined as P1 and P1' positions [29], respectively. These classical cleavage type sites also show different preferences for the P2 and P2' positions [39, 40].

The type 1 cleavage site was thought to be important in defining specificity. With the exception of pepsin, cellular proteases are not capable to efficiently process peptide bonds at the imino side of Pro residue. The unique nature of Pro at the P1' position of retroviral cleavage sites was recognized even before the discovery of the protease and implied that these sites should be processed by a virally-coded enzyme [41]. It should be noted, that many of the cleavage sites do not fit into these classifications (e.g., they contain polar residue at P1 or P1' or contain Pro after a nonaromatic residue), therefore, this classification might be an oversimplification.

1.7 HIV protease: an aspartic protease

HIV-protease was shown to belong to a class of proteases known as aspartic proteases based on various criteria in the past. Examination of the sequence homology of HIV-protease to other aspartic proteases showed that this enzyme has a sequence Asp-Thr-Gly, which is conserved among the aspartic protease enzymes [42]. Mutational analysis of the highly conserved Asp25 has shown that substitution of this residue with either Asn, Thr or Ala, leads to loss of proteolytic activity in this enzyme, as is the case with other enzymes [43-45]. In-vitro inhibition of HIV-protease by pepstatin, a natural product that selectively inhibits members of the aspartic proteases family, provided more support to the classification of this enzyme in the aspartic proteases family [43-46]. The structure of this enzyme also supported its classification in aspartic proteases family [23,47,48]. The dimeric structure, in which each monomer contributes one Asp-Thr-Gly triad to pseudo-symmetric active site, shows an active site that is indistinguishable from those of monomeric aspartyl proteases. However, HIV-protease is considerably smaller, consisting of two identical subunits that, taken together comprise a total of 198 amino acids. In contrast, cellular aspartic proteases

are generally two domain monomeric proteins of 325-350 amino acids. Rao *et al* had compared structural and evolutionary relationships between retroviral and eukaryotic aspartic proteases [49]. In both the retroviral and cellular aspartic proteases, an anti-parallel β sheet is present at the dimer/domain interface. In the cellular enzymes, this β sheet contains six strands arranged in the order A1R1Q1 \diamond Q2R2A2 (\diamond represents the pseudo-dyad and uppercase letters denote the strands in cellular aspartic proteases), where A1, Q1 and R1 are strands from the N-terminal domain and A2, Q2 and R2 are from the C-terminal domain. Retroviral enzymes lack the strands corresponding to Q1 and Q2 (penultimate C terminal strands in cellular aspartic proteases) and therefore the inter-subunit sheet is four stranded with strands arranged in the order a1q2 \diamond q1 a2, where a1 and q1 are strands from the first subunit and a2 and q2 are from the other (lower case letters denote strands in retroviral counterparts). Interestingly, the anti-parallel nature of the inter-domain sheet of cellular aspartic proteases has been preserved in the inter-subunit sheet of the retroviral proteases via strand exchange (q2/R1 and q1/R2) event. Also the orientation of this sheet relative to protein core differs by approximately 40° in the two groups of aspartic proteases. There are two flaps in the retroviral proteases as opposed to single flap in the cellular enzymes.

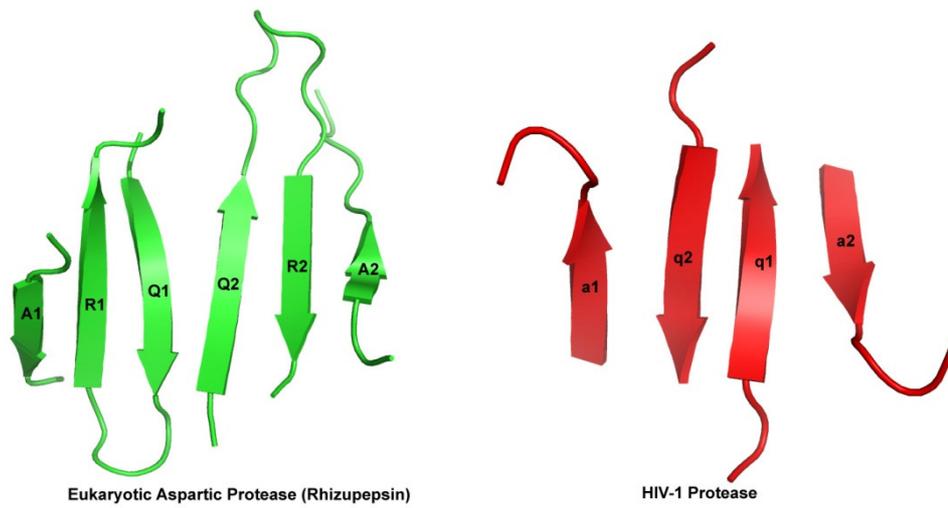


Figure 1.7: Cartoon of the inter domain beta sheet of Eukaryotic aspartic protease (Rhizopepsin) and inter subunit beta sheet of HIV-1 Protease.

1.8 HIV protease reaction mechanism:

Based on computational [50–53], biochemical [54-56] and structural results [57–61], three types of proposals have been made in the past for the catalytic mechanism [reviewed in 62–63]. In the first type, championed mostly by computational studies, the nucleophilic attack on the carbonyl carbon atom of the scissile peptide bond is directly by carboxyl oxygen atom of the catalytic aspartates which leads to the formation of covalent enzyme-substrate tetrahedral intermediate, followed by the expulsion of the amine component. This mechanism is largely discredited now because no evidence for covalent intermediates has been found so far. Second, is the general acid–general base mechanism, where one catalytic aspartate side chain carboxylate acts as a general base to remove a proton from the water molecule nucleophile, while another aspartic acid side chain carboxyl general acid donates a proton to the carbonyl oxygen atom of the scissile peptide bond. In the third ‘kinetic isomechanism’, a 10-membered cyclic structure is formed, involving the two aspartic acid side chain carboxyl groups, with a proton between them, and the water molecule nucleophile; this mechanism allows for energy-inexpensive proton shuffling within the cyclic structure along the reaction coordinate. The last two mechanisms also invoke a low-barrier hydrogen bond (LBHB); in the general acid–general base mechanism, the LBHB would stabilize the transition state, while in the kinetic isomechanism it allows for hydrogen tunneling.

1.9 HIV protease: a drug target

Efforts that have been undergoing for the past many years to design a vaccine have not provided satisfactory results due to virus extreme variability and escape mechanisms from the immune system [64]. In this scenario, therapeutic approaches

aimed at counteracting molecules essential for virus cycle have proved to be very useful. Kramer et al. showed that a frameshift mutation in the protease region of the *pol*-gene prevented cleavage of the gag polyprotein precursor, which is an essential step for the maturation of infectious HIV particles [65]. Subsequently, it was shown that inhibition of HIV polyprotein processing through the inactivation of the viral protease (either by mutation or chemical inhibition) leads to the production of immature viral particles that cannot undergo maturation to an infective form [66,67]. These observations led to the identification of HIV-1 protease as the prime target for structure-assisted or rational drug design.

1.10 FDA-approved HIV-1 protease inhibitors:

Since the identification of HIV as an etiological agent of AIDS, about twenty five different antiretroviral drugs (ARV) have been introduced into the clinical practice. Such drugs inhibit specific steps in viral lifecycle and thus the replication of the virus. ARVs are divided into five different classes depending on their viral target. The introduction of HIV protease inhibitors (PIs) in 1995 and the application of Highly Active Anti-Retroviral Therapy (HAART), *i.e.*, combination of PI with other ARVs, mainly inhibitors of the HIV reverse transcriptase (RT), resulted in a vastly decreased mortality and a prolonged life expectancy of HIV-positive patients. Unlike RT inhibitors, which are only effective in blocking viral replication in acutely infected cells, protease inhibitors can inhibit infectious virus production in both acutely and chronically infected cells [68].

Till now, U. S. Food and Drug Administration (FDA) have approved nine drugs that function as inhibitors of HIV protease: Saquinavir (SQV), Ritonavir (RTV), Indinavir (IDV), Nelfinavir (NFV), Amprenavir (APV), Lopinavir (LPV), Atazanavir (ATV),

Tipranavir (TPV) and Darunavir (DRV). All currently approved HIV-1 protease inhibitors (PIs) are competitive inhibitors that bind in the active site of the protease and, with the exception of TPV, all are peptidomimetics.

The knowledge of inhibitors of other aspartic proteases, such as rennin, greatly facilitated the development of the first generation PIs. Also, early availability of a wealth of information on the enzyme- inhibitor interactions in the protease active site led to the optimization of various lead inhibitors [23, 32, 47]. Initial designs of inhibitors were based on modification of the substrate backbone by replacing the hydrolysis unit in the peptide chain with non-cleavable dipeptide isosteres as core scaffolds and modification of peptidic side chains into non-peptide groups [69]. It was recognized at an early stage that mimics of the tetrahedral intermediate of the reaction catalyzed by the enzyme might be excellent inhibitors. The first inhibitors that were designed were based on gag and gag-pol substrates of HIV-1 protease, and in particular on the hydrolysis site Phe*Pro, specific of retroviral proteases.

Saquinavir (SQV), discovered by Roche [70], was the first HIV-1 PI approved by the FDA in December 1995 for the treatment of HIV-AIDS. It is characteristic of peptide-based HIV-1 protease inhibitors in which a transition-state analog mimics the tetrahedral intermediate formed during cleavage of the scissile Phe*Pro amide bond of the natural substrate. Lead optimization included replacement of the P1-P1' amide bond with non-cleavable hydroxyethylamine based dipeptide isostere and replacement of the P1' proline with a bicyclic decahydroisoquinoline. Indinavir (IDV) and ritonavir (RTV) were introduced soon afterward. The PIs are generally shorter than the peptide substrates and contain hydrophobic groups that bind within the hydrophobic pockets at the S2-S2' subsites of protease. The early PIs were designed with polar groups resembling those of the substrate peptide main chain. One of the

common features of these inhibitors is the presence of a secondary hydroxyl group which mimics the hydroxyl group of tetrahedral reaction intermediate. This group makes critical interactions with the catalytic Asp25/25' residues of the protease and is required for tight inhibitor binding with the protease [70]. The binding affinity of the PIs varies from nanomolar to picomolar. The later inhibitors amprenavir, tipranavir and darunavir were designed to reduce the peptidic backbone but retain the central hydroxyl group [71, 72]. Due to its numerous side effects, RTV is no longer used as a PI on its own. However, RTV is a potent inhibitor of cytochrome P450 (CYP-450) 3A4 isoform [73] which is a liver enzyme involved in drug metabolism. Because of this side activity, low

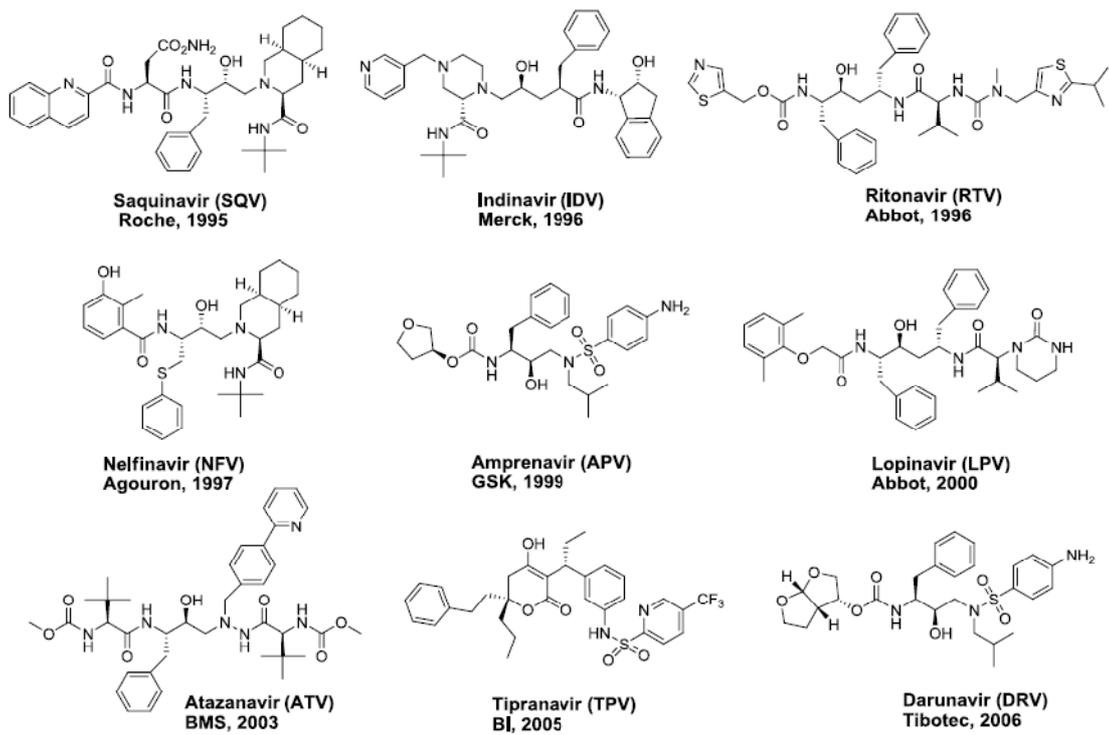


Figure 1.8: FDA approved HIV protease inhibitors

dose of RTV is currently used as a boosting agent in HAART therapy with other PIs. Another common feature in the complexes between peptidomimetic inhibitors and HIV-1 protease is a conserved water molecule, known as flap water, which mediates hydrogen bonds between the P2/P1' carbonyl oxygen atoms of the inhibitors and the amide groups of Ile50/Ile50' of the enzyme. TPV, unlike peptidomimetic inhibitors, makes direct hydrogen bond interactions with Ile50/50' in the flap region of the protease, replacing the conserved flap water. The development and clinical introduction of HIV-1 PIs is regarded as a major success of structure-based rational drug design [32].

1.11 Drug resistance:

The emergence of drug resistance against protease inhibitors limits the long term clinical efficacy of these drugs. The high error rate of HIV reverse transcriptase (about 1 in 10,000 bases) and the high rate of replication of the virus (10^8 – 10^9 virions/day) are responsible for the rapid mutation and the selection of drug-resistant viruses [74]. Computer simulations of mutational stability suggest that even when low level selection pressure is combined with high replication rate compared to the duration of infection, drug resistant mutants will rapidly outgrow the wild type virus [74]. The virus evolves to accumulate a multitude of mutations within the protease that prevent PIs from binding to the protease. More than half the residues within the protease mutate in different combinations and lead to drug resistance [74-76].

From the structural point of view, drug resistant mutations can be classified as active site or non-active site, depending upon whether they occur inside or outside the substrate binding cleft. As the functional HIV-1 protease enzyme is a symmetric dimer, both

monomers contribute to substrate binding. The active site region is primarily formed by residues 25–32, 47–53 and 80–84 from both the monomers. Under protease inhibitor therapy, a majority of initial mutations arise within the active site of the enzyme, directly affecting inhibitor binding and are the primary cause of resistance to PIs. However, the protease mutates extensively in the regions beyond active site, and these non-active site mutations have been known to greatly contribute to drug resistance. The mechanism by which the mutations outside the active site confer resistance remains elusive. Some of these mutations are primary drug resistant mutations and others have been suggested to contribute to drug resistance when present along with other major mutations.

Of the 99 residues in each monomer, about 37 are known to be conserved and 17 residues mutate even in the absence of any inhibitor, ie, they are the sites of non-treatment related polymorphisms [74-77]. About 45 residues in each monomer have been implicated in drug resistance. Out of these 45 residues, mutations of 26 residues have been shown to significantly decrease susceptibility to one or more PIs and the others are polymorphic mutations that occur more frequently when associated with inhibitor therapy [76-78]. Furthermore, almost 60% of these 26 positions fall outside the active site region. Thus excluding the invariant positions and including the polymorphic sites associated with drug resistance, an overwhelming 60–63% of the sequence has been known to vary in patient isolates and almost 40–45% of the protease sequence is implicated in contributing to drug resistance [74-77]. Furthermore, resistance mutations in the protease can be accompanied by mutations in the viral polyprotein cleavage sites to improve the viral fitness. Mutations in the gag precursor cleavage sites NC/p1 and p1/p6 follow the drug resistant mutations in protease [79]. Also, protease drug resistance can emerge due to

mutations in the gag substrate alone, without corresponding mutations in the enzyme [80]. In addition, recent reports indicate a role for amino acid insertions in the drug resistance of HIV protease [81-83].

1.12 Aims:

In view of the burgeoning problem of resistance against anti HIV protease drugs, there is a need for continuous improvement of existing drugs and design of new HIV protease inhibitors. This requires structural understanding of: (1) chemical interactions between HIV protease and its substrates/products (2) mechanism of action of HIV protease, and (3) mechanism of drug resistance. X-ray crystallography is a powerful tool to get such information at atomic level. Keeping these objectives in mind, the present work aims at crystallographic studies on complexes between substrate-peptides/inhibitors and native/drug resistant mutants of HIV-1 protease.

Chapter 2

Protein crystallographic methods

Diffraction from a periodic array of molecules in the crystal can be interpreted as 'reflections' from a set of lattice planes. Bragg's law predicts the angle of reflection ' θ ' of any diffracted ray from specific atomic planes whereby

$$2d \sin \theta = n\lambda$$

Where, d is the interplanar spacing of that set of planes, λ is the wavelength of the X-rays and n is an integer. Bragg's law is a special formulation of the Laue equations which define the condition for diffraction (constructive interference) to occur:

$$\mathbf{a} \cdot \mathbf{S} = h$$

$$\mathbf{b} \cdot \mathbf{S} = k$$

$$\mathbf{c} \cdot \mathbf{S} = l$$

where h, k, l are integers that turn out to be the Miller indices defining Bragg plane hkl of reflection (or diffraction spot), a, b and c are the lattice vectors and \mathbf{S} is the difference between unit vectors along the incident and reflected ray wave vectors for the hkl plane.

The intensity measured for a given reflection hkl is proportional to $|F(hkl)|^2$ where

$$F(hkl) = \sum_{j=1}^N f_j \exp [(2\pi i (hx_j + ky_j + lz_j))]]$$

And f_j is the atomic scattering factor for X-rays for the j th atom of coordinate (x_j, y_j, z_j) expressed as fractions of the cell a, b, c . This equation is the structure factor equation. The structure factor is a complex quantity characterized by a magnitude and a phase; $|F_{hkl}|$ and α_{hkl} . The Fourier inverse of this equation is the electron density (ρ) equation:

$$\rho(xyz) = \frac{1}{V} \sum_{hkl} |F_{hkl}| \exp(i\alpha_{hkl}) \exp[-2\pi i (h_x + k_y + l_z)]$$

Whereby, if the amplitude $|F_{hkl}|$ and phase (α_{hkl}) of the structure factor are known for all hkl planes or reflections, then the electron density can be calculated for all points (x, y, z) in the cell and so the crystal structure is then solved. Of course it is impossible to measure all h, k, l reflections so the summation is usually terminated with a finite number of terms at a certain Bragg resolution limit. The problem of phase determination is the fundamental one in any crystal structure determination since $|F_{hkl}|^2$ is the measured quantity and not α_{hkl} .

Protein structure determination involves the following steps: (i) Preparation of purified protein samples (ii) Crystallization (iii) Measurement of diffraction data (iv) Solution of phase problem (v) Phase and electron density calculations (vi) Map interpretation and model building (vii) Model refinement (viii) Validation

2.1 Protein expression and purification:

Advent of recombinant DNA technology was one of the major advances in the field of protein crystallography. Various kinds of expression systems consisting of a suitable vector carrying the gene insert and corresponding hosts, ranging from bacteria to eukaryotic cells such as yeasts, insect cell lines, and mammalian cell lines are available.

With advanced cloning techniques viz ligation independent cloning, topoisomerase cloning, and Gateway cloning, expression systems can be switched easily. Structural studies by crystallography not only require soluble protein but also conformationally homogenous material. Solubility as well as the demand for conformational purity requires careful selection of suitable expression host. E. coli based bacterial systems are the most common and easy to use. But they are not capable of eukaryotic posttranslational modifications and generally have difficulty expressing properly folded proteins containing multiple disulfide bonds. Yeast cells possess posttranslational modification machinery. But, glycosylation patterns are different for yeasts and mammalian cells, and hyperglycosylation is a problem with some yeast systems. Glycosylation of mammalian proteins is a major source of conformational heterogeneity. Glycosylation-deficient host cell lines can reduce this problem. In vitro cell free transcription-translation systems allow full control over the expression conditions, including addition of folding chaperones. Recombinant DNA techniques allow fusing a number of useful sequences to the target sequence. The most common fusion partners are small affinity purification tags and solubility enhancing fusion partners as well as chaperones. For anomalous phasing methods, incorporation of Se-Met instead of regular methionine residues provides heavy atom labels while maintaining structural isomorphism to the native protein. One method of incorporating Se-Met is expression in a methionine auxotroph host strain and supplementing the minimal medium with Se-Met. In the metabolic inhibition technique, bacteria are grown in regular medium and Met biosynthesis inhibiting amino acids Ile, Lys, and Thr as well as Se-Met are added to the medium before induction.

2.2 Crystallization:

The most popular method of protein crystallization is vapour diffusion in the hanging-drop or sitting-drop format [1]. Other techniques include counter diffusion of the protein and precipitant solutions through membranes, gel or capillaries; microbatch crystallization and dialysis. Microfluidic chips or thin walled capillaries are used for free interface diffusion. The advantage of free interface diffusion methods is wide coverage of the crystallization phase space, and very little material is required in the case of microfluidic chip methods. Macromolecules usually require some precipitant to initiate nucleation and crystal growth. The most popular precipitants are inorganic salts (ammonium sulfate, sodium chloride etc.) and organic polyols (polyethylene glycol of various sizes, methylpentanediol etc.). A number of other parameters can be varied to optimize the crystallization conditions, e.g. the buffer type and pH, temperature, concentration of protein etc. Some proteins require small amounts of special additives such as dioxane, phenol, 2-propanol or various cofactors to produce good-quality crystals. Membrane proteins require the presence of detergents for crystallization, to avoid aggregation and micelle formation. The ingenious way of dealing with membrane proteins is crystallization in the scaffold of the lipidic cubic phase. Various sets of crystallization screening conditions selected by sparse-matrix sampling [2] have been proposed and many ready-made screens are available commercially (Hampton Research, <http://www.hamptonresearch.com/>; Emerald Biostructures, <http://www.emeraldbiostructures.com/>; Jena Bioscience, <http://www.jenabioscience.com/>; Molecular Dimensions, <http://www.moleculardimensions.com/>); Qiagen, <http://www.qiagen.com/>). These days,

crystallization robots can effectively perform the initial crystallization trials in a multitude of conditions with small volume of protein sample (of the order of tens of nanolitres). Crystallization robots can be linked to the imaging systems that automatically inspect all crystallization setups through an optical microscope and take images at specified intervals.

2.3 Measurement of diffraction data:

2.3.1 X-ray sources:

X-rays for crystallographic studies are typically generated by bombarding a metal target with an energetic beam of electrons. The electrons produced by heating a metal filament are accelerated towards the target by a large applied electrical potential between the filament and the target. When the beam of electrons hits the target (or anode), this rapid deceleration of electrons causes the emission of X-ray radiation and a large amount of heat. Two types of X-rays emitted in this process are a continuous band of white radiation or Bremsstrahlung as well as a series of discrete lines that are characteristic of the target material. Copper anode is most often used for proteins since it is hard, an efficient conductor of heat and the CuK α emission is relatively intense. The wavelength of the X-rays produced is 1.54 Å.

2.3.1.1 Sealed-tube X-ray generator:

It uses a stationary anode. Since the amount of heat generated in the system is dissipated through water cooling, limited power can be applied to the tube. This limits the flux of X-rays generated. Sealed-tube sources with their low maintenance requirements are generally adequate for small molecule crystallography but their use for the

macromolecules is limited.

2.3.1.2 Rotating anode X-ray generator:

In these generators, heat dissipating ability of the system and thus the X-ray beam intensity is increased by rotating the anode surface so that the beam of electrons continually hits a new region of the anode [3]. Rotating anode generators require a considerable amount of maintenance to replace filaments, and repair or replace the anode bearings as well as vacuum and water seals. To keep from burning the filament, it must remain in a high vacuum. The anode with its constant flow of cooling water must be continuously rotating at speeds of 6000 rpm or more. Special Ferro-fluidic seals are used to maintain the vacuum along the rotating shaft of the anode. The recently introduced chromium based ($\lambda=2.23 \text{ \AA}$) rotating anode [4] appropriately answers the current interest in phasing based on the anomalous scattering signal of sulphur and other relatively light elements.

2.3.1.3 Micro-focus generator:

Another type of sealed-tube source that produces beam fluxes comparable to rotating-anode systems is a micro-focus generator [5]. Because heat dissipates rather quickly in a metal block, manufacturers have found that when the focal size is reduced to 10-300 μm then the power can be increased to make the beam flux much higher than for normal or even fine-focus sealed tube sources. One of the great advantages of a micro-focus radiation source is that the electrical power needs are in the range of 30-80 Watts, and not the 2-3 kWatts that are required of a typical sealed tube generator, or the 3-12 kWatts required by a rotating anode generator.

2.3.1.4 Metal jet anode:

A new method of generating X-rays uses an electron-impact beam impinging on a stream of liquid gallium. The advantage of the metal-jet anode is that the maximum of the electron-beam power density is significantly increased. (<http://www.excillum.com>)

2.3.1.5 Synchrotron radiation:

Whenever electrically charged particles such as electrons or positrons of high energy are kept under the influence of magnetic fields and travel in a pseudo circular trajectory, synchrotron radiation is emitted. For relativistic electrons with energy E , the electromagnetic radiation is compressed into a fan-shaped beam tangential to the orbit with a vertical opening angle $\Psi = mc^2/E$, i.e., 0.1 mrad for $E=5$ GeV. As this fan rotates with the circulating electrons or positrons, a stationary observer will see n flashes of radiation every $2\pi R/c$ s, the duration of each flash being less than 1 ns. The spectral distribution of synchrotron radiation extends from the infrared to the X-ray region. An important parameter is the median of the distribution of power over the spectral region, called the “critical photon energy” E_c , which divides the power spectrum into two equal parts. Taking the wavelength λ instead of the photon energy E , the critical wavelength is given by,

$$\lambda_c = 18.64/(BE)^2$$

Where, B ($= 3.34 E/R$) is the magnetic bending field in Tesla, E is in GeV, and R in meters.

The particles are injected into the storage ring directly from a linear accelerator or through a booster ring. Storage ring consists of an evacuated chamber around which two

basic types of magnets are arranged: focusing and bending magnets. A focusing magnet's multiple poles set up a non-uniform magnetic field that acts like a lens, confining the electrons in a tight beam as they travel around the chamber. A bending magnet typically establishes a uniform magnetic field that is at right angles to the direction in which electrons are moving. Such a field bends the path of the electrons. Synchrotron radiation produced by electrons curving in bending magnets emerges into tangential beamlines. An oscillating electromagnetic field established in a radio-frequency cavity provides energy to maintain the particles at relativistic speeds. They also divide the circulating electron beam into electron bunches. As the particle beam traverses each magnet, the path of the beam is altered, and synchrotron radiation is emitted. Synchrotron radiation is highly polarized. In an ideal ring, where all electrons are parallel to one another in a central orbit, the radiation in the orbital plane is linearly polarized, with the electric vector lying in this plane. Outside this plane, the radiation is elliptically polarized. The synchrotron radiation can be channelled through different beamlines for use in research. The ring may also include other magnetic devices, known as wigglers and undulators that substantially increase the brightness of the radiation. These insertion devices cause electrons to follow a sinusoidal path instead of a curved one by establishing a series of magnetic fields that alternate in polarity and that are perpendicular to the electron's direction of travel. A wiggler enhances the brightness of the radiation produced by a given electron beam by a factor roughly equal to twice the number of full oscillations the beam undergoes. A wiggler consists of one or more dipole magnets with alternating magnetic field directions aligned transverse to the orbit. The critical wavelength can thus be shifted towards shorter values because the bending radius can be decreased over a short section. Wigglers cause

a continuous spectrum of radiation which can be easily tuned to the desired wavelength. Such a device is called a “wavelength shifter”. A series of N dipole magnets constitutes a multipole wiggler. The flux of N dipoles adds up to N times the flux of a single dipole. A multipole wiggler becomes an undulator if the magnet poles have a short period and $2\alpha < \Psi$. Interference takes place between radiation of wavelength λ_0 emitted at two points λ_0 apart on the electron trajectory. The spectrum at an angle θ to the axis observed through a pinhole has a peak at a specific wavelength and a few harmonics. The importance of synchrotron radiation for macromolecular crystallography lies in the high brilliance (photons $\text{s}^{-1} \text{mrad}^{-2} \text{mm}^{-2}$ per $\Delta\lambda/\lambda$) of the beam, the high intensity, and the tunability of the wavelength in the relevant range from 0.5 to 3.0 Å. The time structure of the beam is of interest for time-resolved crystallography. The particles circulate in bunches with widths of 50 to 150 ps, and repeat every few microseconds. About 22 synchrotron radiation facilities equipped with beamlines for macromolecular crystallography are available worldwide, and are operated at energies from 1.5 to 8 GeV for third-generation machines. The European Synchrotron Radiation Facility (ESRF) storage ring is operated at 6 GeV and has a circumference of 844.39 m. Its critical wavelength, λ_c , is 0.6 Å. In India, Indus-2 is a 2.5 GeV Synchrotron Radiation Source with critical wavelength of about 2 Å and a circumference of 172.4743 m.

2.3.1.6 Free electron laser (FEL):

The Free electron laser works on SASE (self amplified stimulated emission) principle and requires a high energy electron beam that is sent through a very long undulator. The alternating magnetic field vectors in the undulator impose alternating accelerating forces on the electrons, which emit X-rays in phase. The resulting coherent X-ray flash is

extremely brilliant, and could be potentially used for determining molecular envelopes by single molecule scattering. X-ray FEL facilities are being built at SLAC, Stanford, CA, USA and DESY, Hamburg, Germany.

2.3.1.7 Tabletop synchrotrons:

These machines are still in the developmental stages. Two novel compact sources of X-rays have been reported in the last two decades:

2.3.1.7.1 Compton sources:

In a Compton source, electrons in a small room sized storage ring (MeV range) are undulated by a high intensity laser focussed into the oncoming electron beam. In the collision zone, the electrons inelastically backscatter high energy X-ray photons (Compton Effect). The resulting X-rays are emitted in a narrow energy bandwidth range of a few hundred eV and can be tightly focused and monochromated, but currently the repetition rate of commercial high intensity (teraWatt) lasers is not high enough to produce continuous X-rays sufficiently intense for routine use in diffraction experiments. (<http://www.lynceantech.com>)

2.3.1.7.2 Laser wakefield acceleration:

A high power femtosecond laser pulse travelling through dilute plasma will locally induce a high degree of charge separation as the electric field of laser pulse strips electrons from the nuclear cores of the atoms making up the plasma. Because electrons are lighter, they move away quickly, resulting in a high degree of charge separation and an associated very strong potential gradient. This wakefield propagates through the

plasma with the same group velocity as the laser pulse and will accelerate any electrons that are injected into it.

2.3.2 X-ray detectors:

2.3.2.1 X-ray photographic film:

It is a plastic sheet coated with an emulsion in which silver halide (generally AgBr) crystals are dispersed. When an X-ray photon strikes the film, a small number of silver ions in the excited crystallites get converted to metallic silver. For visualizing the diffraction spots, X-ray film has to be developed, washed, fixed and dried manually. Advantages of X-ray film include large area and high resolution while disadvantages of using film are poor dynamic range, high background, moderate sensitivity, time-consuming process of development and scanning.

2.3.2.2 Image plate:

The imaging plate is an area detector, qualitatively similar to the photographic film. It consists of a support (either flexible plastic plate or metal base) coated with an emulsion, which contains crystallites of barium fluoride bromide or barium fluoride iodide, with artificially introduced impurities of Eu^{2+} ions. During the preparation of these crystals a large number of vacancies are created at the sites of fluoride and bromide (or iodide) negative ions. When a photon strikes the imaging plate, Eu^{2+} ions are ionized further to Eu^{3+} and the “detached” electrons are raised to the conduction band. When so excited, the electrons are trapped at the vacancies and thereby produce temporary colour centres. When plate is scanned by a He–Ne laser ($\lambda=633$ nm), the trapped electrons fall down to the valence band via an excited state, and recombine with Eu^{3+} to form Eu^{2+} . This

transition is accompanied by a release of energy, which corresponds to the emission of blue light ($\lambda=390\text{nm}$). The intensity of this photo stimulated luminescence, measured with a photomultiplier, is proportional to the intensity of the X-rays which gave rise to the colour centres. Advantages of image plate are large area, wide dynamic range, high sensitivity, high resolution and low background while disadvantages are slow readout and non-linear response.

2.3.2.3 Charge-coupled devices (CCD):

This type of detector is a semiconductor in which incident radiation produces electron-hole pairs; the electrons are trapped in potential wells and then read out as currents. An X-ray phosphor (commonly $\text{Gd}_2\text{O}_2\text{S: Tb}$) is attached to a fibre-optic faceplate, which is tightly connected to a fibre-optic taper. The X-ray-sensitive phosphor surfaces at the front convert the incident X-rays into a burst of visible-light photons. The light emitted by the screen is conducted by a tapered bundle of optical fibres and strikes the array of pixels, each of which is a metal-oxide-semiconductor (MOS) capacitor. When a light photon strikes an MOS pixel, an electron is emitted owing to the photoelectric effect and is stored in the capacitor (an electron-hole pair is produced). Therefore, the charge distribution throughout the whole CCD follows the distribution of radiation scattered from the crystal. The charge is subsequently transferred to an electronic circuit, and converted into an array of pulses the height of which is proportional to the intensity of X-rays that fell on the phosphor screen. This digital information is transferred to a computer, which records the pattern of diffracted intensity on a relative scale. The CCD must be cooled to temperatures ranging from -40°C to -90°C , depending on the various systems. The great advantage of CCD detectors is their short readout time, which lies in

the range from 1 to a few seconds. Advantages of CCD include wide dynamic range, relatively fast, high resolution while disadvantages are small area, expensive, distortions, blackouts.

2.3.2.4 Pixel array detector (PAD):

The PAD is a two-layer device, consisting of an X-ray sensitive array of silicon photodiodes bonded to a pixelated complementary metal–oxide–semiconductor (CMOS) chip (with pixels limited in size to the same area as the detector pixels). Each pixel of the electronics layer integrates the current generated by the X-rays converted in the photodiode. Here, no converting layer and taper are needed. The advantages are high dynamic range, low dead time and fast readouts.

2.3.3 Data collection methods:

2.3.3.1 Precession method:

Precession photography involves making a crystal precess at a fixed angle around a defined axis. If the crystal is precisely aligned such that a real space unit cell axis lies along the rotation axis, a precession photograph can be arranged to provide an undistorted view of a single plane through diffraction space. Therefore, symmetry in the reciprocal lattice can directly be inferred from precession photographs. However, since isolating a single reciprocal lattice layer involves introducing a metal layer screen that blocks most of the diffraction that is happening and only allows passage of that from the desired layer, it is not very efficient way of collecting data. Precession photography was used in the early days of protein crystallography, before advanced algorithms made auto-indexing and interpretation of oscillation photographs very straightforward.

2.3.3.2 Screenless oscillation (rotation) method:

This method involves rotating the crystal through a small solid angle about a single axis. The pattern recorded corresponds to planes in reciprocal (diffraction) space slicing through the Ewald sphere so that only a limited amount of each lattice plane is in diffraction condition within the oscillation range. Entire datasets are built up by collecting contiguous series of such images to form a solid volume of rotation. This method was used to collect all the data in the present work. The ability to auto-index oscillation data has considerably enhanced the usability of this method.

2.3.3.3 Weissenberg method:

Weissenberg data collection combines the rotation/oscillation method with a coupled translation of the detector along the rotation axis. This is used to reduce the overlap of spots that can occur with larger oscillation ranges or larger unit cells. In practice you need to align your crystal accurately in order to make the most of Weissenberg photography and data collection is rather tedious and the diffraction pattern more difficult to interpret. Weissenberg cameras use cylindrical drums coaxial with crystal rotation axis to house the detecting medium.

2.3.3.4 Laue (polychromatic) method:

In contrast to methods that have been discussed before, all of which use monochromatic X-rays, Laue photography specifically uses polychromatic X-rays over a wide wavelength range. The advantage of Laue is that many diffraction maxima are in diffraction condition at the same time, so we can collect the data in one or just a few images. Laue data collection held promise in the early days, especially for high-symmetry space groups and time-resolved studies, but the inherent difficulties in indexing the

diffraction images from these systems, with multiple overlapped spots from multiple wavelengths, has essentially rendered it useless for routine data collection.

2.4 Oscillation data processing:

Data processing involves specification of the diffraction experiment, recognition of image background and gain, localizing diffraction spots, extraction of basis vectors of the lattice explaining the spots, refinement of lattice parameters, determination of reflection profile, integration, assessment of data quality, space group determination, data corrections and scaling.

2.4.1 Autoindexing:

One of the most significant advances in data processing was the introduction of autoindexing. Prior to this, the unit cell was normally determined by precession method since in this method, the recorded patterns represented undistorted single layers in the reciprocal lattice. This process was both laborious and time consuming. In contrast, in the screenless rotation/oscillation method in vogue today, reciprocal lattice is represented on the two dimensional flat detectors in a distorted way. Two methods of autoindexing are in common use in different data processing programmes. Both require mapping the two dimensional detector co-ordinates onto a three dimensional reciprocal lattice. Both methods give unit cell dimensions and the crystal orientation. In Difference vectors method, vectors between reciprocal lattice points that occur most often and reinforce each other are determined. This approach is used in the programme XDS [6]. In the Fourier Transform methods one calculates either a 3D fast Fourier transform from the reciprocal lattice points or a 1D fast Fourier transform from the projections of the reciprocal lattice

onto directions, systematically varied around a hemisphere in reciprocal space. In either case, linear transformation operations are applied, to the obtained primitive triclinic basis solution to generate Bravais Lattices, to produce a putative list of solutions, and score these according to the distortion required.

2.4.2 Integration:

There are two basic types of integration procedures: Summation integration (box-sum) and Profile fitted integration.

2.4.2.1 Summation integration:

The counts in each pixel are added in the reflection spot and background underneath the spot is subtracted. For Integration in the absence of X-ray background or detector noise, each pixel is assigned to the nearest reciprocal lattice point. All pixel counts are added to obtain the integrated intensity. There is no penalty (in $I/\sigma(I)$) for including pixels beyond the physical extent of the diffraction spot. This situation is never realized in practice, even for very strong spots. For Integration in the presence of X-ray background, we need to subtract the X-ray background. This requires the definition of the peak and background regions; this definition may need to vary across the detector to allow for the variation in spot shape and size. The size of a spot does not depend on its intensity.

2.4.2.2 Integration by profile fitting:

Summation integration is unbiased (providing peak and background regions are correctly defined) but gives poor signal to noise ($I/\sigma(I)$) for weak reflections. Profile fitting can improve the estimation of weak intensities. A profile describes a peak boundary or shape,

as well as the intensity distribution within that shape. Reference profiles are created from the observed profiles of a collection of reflections. Next, each observed profile is fitted with a least-squares procedure to the reference profile and the estimated intensity derived from this minimization.

Two- and three-dimensional integration; in two dimensional integration as implemented in programs MOSFLM and HKL [7, 8], each image is considered separately and reflections are integrated in two dimensions first and integrated contributions from individual images are summed in a subsequent step. On the other hand, three dimensional integration packages, viz XDS, d*TREK [6, 9] work with multiple images. These programs treat all reflections in three dimensions. Two dimensions are from the area detector itself, while the third dimension arises from the rotation angle of the crystal. One advantage of three-dimensional integration is that the reflection centroid in the rotation-angle direction at the time of integration is known accurately. In the two-dimensional case, a subsequent step known as post-refinement determines accurate rotation angles for reflections. Accuracy is important when predicting reflection positions and fitting reflection profiles. In the three-dimensional profile fitting employed by XDS, the profiles of individual spots are mapped back to a new coordinate system based on the scattering vector for each spot [6]. This procedure eliminates the wide variation in the widths of different reflections in the ϕ direction, simplifying the tasks of both forming the standard profiles and fitting these to individual reflections. The same approach has been adopted in d*TREK [9].

2.4.3 Intensity estimation:

If p_i is the predicted profile normalized in such a way that $\sum_i p_i = 1$, where sum is over all the pixels i within a profile, the reflection intensity I is estimated as

$$I = \frac{\sum_i (c_i - b_i)p_i/V_i}{\sum_i p_i^2/V_i}$$

which minimizes the function

$$\psi(I) = \sum_i (c_i - I \cdot p_i - b_i)^2/V_i$$

b_i ; c_i ; V_i are background, contents and variance of pixels.

2.4.4 Scaling and merging:

Scaling and merging the data is the next step following integration. This is an important step because it attempts to put all observations on a common scale and, it provides the main diagnostics of data quality. The integrated intensities of reflections need to be corrected by various factors arising from the following:

1. Changes in beam intensity and illuminated crystal volume
2. Absorption of incident and diffracted beams
3. Radiation damage
4. Variations in sensitivity within detector plane.
5. Different crystal sizes and crystalline order if data come from several crystals

Most of the scaling procedures in macromolecular crystallography [10-12] exploit the prior knowledge that symmetry-related and duplicate measurements of a reflection are equal to determine a correction factor for each observed intensity. However, the scaling programs differ in the details of their scaling models, i.e. the parameterization and methods used for determination of the correction factors. These procedures make the data internally consistent. After scaling, the remaining differences between observations can be analyzed in the form of various R-factors to give an indication of data quality.

2.4.5 Post refinement:

The method derives complete intensities of reflections that are only partially recorded on an image, from accurate estimates for the fractions of observed intensity: the 'partiality'. The partiality of each reflection can always be calculated as a function of orientation, unit-cell metric, mosaic spread of the crystal and model intensity distributions. The accuracy of the estimated full reflection intensity obviously then strongly depends on a precise knowledge of the parameters describing the diffraction experiment. Usually, symmetry-related fully recorded reflections can be found for many of the partial reflections and the list of such pairs of intensity observations can be used to refine the required parameters using a least-squares procedure.

2.5 Solution of phase problem:

2.5.1 Direct methods:

Direct methods are based on the assumptions that the electron density is positive everywhere in the unit cell (positivity) and that all atoms are resolved (atomicity); which lead to the tangent formula

$$\tan \alpha_{\mathbf{h}} = \frac{\langle E_{\mathbf{h}'} E_{\mathbf{h}-\mathbf{h}'} \sin(\alpha_{\mathbf{h}'} + \alpha_{\mathbf{h}-\mathbf{h}'}) \rangle_{\mathbf{h}'}}{\langle E_{\mathbf{h}'} E_{\mathbf{h}-\mathbf{h}'} \cos(\alpha_{\mathbf{h}'} + \alpha_{\mathbf{h}-\mathbf{h}'}) \rangle_{\mathbf{h}'}}$$

Where, E represents normalized structure factor corresponding to point atoms at rest. Tangent formula postulates that squaring in direct space followed by a Fourier transformation leads to structure factors that are proportional to the original structure factors. The phases are identical. Small set of starting phases are used to construct a more complete phase set.

Requirement of very high resolution data ($< 1.2 \text{ \AA}$) has limited the usefulness of direct methods in protein crystallography. However, direct methods are used routinely to find the heavy-atom substructure, in computer programmes Shake-and-Bake [13], SHELXD [14] and SHARP [15], and even subsequent phase determination from the substructure with programs such as SHELXE [16] and ACORN [17].

2.5.2 Experimental phase determination:

Experimental methods for phase determination include isomorphous replacement and anomalous scattering methods which rely on information that can be derived from small

differences between diffraction datasets of native and derivatized crystals. The first step in both methods is the location of the heavy atoms or anomalous scatterers, generally termed the substructure, in the crystallographic asymmetric unit.

2.5.2.1 Isomorphous replacement:

The amplitudes of reflections are measured for the native crystal, $|F_P|$, and for the derivative crystal, $|F_{PH}|$. The isomorphous difference, $|FH| = | |F_{PH}| - |F_P| |$, is then used as an estimate of the heavy-atom structure-factor amplitude to determine the heavy-atom positions using Patterson and/or direct methods. Once located, the heavy-atom parameters (xyz positions, occupancies and Debye Waller thermal factors B) can be refined and used to calculate a more accurate $|FH|$ and its corresponding phase α_H . The native protein phase, α_P , can be estimated using the cosine rule, leading to two possible solutions symmetrically distributed about the heavy-atom phase.

2.5.2.2 Anomalous differences:

The anomalous or Bijvoet differences can be used in the same way as the isomorphous differences to locate the anomalous scatterers and to derive the phase information. Anomalous scattering from the same crystal can be used to break the phase ambiguity in a single isomorphous replacement experiment, leading to SIRAS (single isomorphous replacement with anomalous scattering). Because of the 90° phase advance of the f' term, anomalous scattering provides orthogonal phase information to the isomorphous term. There are two possible phase values symmetrically located about f' and two possible phase values symmetrically located about FH . Multiwavelength anomalous

diffraction (MAD) provides orthogonal dispersive and anomalous differences from the same crystal and phase ambiguity can be resolved directly from the phasing equation.

2.5.3 Molecular replacement:

Molecular replacement method is commonly used to solve structures for which a homologous structure is already known. Each molecule needs six parameters to define its orientation and position: three rotation angles and three translations (e.g. α , β , γ ; t_x , t_y , t_z). If there are N molecules in the asymmetric unit, then a total of $6N$ parameters are required to define the solution. This number is enormously reduced if the two searches can be separated and the translation search is carried out only for the best solutions found in the rotation search. For this reason, most programs split the search in this way and pick a relatively small number of good solutions from the rotation search to test in translation searches. But in a six-dimensional search or the second three-dimensional search, all parameters (α , β , γ ; t_x , t_y , t_z) are defined at each search point, so the correct structure factor $F_{\text{calc}}(\alpha, \beta, \gamma; t_x, t_y, t_z)$ can be calculated and then compared with the observed F_{obs} in a scoring function. However, in the first three-dimensional search on rotation, the correct $F_{\text{calc}}(\alpha, \beta, \gamma)$ cannot be calculated with an unknown translation and so cannot be compared directly with F_{obs} . There are two ways to overcome this problem, using different approaches and different scoring functions: (i) Traditional rotation searches are based on the Patterson function, scoring the overlap between observed and model Pattersons in a region around the origin where the function is dominated by self-vectors from within the molecule which are independent of translation (ii) Maximum-likelihood methods use a statistical approach in reciprocal space to average over all possible values of the unknown translation.

2.6 Electron density interpretation- Molecular model building:

These days, molecular model is built into the electron density map almost exclusively with graphical model building programs. When interactive computer graphics were first introduced, the protein models were built by fitting individual atoms into the displayed electron-density maps. In the currently used programs, many tasks are performed in a semiautomatic fashion and the human operator needs to only check visually the correctness of the resulting action. The most widely used graphics systems are O [18], Xtal View [19], QUANTA [20], COOT [21, 22] and MAIN [23, 24]. In fact, graphics programs are used mainly to check the correctness of the model and its fit into the electron density or to modify selected fragments of the structure, since the main task of the primary construction of the protein chain is very often performed without human intervention by automatic model building programs such as ARP/wARP [25], RESOLVE [26], MAID [27] and BUCCANEER [28]. These programs are evolving constantly and becoming more powerful and capable of tracing correct models in low-resolution maps. Most of these programs are moving towards integration of tasks which traditionally used to be executed sequentially, e.g. real-space fitting and refinement with concomitant validation etc.

2.7 Crystallographic refinement:

During refinement, the parameters describing a molecular model are adjusted so that the fit between discrete experimental observations and their computed values is optimized, as calculated from a target function. Observations include everything known about the crystal prior to the refinement. These may include experimental observations such as

structure factor amplitudes or general observations which are valid for all the models such as standardized stereochemistry. Stereochemical information consistent with their scatter in accurate structures is applied in the form of restraints in the refinement to improve data to parameter ratio. But at higher resolutions this information can become redundant and can be dropped. That is why geometric restraints are weighted relative to the experimental data in the refinement process. Traditionally a set of parameters describing a macromolecular model include three x,y,z coordinates describing the location of an atom in the unit cell, B factor describing the extent of vibration of an atom about its mean position, and occupancy. A model can be parameterized in different ways depending upon the amount of data we have to improve the data to parameter ratio. Eg, occupancy can be constrained to 1.0 as most atoms in the crystal are chemically bonded. But this assumption may not be true for individual water molecules. Likewise B factors can be refined as isotropic or anisotropic. Isotropic B factor means that atom vibrates equally in all directions, therefore can be described by only one number. While anisotropic B-factor describes the vibration of the atom within an ellipsoid centred at the atomic coordinate. Six parameters are required to describe such an ellipsoid. Other forms of parameterizations used in macromolecular refinement are described below:

2.7.1 Rigid body parameterization:

Rigid body parameterization is used when one usually knows the internal structure of the molecule but its orientation and location in the crystal is unknown. So, a set of six parameters, three for rotation which rotates the molecule into the correct orientation and three for translation which places the correctly oriented molecule in the unit cell are refined.

2.7.2 Non crystallographic symmetry (NCS) constrained parameterization:

When asymmetric unit contains multiple copies of the same molecule and data resolution is limited or in early stages of refinement, it is useful to use NCS constrained parameterization. In this case only one of the molecules needs to be parameterized and other can be generated from the refined copy by application of NCS operators. In programme SHELXL, there is an option for restraining NCS by suggesting that torsion angles of related molecules are similar rather than the position of atoms being similar after rotation and translation.

2.7.3 Torsion-angle parameterization:

There are on an average about five torsion angles and about eight atoms per amino acid. Changing from an atomic model to torsion angle model decreases the number of parameters from 24 positional to five angular. This greatly improves the data to parameter ratio.

2.7.4 TLS B-factor parameterization:

This parameterization helps describing anisotropic motion of atoms with fewer parameters. Group of atoms are assumed to move as a rigid unit. This motion is described by three matrices, one for translation vibration of the group, second for liberation (wobbling) about a fixed point and third for translation and liberation that occur in concert. Here, twenty parameters are used to describe the motion of a group of atoms. Therefore if more than three atoms are considered in the group, TLS parameterization improves the observation to parameter ratio.

2.7.5 Hydrogen atoms and riding restraints:

The locations of hydrogen atoms covalently bound to the refined N, C, and O atoms of amino acids and ligands are well known from basic stereochemistry and can thus be computed. As these hydrogen atoms move rigidly with the atoms they are attached to, they do not need to be refined and add no parameters to the refinement. Although the empirical stereochemical restraints implicitly include the contributions of hydrogen atoms, the inclusion of hydrogen atoms does improve the accuracy of the structure factor calculation and allows calculation of specific hydrogen bonded and specific non-bonded anti-bump restraints. The inclusion of riding hydrogen atoms generally leads to small but significant improvements in model quality.

2.7.6 Jelly body restraints/Deformable elastic network (DEN):

It is equivalent to adding springs between atom pairs. The term added to the target function is

$$\sum_{pairs} w (|d| - |d_{current}|)^2$$

Summation is over all the pairs in the same chain and within a given distance. $d_{current}$ is recalculated at every cycle. This function does not contribute to the gradients. It only contributes to the second derivative matrix. If all interatomic distances were constrained, then individual atomic refinement would become rigid-body refinement. The effect of ‘jelly-body’ restraints is the implicit parameterization between the rigid body and individual atoms. This approach is implemented in the latest versions of CNS and REFMAC [29, 30].

2.7.7 Maximum likelihood target function:

Most of the present day refinement programs, such as CNS, PHENIX, REFMAC BUSTER/TNT [30-33] are based on maximum-likelihood principles. Maximum likelihood target functions account for incompleteness and errors in the model. Therefore, they are superior to basic least squares target functions, particularly in the early, error prone stages of the refinement. Maximum likelihood is an approximation of the general Bayesian inference procedure. Bayes' theorem states that the probability of a model given the data (ie, posterior probability) is proportional to the probability of data given the model (ie, likelihood function) multiplied by the prior probability of the model. For crystallographic refinement, prior probability comes from the standard stereochemical information and the likelihood comes from the X-ray diffraction experiment. Likelihood function, $\text{prob}(F_{\text{obs}} | F_{\text{calc}})$, allows the inclusion of model errors when estimating the expectation value of an observation in the form of F_{calc} from the structure model. The effect of incompleteness on the expectation value can be estimated by Sim weights 'm' derived from likelihood distributions under the assumption of an incomplete model, while positional errors of model atoms are accounted for by the Luzzati D -factor. Incorporation of bivariate errors (phase and amplitude) as complex Gaussian function and transformation into normalized structure factors shows that the Luzzati D -factor and the complex variance σ_{Δ}^2 are anticorrelated, and can be represented by the single parameter σ_A .

2.7.8 The optimization methods:

Various optimization algorithms can be used to achieve the best fit between parameterized model and all observations, which include experimental data and restraints. Optimization algorithms are roughly divided into analytical or deterministic procedures and stochastic procedures. Deterministic optimizations such as gradient based maximum likelihood methods are fast and work well when we are reasonably close to a correct model, at the price of becoming trapped in local minima. On the other hand stochastic procedures employ a random search that also allows movements away from local minima. They are slow but compensate for it with a large radius of convergence. Deterministic optimizations can be classified depending on how they evaluate the second derivative matrix. They generally descend in several steps or cycles from a starting parameter set (model) downhill towards a hopefully-but not necessarily- global minimum.

2.8 Validation:

Information not used for the refinement of model is very valuable for its validation. The R_{free} value [34], which is based on reflections not contributing to the refinement, is much more informative and provides a more objective quality criterion than the standard R value based on all reflections, which can be easily abused [35]. However, R_{free} is a global parameter. Similarly, since the torsion angles are usually not restrained during refinement, their agreement with expected values in the form of the Ramachandran plot [36] or clustering of side chain rotamers is extremely useful for the purpose of model validation. Several programs have been especially developed for checking the correctness

of protein models, of which the most popular and comprehensive are PROCHECK [36], WHAT_CHECK [37], SFCHECK [39] and MOLPROBITY [40]. In fact, many validation tools are built into all contemporary refinement and graphics display programs.

2.9 Site directed mutagenesis:

QuikChange Multi site-directed mutagenesis method was used to make all the mutations in the present work. This is a three step protocol in which Step 1 uses a thermal cycling procedure to achieve multiple rounds of mutant strand synthesis. Components of the thermal cycling reaction include a supercoiled double-stranded DNA template, two or more synthetic oligonucleotide primers containing the desired mutations, and the kit-provided enzyme blend featuring *PfuTurbo* DNA polymerase. First the mutagenic primers are annealed to denatured template DNA. *PfuTurbo* DNA polymerase then extends the mutagenic primers with high fidelity and without primer displacement, generating ds-DNA molecules with one strand bearing multiple mutations and containing nicks. The nicks are sealed by components in the enzyme blend. In Step 2 of the procedure, the thermal cycling reaction products are treated with the restriction endonuclease *Dpn* I. The *Dpn* I endonuclease (target sequence: 5'-Gm6ATC-3') is specific for methylated and hemimethylated DNA and is used to digest the parental DNA template. DNA isolated from almost all *Escherichia coli* strains is dam methylated and therefore susceptible to digestion. In Step 3, the reaction mixture, enriched for multiply mutated single stranded DNA, is transformed into XL10-Gold® ultracompetent cells, where the mutant closed circle ss-DNA is converted into duplex form in vivo. Double stranded plasmid DNA may then be prepared from the transformants and analyzed by appropriate methods to identify clones bearing each of the desired mutations.

2.10 Protease assay:

The chromogenic substrate of amino acid sequence His-Lys-Ala-Arg-Val-Leu*NPhe-Glu-Ala-Nle-Ser (where * denotes the cleavage site, and NPhe and Nle denote *p*-nitrophenylalanine and norleucine residues respectively) was synthesized and purified by Dr. Smita Mahale at the National Institute for Research in Reproductive Health, Parel, Mumbai, India. The spectrophotometric assay with this peptide was performed using a Jenway 6505 UV-Visible spectrophotometer. The reaction was carried out by incubating the varying concentrations of substrate (50-800 μM) with the enzyme (50nM) at room temperature in 50 mM sodium acetate buffer (pH 5.0). The decrease in absorbance at 300 nm resulting from hydrolysis of the scissile peptide bond was monitored. Values for k_{cat} and K_{m} of native tethered dimer of HIV-1 protease were obtained by fitting the curves to the Michaelis-Menten equation using Graphpad Prism software (Figure 2.1).

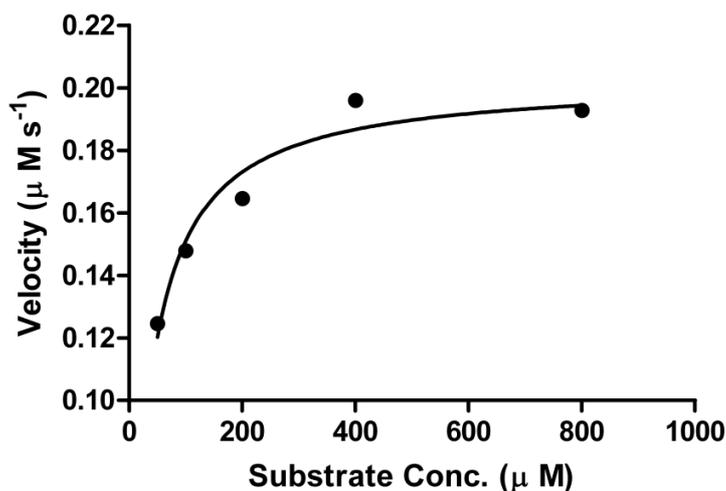


Figure 2.1: Velocity vs. Substrate concentration curve for native tethered dimer of HIV-1 protease.

Chapter 3

Catalytic water co-existing with a product peptide in the active site of HIV-1 protease revealed by X-ray structure analysis

3.1 Introduction:

Understanding the catalytic mechanism of HIV protease and the structure and interactions of the transition state would contribute significantly in the development of novel inhibitors. Based on computational [1-4], biochemical [5-7] and structural results [8-12], two types of proposals have been made in the past for the catalytic mechanism: direct and indirect [reviewed in 13,14]. In the direct type, championed mostly by computational studies, the nucleophilic attack on the carbonyl carbon atom of the scissile peptide bond is directly by carboxyl oxygen atom of the catalytic aspartates. In the indirect type, the attack is by a water molecule [15]. The position and hydrogen bonding patterns from this water molecule at the time of attack are different in different proposals of the catalytic mechanism, and therefore knowing the location and interactions of nucleophilic water molecule would be a step in establishing the correct mechanism for this enzyme.

HIV-1 protease is a homodimeric enzyme in which the active site is located at the subunit interface, with each subunit contributing one aspartic acid to the catalytic center. The active site is covered on the top by two flaps, which become ordered into a closed conformation whenever a substrate or inhibitor is bound in the active site. During virus maturation, HIV-1 protease cleaves viral polyproteins at nine different sites of varying amino acid sequences. A water molecule found symmetrically hydrogen bonded to

carboxyl oxygen atoms of both catalytic aspartates in the high resolution crystal structures of unliganded enzyme, (PDB Id 1LV1 and 2G69) is believed to be the nucleophile. This belief has been questioned [16] recently on the grounds that in the crystal structures of enzyme-ligand complexes, this water molecule has not been found to coexist with the ligand. Thus the location of nucleophilic water in the active site of HIV-1 protease is still an open question. In this respect, we have been pursuing crystallographic studies on active HIV-1 protease complexed with different substrate peptides [17-19]. We have been able to carry out such studies because of our discovery of closed-flap conformation of the enzyme in hexagonal crystals of HIV-1 protease even when the enzyme is unliganded [20,21]. Complexes with oligopeptide substrates could then be prepared by soaking these native crystals into aqueous solutions of the substrates. The chemical conditions, pH for example, of these solutions could be varied to try trapping the reactants at different stages of the reaction. In the present study, native crystals were soaked into solution of the substrate of amino acid sequence His-Lys-Ala-Arg-Val-Leu*-NPhe-Glu-Ala-Nle-Ser (where * denotes the cleavage site and NPhe & Nle denote p-nitrophenylalanine and norleucine, respectively) at pH 7.0. It was found that the full length substrate was cleaved at the specific cleavage site (Leu-p-nitro-Phe). The N-terminal product peptide (P product) had diffused out leaving behind only the C-terminal product peptide (Q product) still bound in the enzyme active site. A set of water molecules had moved into the region vacated by the P product peptide. One of these water molecules (WAT1) is optimally positioned to be the nucleophile. In this position, the water molecule does not accept any hydrogen bond through its lone pair and also is a donor in two strong hydrogen bonds, two features that contribute significantly towards

activation of the water molecule for nucleophilic attack [22]. This position is shifted by about 1.4 Å from that observed in all unliganded structures of HIV-1 protease. The position WAT1 overlaps exactly the hydroxyl group of the picomolar transition-state mimic inhibitor KNI-272. Adachi et al. have suggested this hydroxyl oxygen of KNI-272 to be an ideal position for a water molecule to launch nucleophilic attack on the scissile peptide bond [23]. Thus the present report of HIV-1 protease product complex is the first observation of putative catalytic water coexisting with the product peptide. This structure further suggests that the transition-state-mimics, such as KNI-272, should be so designed that they bind the catalytic aspartates with a hydrogen-bonding pattern similar to that of WAT1.

3.2 Materials and methods:

3.2.1 Protein expression and purification:

HIV-1 protease tethered dimer used in the present study contains five residue linker, Gly-Gly-Ser-Ser-Gly (GGSSG), covalently linking the two monomers [24] (Figure 3.1). The cloned insert also contains 57 extra codons in the beginning, which is a part of N-terminal polyprotein of pol gene. Therefore, the insert codes for a 29 kDa precursor protein, containing natural cleavage site for HIV-1 protease, which after self-cleavage results in a mature protein of 22 kDa.



Figure 3.1: HIV-1 protease precursor coded by tethered dimer insert

HIV-1 protease tethered dimer insert is cloned into plasmid expression vector pET11a, bearing ampicillin resistance and the T7 promoter. This plasmid contains unique NdeI and BamHI restriction sites at the 5' and 3' ends of the gene.

Expression, purification and crystallization of HIV-1 protease tethered dimer followed the procedures reported earlier [20-21]. Briefly, *Escherichia coli* (BL21 (DE3) strain) cells with mutant insert carrying plasmid were grown at 37 °C in Luria Broth (LB) medium containing 100µg/ml ampicillin. When the optical density of the culture medium measured at 600 nm (OD_{600}) reached about 0.6, the protease expression was induced by adding Isopropyl-β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 1 mM. Two hours after induction, cells were harvested by centrifuging at 6500 x g for 10 minutes. Cells were then resuspended in lysis/wash buffer (20mM Tris-HCl pH 7.5, 10mM ethylenediaminetetraacetic acid, 1% Triton-X-100) and ruptured using sonication. Inclusion bodies were collected by centrifuging the lysate at 10,000 x g for 10 minutes. Inclusion bodies were thoroughly washed 3-4 times with wash buffer and solubilised in 67% acetic acid. Extract was diluted to 2-5% acetic acid and dialyzed overnight against water. This was followed by dialysis against refolding buffer of pH 6.5, containing 20 mM piperazine-N, N'-bis(2-ethanesulfonic acid), 100 mM sodium chloride, 1 mM dithiothreitol, and 10% glycerol. Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS PAGE) result for various stages of protein purification is shown in Figure 3.2.

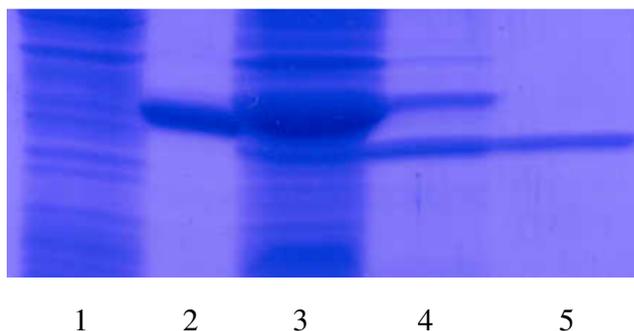


Figure 3.2: SDS PAGE at various stages of protein purification

Lane Nos: **1.**Bacterial cell extract before induction; **2.** 29 kDa Marker; **3.**Bacterial cell extract after induction; **4.**After dialysis to remove denaturant; **5.**After dialysis against Refolding buffer (only 22kDa band representing active protease was seen)

3.2.2 Crystallization and soaking:

Crystals were grown by hanging drop vapour diffusion method. Equal volumes of Protein (5 mg/ml in 50 mM sodium acetate, pH 4.5, containing 1 mM dithiothreitol) and reservoir solution (1% saturated ammonium sulfate, 200 mM sodium phosphate, and 100 mM sodium citrate at pH 6.2) were mixed on a cover slip and sealed over the reservoir well at room temperature.

The 11-residue substrate peptide of amino acid sequence His-Lys-Ala-Arg-Val-Leu-NPhe-Glu-Ala-Nle-Ser was synthesized by Dr. Smita Mahale at the National Institute for Research in Reproductive Health, Parel, Mumbai, using an automatic peptide synthesizer. The peptide was dissolved in water to prepare a 5 mM stock solution. This stock solution was diluted 5-fold into the reservoir solution (pH 7.0) to prepare the soaking drop. Protease crystal was transferred first to a fresh reservoir solution (pH 7.0) drop to wash the crystal and then to the soaking drop using a cryoloop. The cover slip was inverted and sealed over the same reservoir well in which crystals had been grown.

3.2.3 X-ray data collection and refinement:

At the end of 72 h of soaking at room temperature, the crystal was equilibrated in the cryo-protectant (25% glycerol and 75% reservoir buffer) before flash freezing, for exposure to X-rays on the FIP-BM30A beam line [25]. The crystals diffracted to 1.69Å resolution. The diffraction data were indexed, integrated, and scaled by using the computer program XDS [26].

Computer program PHASER [27-28] from CCP4 suite was used to obtain molecular replacement solution using the structure 1LV1 [21, 29] as the search model. The structure was refined in Crystallography and NMR System (CNS) by using standard simulated annealing protocols and the amplitude-based maximum likelihood target function [30,31]. A test set containing 5.0% of randomly chosen reflections were reserved for determination of R_{free} [32], which is an indicator of gainful refinement. Occupancies of ligand molecules in the two orientations were systematically varied, in steps of 0.1, subject to the constraint that their sum be 1.0. Electron density maps of all types were calculated using CNS. All interactive model building and molecular superpositions were carried out using the graphics software O [33]. Structural comparisons are based on superpositions of protein C α atoms. All figures were drawn using program PYMOL [34]. Atomic co-ordinates and structure factors have been deposited in the Protein Data Bank under the PDB Id 2WHH.

3.3 Results:

3.3.1 The Model of the complex:

HIV-1 protease tethered dimer used here contains a five residue linker, GGSSG, linking the N-terminus of second monomer to C-terminus of the first monomer [24]. Residues in the first monomer are numbered as 1-99 and those in the second monomer are numbered 1001-1099. Residues of the linker are numbered as 101-105. Crystal and intensity data statistics are given in Table 3.1.

Table 3.1: Data collection and refinement statistics

Space group	P6 ₁
Unit cell parameters (Å)	a= b = 62.54,c= 81.99
Wavelength used (Å)	0.97945
Resolution (Å)	50.0-1.69 (1.72 – 1.69)*
Number of unique reflections	20251 (922) *
I/σ(I)	12.61 (3.35) *
R_{merge} (%)	8.2 (49.0) *
Completeness (%)	97.9 (94.1) *
<u>Refinement statistics</u>	
R _{work} / R _{free} (%)	21.9 (30.2) * / 25.6 (31.5) *
<u>RMS deviations from ideal values</u>	
Bond lengths (Å)	0.01
Bond angles (°)	1.5

*Data for highest resolution shell are given in the parenthesis.

$$R_{merge} = \frac{\sum_{hkl} \sum_j |I_{hkl,j} - \langle I_{hkl} \rangle|}{\sum_{hkl} \sum_j I_{hkl,j}}$$

On refinement of the protein structure, difference density was found in the active site region of the enzyme (Figure 3.3), and this difference density represented the soaked-in substrate cleaved at the linkage connecting Leu and Nphe residues in the sequence. As per convention, residues in the C-terminal product (Q product) counted from the scissile bond were designated as P1'-P5', and those in the N-terminal product (P product) as P1 – P6. The density for residues P1 – P6 was very weak suggesting that the P product peptide had diffused out leaving behind only the Q product peptide still bound in the enzyme active site. A set of water molecules had substituted the P product peptide. Electron density for residues beyond P2' in the Q product was also very weak. The Q product and the water molecules were placed in two orientations, consistent with the pseudo-symmetry of HIV-1 protease active site. The lowest R_{free} was obtained when the occupancies for the two orientations were 0.7 and 0.3. The B-factor averaged over all atoms of the product peptide was 43.7 \AA^2 and 42.2 \AA^2 respectively for the two orientations. The electron density suggested that the side chains of few protein residues existed in multiple conformations in the crystal. Alternate conformations were modeled for the residues Val 82, Ile 84, Val 1082 and Ile 1084. There was no visible density in the 2Fo-Fc map for the linker region between residues 99 and 1001 of the tethered dimer under study, suggesting that the linker region was not ordered in the crystal. The final molecular model thus consisted of 1514 protein atoms, 181 water molecules and Q product peptide bound in two orientations with occupancies of 70% and 30% respectively. Conformationally, more than 90% of non-glycine residues were in the most favored regions of Ramachandran plot. The final refined 2Fo-Fc map for P1' p-nitro-

phenylalanine and P2' glutamic acid residues and the active site water molecules in the two orientations are shown in Figure 3.4.

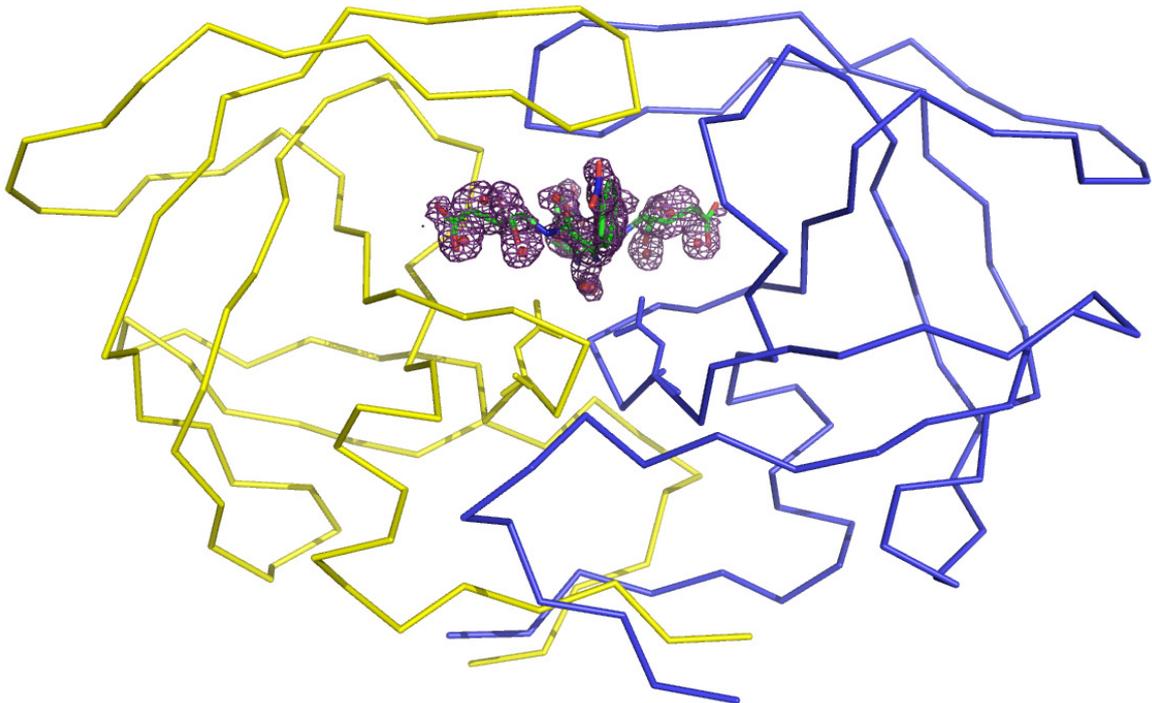


Figure 3.3: F_O-F_C difference density map contoured at 2.0σ level in the active site of HIV-1 protease/product complex. The enzyme is shown as a Ca trace. The side chains of catalytic aspartates are also drawn.

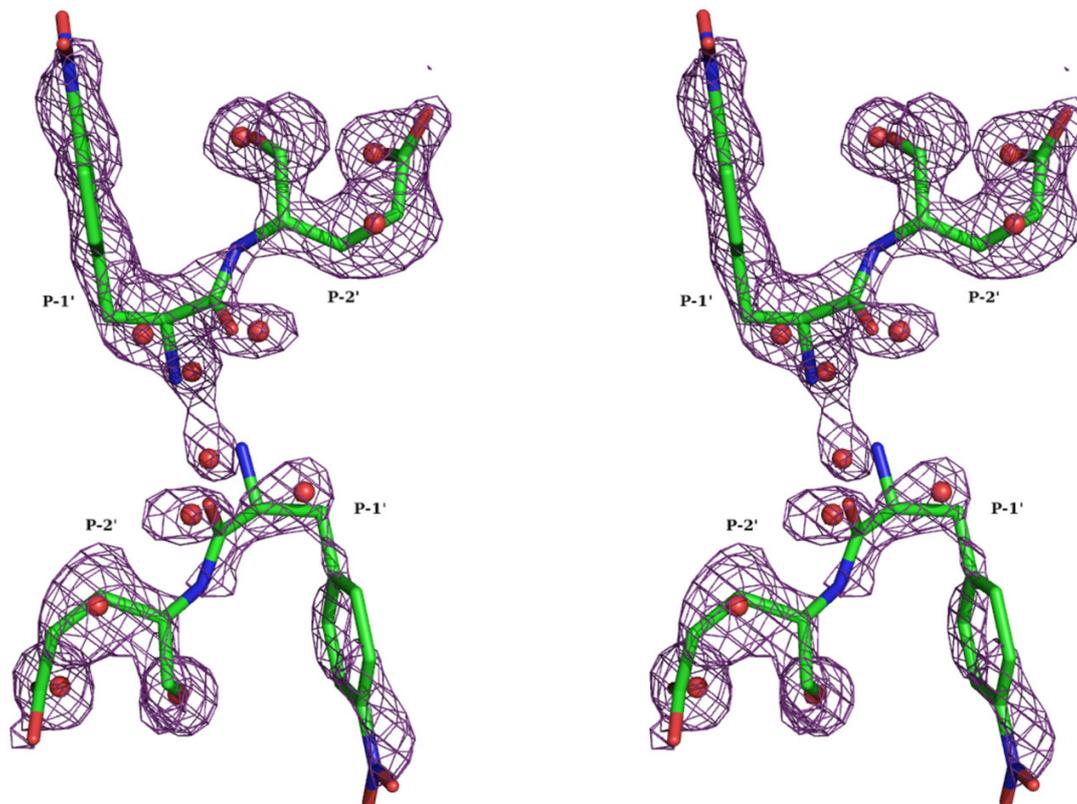


Figure 3.4: Fit of carboxyl terminal (Q) product peptide and active site water molecules into 2Fo-Fc electron density. The electron density map is contoured at 1.0σ level. The carboxyl product peptide (violet purple) and the active site water molecules are shown in the two orientations.

3.3.2 Protease- Q Product peptide interactions:

Hydrogen bonding interactions between P1', P2' residues of the Q product and the protein residues in the active site are shown in Figure 3.5a. The Q product is held in the active site through 11 hydrogen bonds, some of which are through bridging water molecules. Terminal nitrogen of product peptide in both the orientations forms hydrogen bond to the outer oxygen (OD2) of ASP- 1025/ASP- 25 (Figure 3.5a). The side chain of P2' GLU forms hydrogen bonds with main chain amide nitrogen and side chain carboxyl oxygen of ASP- 30 or ASP- 1030 depending upon the orientation. One very well ordered

water molecule forms the bridge between product peptide and the amide group of Ile 50/Ile 1050. One of the oxygens of P1' nitro group forms hydrogen bond with the Arg 8 while the other oxygen is bridged by two water molecules to the carbonyl oxygen of Gly 49.

3.3.3 Water molecules in the active site:

A set of water molecules had substituted the P product peptide. These water molecules are held in place through hydrogen bonds among themselves and also with the protein (Figure 3.5b). One of these water molecules, WAT1, which is within hydrogen bonding distance from the oxygens of both catalytic aspartates, may be of functional importance. The OMIT density for this water molecule is shown in Figure 3.6. WAT1 also makes a short hydrogen bond with the N-atom of the Q product peptide. WAT1 is shifted by about 1.4 Å from the corresponding water molecule coordinating both catalytic aspartates in the unliganded structures (PDB Id 1LV1 and 2G69). This water molecule is at an average distance of about 2.7 Å from the scissile carbon of the modeled substrate peptides (Figure 3.6).

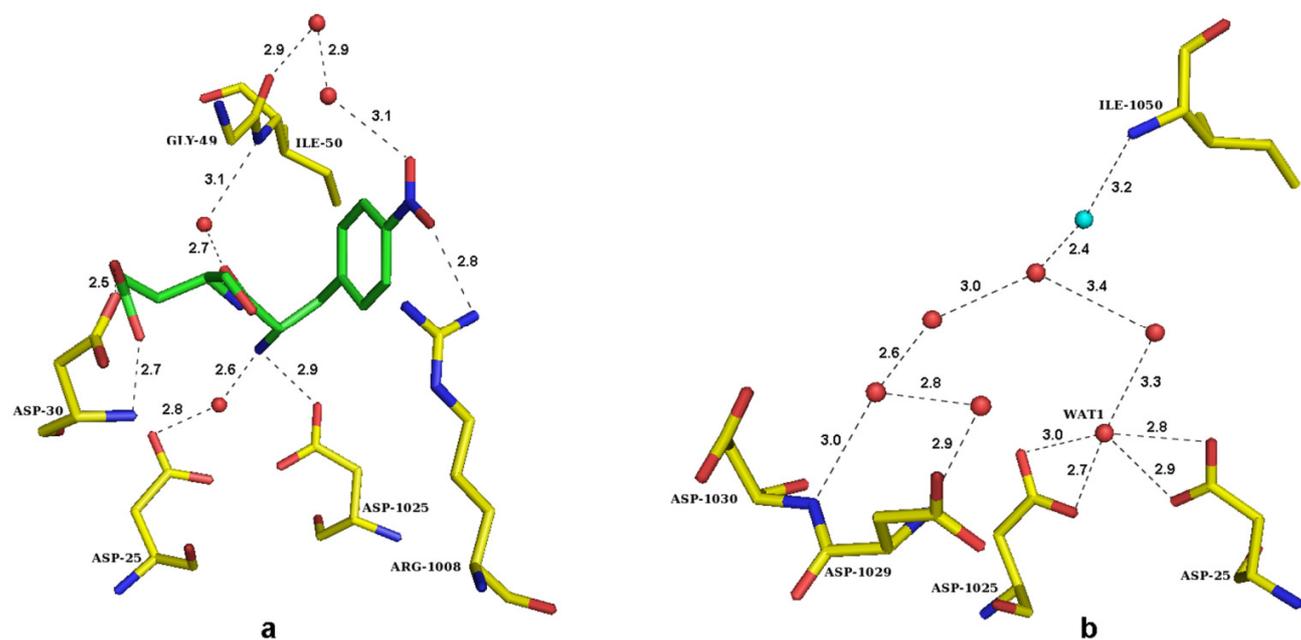


Figure 3.5: Hydrogen bonding interactions at the active site: a) between the protease and the P1'-P2' residues of the Q product peptide, b) between the protease and the water molecules. Possible hydrogen bonds are drawn as dotted lines, and the lengths are indicated.

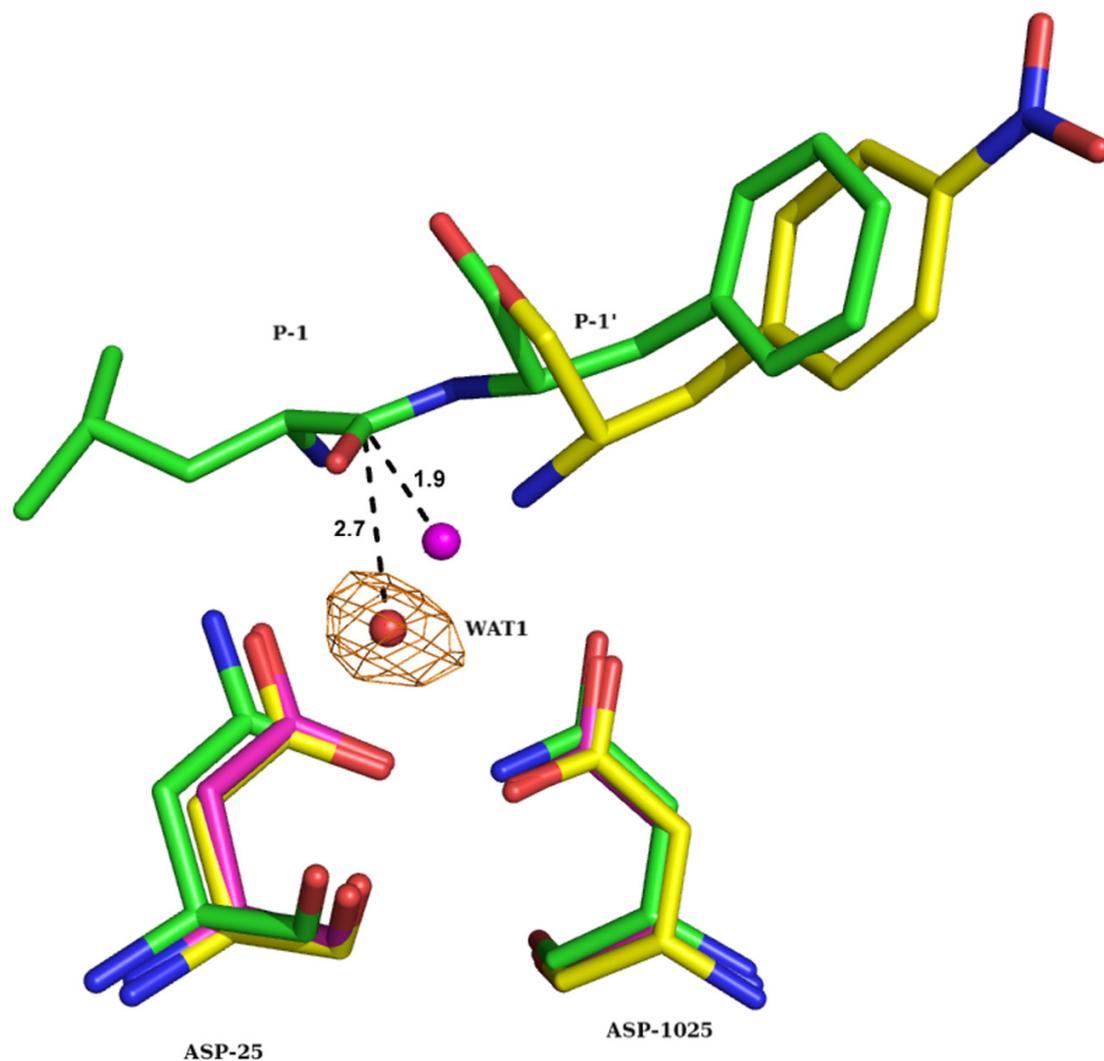


Figure 3.6: Relative positions of WAT1 and the modelled substrate in the active site: Diagram showing superposition of three structures: 1) present structure (yellow carbon), 2) unliganded HIV-1 protease (magenta carbon, PDB Id 1LV1) and 3) inactive HIV-1 protease/substrate complex (green carbon, PDB Id 1KJH). Water molecule observed in unliganded HIV-1 protease is also shown (magenta). The distances to the scissile carbon atom are indicated. SA OMIT density contoured at 3.0σ level is also shown for WAT1.

3.4 Discussion:

3.4.1 Position of attacking water molecule:

In the hydrolysis reaction catalyzed by HIV-1 protease there are two substrates: 1) an oligopeptide of appropriate amino acid sequence and 2) the nucleophilic water molecule. At the start of the reaction, both these are bound in the active site leading to formation of Michaelis complex. At the end of the reaction, but before product release, the nucleophilic water is used up and hence should not be present in the active site. The presence of WAT1 in the active site places the present structure somewhere near the beginning steps of the reaction. Presently there is no crystal structure report of a Michaelis complex between active HIV-1 protease and substrate peptide. However, the present structure can be considered a close approximation to Michaelis complex since a part of the substrate peptide is present in the active site along with the water molecule. We have earlier reported the structures of HIV-1 protease complexed with two different substrate oligopeptides corresponding in amino acid sequence to the junctions CA-p2 [17] and RT-RH [18] in the polyprotein substrate. While the substrate is converted into a tetrahedral intermediate in the complex with CA-p2, the RT-RH peptide is cleaved, with both product peptides still bound in the active site. The water molecule, WAT1, is at a distance of 0.9 Å from one gem-diol hydroxyl in the tetrahedral intermediate complex. Similarly WAT1 is at a distance of about 1.0 Å from one of the carboxyl oxygens in the product peptide complex (PDB Id 2NPH) (Figure 3.7). Because of these proximities, we suggest that the water molecule serving as the nucleophile in peptide bond hydrolysis does so from the position WAT1 observed in the present structure. Such a hypothesis

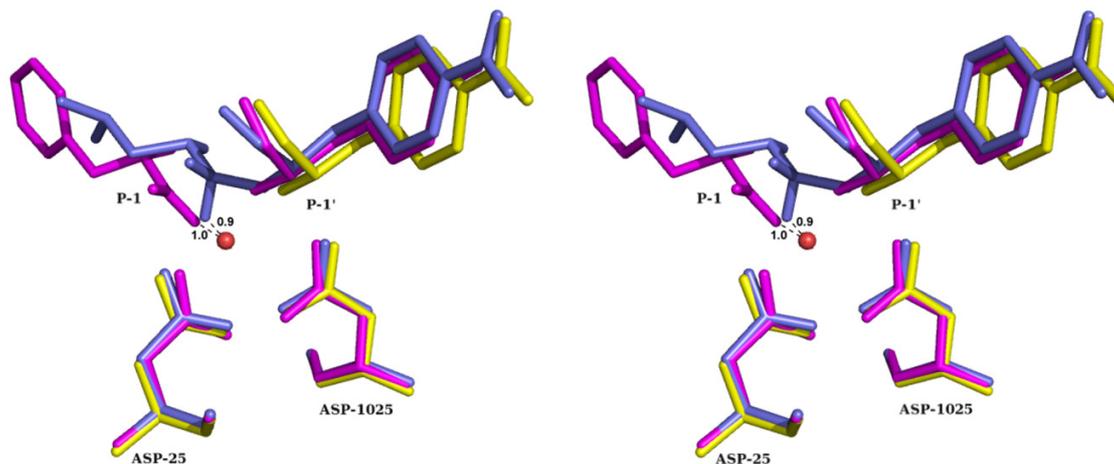


Figure 3.7: Structural comparison of present complex with tetrahedral intermediate complex [17] and product peptide complex [18]: Stereo diagram showing the ligand atoms at the catalytic centre along with catalytic aspartates. Protein C α atoms are used in the structural superposition. WAT1 is within 1 Å from an oxygen atom in the newly generated gem-diol [17] or carboxyl group [18].

would be consistent with the principle of least nuclear motion for chemical reactions [35]. To further explore this idea, we have investigated by molecular modeling, if the scissile peptide bond of a substrate bound in the active site would be accessible to WAT1 for attack. We have superposed separately the present complex with reported complexes between D25N inactive enzyme and two different substrate oligopeptides (PDB Id 1KJH and 3BXR) [36, 37]. Using only protein C α atoms for structural superposition, equivalent positions of the substrate molecules were derived. Figure 3.6 shows the derived positions relative to WAT1. It is clear that the scissile peptide bond is optimally accessible to WAT1 for nucleophilic attack, the WAT1...C-O and WAT1...C-N angles being 69° and 104° respectively. Further, the distance of WAT1 to the scissile carbon atom is 2.7 Å, which is reasonable for a nucleophilic attack. Figure 3.6 also shows the position of the

catalytic water observed in the structure of unliganded HIV-1 protease (PDB Id 1LV1). The separation of this water molecule from the scissile carbon atom is only 1.9 Å, which is too short a distance for the water molecule to stay in this position along with the substrate. Unlike in unliganded structures, the position of WAT1 is asymmetric with respect to the catalytic aspartates. WAT1 forms two short hydrogen bonds to outer and inner carboxyl oxygens of ASP- 25 and ASP-1025 respectively. Further, WAT1 does not accept any hydrogen bond and is a donor (see below) in two strong hydrogen bonds with catalytic aspartates. Both these features should increase the nucleophilicity of WAT1 [22]. From all these considerations, WAT1 appears to be a reasonable position for the water molecule from where nucleophilic attack takes place during bond breakage. This hypothesis is also consistent with the structure of HIV-1 protease /KNI-272 complex reported recently [23]. KNI-272 is one of very few highly selective and potent inhibitors of HIV-1 protease with a picomolar inhibitory constant. The high potency is suggested to be due to its pre-organized rigid structure that very closely resembles the transition state. The structure of the complex has been determined to very high resolution using X-ray and neutron diffraction techniques. According to the authors of this study, the position of the hydroxyl group in the hydroxymethylcarbonyl part of KNI-272 is ideal to mimic the location of the attacking water molecule in catalysis. Figure 3.8 shows the superposition of the present structure with the HIV-1 protease/KNI-272 complex mentioned above (PDB Id 3FX5). It is very interesting that WAT1 perfectly overlaps the hydroxyl group of KNI-272 in the complex. Since this overlap guarantees the adherence to the principle of least nuclear motion, KNI-272 is a very potent inhibitor of HIV-1 protease. In addition to hydrogen bonds to catalytic aspartates, WAT1 is hydrogen bonded to the terminal N atom

of the Q product peptide. Once the Q-product leaves the active site, WAT1 will move back to the position observed in the structures of unliganded HIV1-protease. The relative positions of the substrate, nucleophile and catalytic aspartates at different stages of the cleavage reaction, according to our proposal, are shown in Figure 3.9 (a-f).

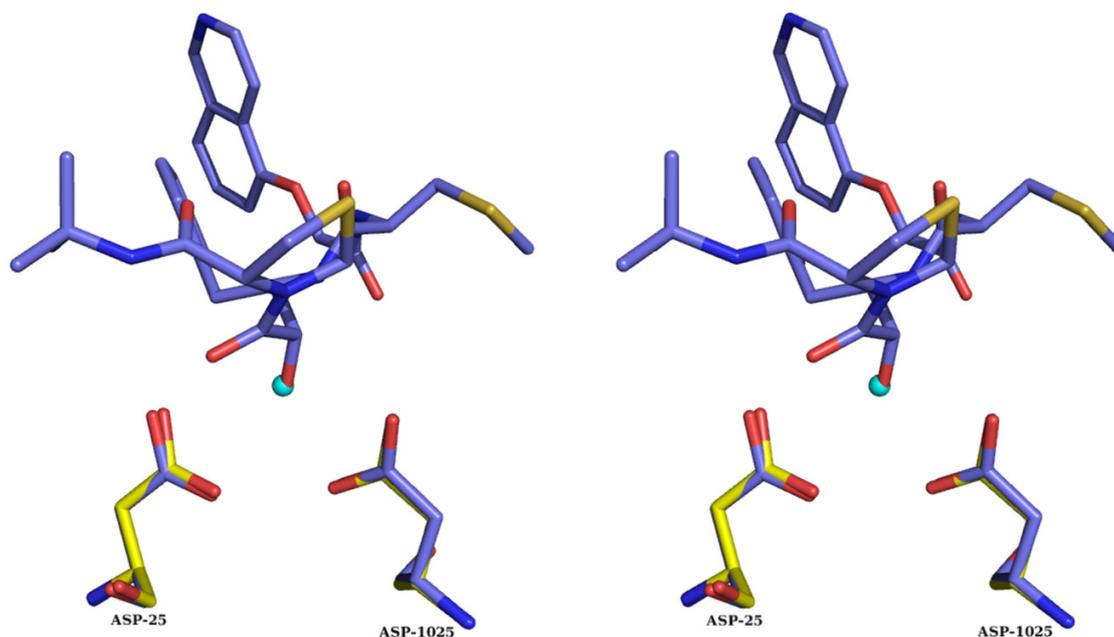


Figure 3.8: Position of the nucleophilic water molecule. Stereo diagram showing the overlap between the hydroxyl group (red) of transition-state mimetic inhibitor KNI-272 and WAT1 (cyan) of the present structure. The Q product peptide of the present structure is not shown. Note the perfect overlap of WAT1 and the hydroxyl oxygen of KNI-272.

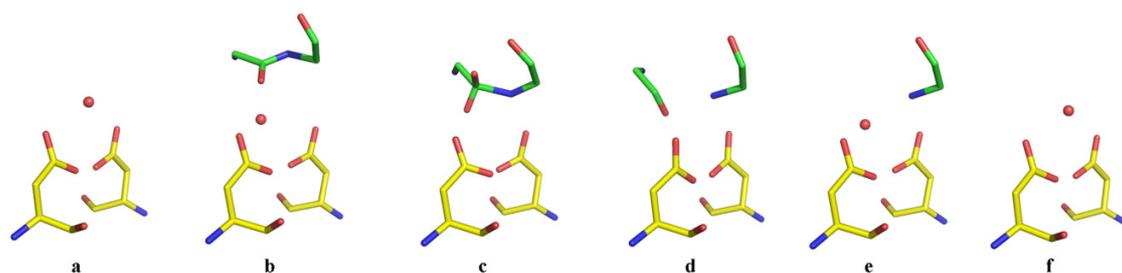


Figure 3.9: Proposed sequence of steps (a-f) in the cleavage of peptide bond by HIV-1 protease. Only main chain atoms of substrate peptide (green carbon) are shown. Each figure is based on the structure indicated: (a) catalytic water molecule bound symmetrically to the two aspartates (structure PDB Id 1LV1), (b) movement of catalytic water to position WAT1 with binding of substrate (modelled), (c) attack by WAT1 and formation of tetrahedral intermediate (Structure TI), (d) cleavage of the peptide bond with product peptides still bound in the active site (structure PDB Id 2NPH), (e) diffusion of P product out of and WAT1 into active site (present structure), and (f) release of Q product and movement of WAT1 into original position (PDB Id 1LV1).

3.4.2 Protonation state of catalytic aspartates:

In the process of inhibitor design, it is important to both structurally mimic the transition state intermediate and to maximize interactions between the inhibitor and the catalytic aspartates. In this context, it is essential to know the protonation states of the catalytic aspartates so that appropriate functional groups are chosen in the inhibitor being designed. Even though hydrogen atoms are not located in the present study, the observed strong hydrogen bonds involving the catalytic aspartates provide a clue to the protonation states of the aspartates. Only O...O/N separation shorter than 2.8/2.9 Å are considered as definite hydrogen bonds [38]. There are four such distances at the catalytic centre in the present structure: i) ASP-25 OD1...ASP-1025 OD1, ii) WAT1... ASP-25 OD1, iii) WAT1...ASP-1025 OD2 and iv) ASP-1025 OD2...N-terminus of P1' residue. The angle ASP-25 OD2 –WAT1- ASP-1025 OD1 is 101° which is very close to the H-O-H angle

(104°) in a water molecule indicating that in the hydrogen bonding to the water molecule WAT1, the aspartate oxygens act as acceptors. Since the substrate is already cleaved, N-terminus of P1' residue is already protonated and it would be a donor in the hydrogen bond with ASP-1025 OD2 atom. Thus the aspartic dyad is monoprotinated with the proton shared between inner oxygens of the two aspartates. We therefore suggest that just prior to the formation of the transition state the aspartates are in this state of protonation. Since on inhibitor binding the protonation state is not likely to change, the hydrogen-bonding group should be chosen appropriately on the inhibitor to maximize interactions with aspartates in this state of protonation.

3.4.3 Effect of pH on conformation and interactions of ASP-25 and ASP-1025:

HIV protease is known to be active over a wide range of pHs. In our earlier study of the crystal structure of HIV-1 protease complexed with the undecapeptide substrate (His-Lys-Ala-Arg-Val-Leu*-NPhe-Glu-Ala-Nle-Ser) at a pH value of 2.0, the substrate bound in the active site had transformed into a tetrahedral intermediate through nucleophilic attack by a water molecule [17]. In contrast, in the present study carried out at pH 7.0, the substrate molecule of the same sequence is found cleaved at the correct scissile bond, and the N-terminal P product peptide has diffused out of the enzyme active site. The conformations and interactions of the catalytic aspartates at the two pHs are compared in Tables 3.2 and 3.3. The changes in the conformations around main-chain and side-chain torsions of the two aspartates are very small, but these small changes have synergistically

Table 3.2: Comparison of conformation of catalytic aspartates in the structures at pH 2.0 and pH 7.0

Parameter compared	pH	χ_1 ($^\circ$)	χ_2 ($^\circ$)	Φ ($^\circ$)	Ψ ($^\circ$)
Conformation of Residue ASP-25	2.0	-173	-25	-126	-85
	7.0	-171	-14	-120	-86
Conformation of Residue ASP-1025	2.0	-175	166	-116	-84
	7.0	-173	173	-121	-83

Table 3.3: Comparison of interactions of catalytic aspartates in the structures at pH 2.0 and pH 7.0

Parameter compared	pH 2.0	pH 7.0
Interaction distance ASP-25 OD1- GLY-27 N	2.8 Å	2.9 Å
Interaction distance ASP-1025 OD1- GLY-1027 N	2.7 Å	2.9 Å
Interaction distance ASP-25 OD1- ASP-1025 OD1	3.0 Å	2.7 Å
Virtual Dihedral angle OD2(25)-OD1(25)-OD1(1025)-OD2(1025)	78 $^\circ$	56 $^\circ$

caused differences in the interaction distances, which could be significant. The hydrogen bonds from inner oxygen (OD1) atoms to N atom of corresponding Gly-27/1027 residues have become longer for both aspartates at pH 7.0. The distance between the two inner oxygen atoms, on the other hand, has changed in the opposite direction, that is, to shorter value, at pH 7.0. If the length of a hydrogen bond is assumed to reflect its strength, the changes in lengths mentioned above appear to preserve the total hydrogen-bonding ability of each OD1 atom. There is a significant change in the virtual dihedral angle OD2 (25)-OD1 (25)-OD1 (1025)-OD2 (1025), which is a measure of the co-planarity of the two aspartic acid side chains [11]. The two side chains tend toward being more co-planar at pH 7.0. There also appears to be a correlation between the co-planarity of the two aspartates and the strength of the hydrogen bond between the OD1 atoms of catalytic aspartates; the higher co-planarity leading to stronger hydrogen bond. In the structure of HIV-1 protease product complex [18] determined at a pH of 6.2 the aspartates are more co-planar with a virtual dihedral angle of 22° while the OD1 distance of the postulated hydrogen bond is only 2.3 Å.

3.4.4 Product release:

The patterns of product inhibition are dependent on the enzyme mechanism. Based on product inhibition and solvent isotope effects, in the cleavage reaction by HIV-1 protease the product peptides are proposed to be released in an ordered manner, with the P product peptide released first [7]. The presence of only carboxyl terminal product in the present structure is consistent with this expectation. The Q product peptide is tending to diffuse out of the active site, although more slowly, since the distance between C α atom of P1' residue and C γ atom of distal aspartate has increased from 5.0 Å in the tetrahedral

intermediate structure [17] to 5.5 Å in the present structure (Figure 3.7). It is interesting that it is the N-terminal P product which is bound when active HIV-1 protease is cocrystallised with a constrained hexapeptide [37]. This difference may be due to the different approach taken for preparing crystalline enzyme/substrate complex. The constrained hexapeptide is cleaved during cocrystallisation, and from among the two products released into solution the P product is selectively bound in the active site because of its increased hydrogen bonding ability coming from the newly formed carboxyl group. Similarly, on cocrystallisation, the presence of an amino group in the product peptide PIV-CONH₂ resulted in binding of PIV-CONH₂ in the active site of HIV-1 protease, in an unexpected mode [9, 37].

Chapter 4

Insights into the mechanism of drug resistance: X-ray structure analysis of G48V/C95F tethered HIV-1 protease dimer/saquinavir complex

4.1 Introduction:

Acquired Immunodeficiency Syndrome (AIDS) is the major pandemic of recent times affecting most of the continents [1], and is caused by Human Immunodeficiency Virus (HIV). Since its discovery in the 1980's [2, 3], elucidation of the molecular biology of the viral life cycle has resulted in the identification of different viral drug targets [4]. HIV-1 protease is one such major target enzyme because of the crucial role it plays in the processing of the viral polyproteins [5]. Until now, 9 inhibitors of HIV-1 protease have been approved by the US Food and Drug Administration (FDA) for use as drugs in the treatment of infected patients. However, the clinical efficacy of these drugs has been limited by the emergence of drug resistant variants of HIV in response to the selection pressure of HIV-1 protease inhibitors (PI's). Mutations of flap residues 46, 47, 48, 50, 53, and 54 are frequently observed in drug-resistant mutants of HIV-1 protease, and these mutants show various levels of reduced susceptibility to different drugs. The drug saquinavir (Ro31-8959) (SQV) was designed to target the wild-type protease, and its chemical structure contains a number of peptidic main chain groups mimicking a natural substrate. It was the first HIV-1 protease inhibitor to be released in the market. It is a sub-nanomolar inhibitor of HIV-1 protease with dissociation constant (K_i) of 0.12 nM [6].

However, drug-resistance against SQV quickly emerges, with mutations occurring independently at positions 48, and 90. The mutation G48V is present in the flap region, and is considered as the major active site mutation against SQV. Generally, single mutations confer only a modest reduction in drug susceptibility and a stepwise accumulation of several mutations is required for the development of high-level drug resistance. Very often secondary mutations also occur to enhance catalytic activity of resistant variants. From a statistical analysis of sequences of protease enzyme isolated from AIDS patients, both treated and untreated with PI's, it has been found that the mutation C95F occurs as a mutation cluster either with L90M and I93L or with G48V and V82A [7]. Further, the mutation C95F was found only in patients treated with SQV and was non-existent in patients not administered any protease inhibitor. Thus C95F is the major non-polymorphic mutation associated with the major mutation G48V against SQV [8]. Detailed three dimensional structures of these mutant proteins and of their complexes with SQV would be helpful in understanding the mechanism of drug-resistance for design of next generation drugs, and also in understanding the phenomenon of 'mutation-clustering'. However, the only structure experimentally determined so far is that of the complex between SQV and the double mutant G48V/L90M [9]. The bottleneck has been the inability to grow single crystals of the other mutants or of mutant/SQV complexes. Hybrid quantum / molecular mechanics studies of structural effects of the G48V mutation toward SQV binding have been published [10,11].

Crystal structure of G48V/C95F tethered HIV-1 protease, in complex with SQV is reported here. This structure was then compared with the wild-type complex (PDB Id 1HXB) and also with G48V/L90M complex (PDB Id 1FB7) to reveal effects of the

mutations on protein structure and inhibitor binding. Loss of direct hydrophobic interactions between SQV and enzyme residues and loss of water-mediated hydrogen bonds between drug and flaps are observed in G48V/C95F complex. Analyses of the CD spectra recorded on wild-type and mutant enzyme show that the stability of the mutant enzyme is reduced, and this loss of stability also could contribute toward drug resistance. The three dimensional structure also provides a rationale for how the mutation V82A in the cluster C95F, G48V and V82A can bring-in additional resistance against SQV.

4.2 Materials and methods:

4.2.1 Site directed mutagenesis:

Site directed mutagenesis was carried out using QuikChangeTM Multi Site-Directed Mutagenesis kit (Stratagene, La Jolla, CA). The oligonucleotide primers were designed with the aid of the Primer X software package. Oligonucleotide primers 5'-TGAAACCACCAATACCAACGATCATCTTCGGCTTC-3' and 5'-CCCGCCAAA GTTCAGAGTAAAACCGATCTGAGTCAGCAGG-3' were used for making G48V and C95F mutations, respectively. Site-directed mutagenesis was performed according to the manufacturer's instructions. Briefly, double stranded pET 11a plasmid DNA with protease insert was used as the template and PCR reactions were performed in a total volume of 25 μ l. The PCR product was generated through thirty amplification cycles of one minute at 95 °C to denature the DNA, one minute at 55 °C to anneal the primers and 13 minutes at 65 °C for DNA extension. The reaction contained 10 X reaction buffer, 100 ng of double stranded DNA template, 100 ng of each oligonucleotide primers and 10 mM dNTP mix. QuikChangeTM Multi enzyme blend (2.5 U/ μ l) was then added. The PCR

mixture was treated using *DpnI* (10 U/ μ l) to digest the methylated DNA template. The plasmid was recircularized at room temperature and used to transform *Escherichia coli* XL1-gold supercompetent cells supplied with the kit. Cells were plated onto Luria agar plates supplemented with ampicillin (100 μ g/ml). Putative mutant plasmid DNA from picked colonies was prepared and sent for sequencing to ensure that no other mutations were incorporated in the cDNA during the thermal cycling reaction. Presence of mutations was confirmed through DNA sequencing.

4.2.2 Protein expression and purification:

HIV-1 protease tethered dimer used in the present study contains a five residue linker, GGSSG, covalently linking the two monomers [12]. Expression, and purification of HIV-1 protease tethered dimer followed the procedures reported earlier [13] and explained in chapter 3 of this thesis.

4.2.3 Crystallization:

Single crystals were obtained by the hanging drop vapour diffusion method. Protein (5 mg/ml in 50mM sodium acetate, pH 4.5, containing 1 mM dithiothreitol) was reacted at room temperature for 30 minutes with ten-fold molar excess of SQV dissolved in dimethyl sulfoxide. For crystallization, equal volumes of reaction mixture and reservoir solution (1% saturated ammonium sulfate, 200 mM sodium phosphate, and 100 mM sodium citrate at pH 6.2) were mixed on a cover slip and sealed over the reservoir well at room temperature.

4.2.4 X-ray data collection and refinement:

The crystals were equilibrated in the cryo-protectant (25% glycerol and 75% reservoir buffer) before flash freezing, for exposure to X-rays on the Swiss Light Source beam line, X06DA. The diffraction data were indexed, integrated, and scaled by using the computer programmes MOSFLM and SCALA [14,15]. Crystal and intensity data statistics are given in Table 4.1.

Crystal structure was solved by difference Fourier method using native coordinates (PDB Id 1LV1). The structure was refined in computer programme PHENIX by using standard simulated annealing protocols and the amplitude-based maximum likelihood target function [16]. A test set containing 4.57% of randomly chosen reflections were reserved for determination of R_{free} [17]. All interactive model building and molecular superpositions were carried out using the molecular modeling software O [18]. Structural comparisons were based on superpositions of protein $C\alpha$ atoms. The quality of model was checked using the software PROCHECK [19]. All figures were drawn using computer programme PYMOL [20]. The atomic coordinates and structure factors have been deposited in the Protein Data Bank (www.rcsb.org) under the PDB Id 3N3I.

4.2.5 Protease assay:

Assay with chromogenic substrate of amino acid sequence His-Lys-Ala-Arg-Val-Leu*NPhe-Glu-Ala-Nle-Ser (where * denotes the cleavage site, and NPhe and Nle denote *p*-nitrophenylalanine and norleucine residues respectively) was carried out as explained in chapter 2 of this thesis. The decrease in absorbance at 300 nm resulting from hydrolysis of the scissile peptide bond was monitored. Values for k_{cat} and K_{m} were

obtained by fitting the curves to the Michaelis-Menten equation using Graphad Prism software (Figure 4.1).

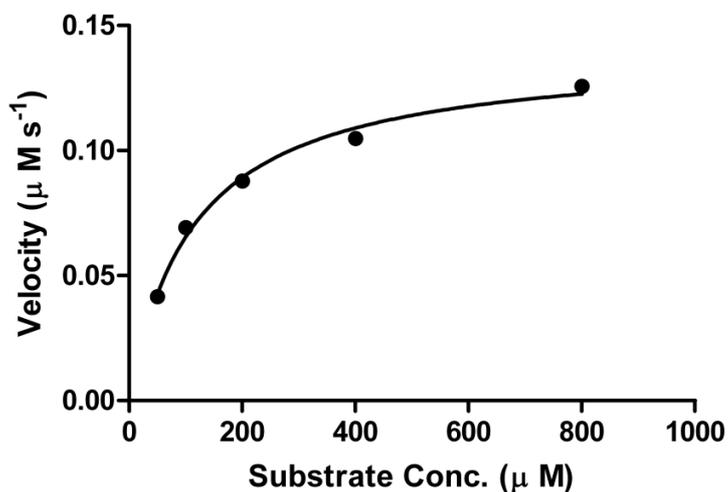


Figure 4.1: Determination of kinetic constants: Velocity vs. Substrate concentration curve for G48V/C95F mutant.

4.2.6 Circular dichroism spectroscopy:

Near and far UV CD spectra on wild type and mutant proteins, both in the absence and presence of molar excess of SQV, were recorded using a Jasco J-815 CD spectrometer. Protein solutions were used at a concentration of 10 μ M in 50mM sodium acetate buffer, pH 4.5. For studying the protein-drug complex, 7 μ M of protein solution was mixed with 10 μ M of SQV. The mixture was incubated for one hour before the start of unfolding experiment. Thermal denaturation experiments were carried out by monitoring the CD signal at 215 nm over a temperature range of 20 $^{\circ}$ C to 110 $^{\circ}$ C with a scan rate of 2 $^{\circ}$ C/min. Temperatures of midpoint unfolding transitions (T_m) were estimated by

evaluating the maximum of the first derivative of $[\Theta]_{215}$ in relation to the temperature data [21].

4.3 Results and discussion:

4.3.1 Kinetic constants:

G48V/C95F mutant exhibited a K_m and k_{cat} of $64.7 \pm 4.1 \mu\text{M}$ and $3.1 \pm 0.1\text{s}^{-1}$, respectively. The catalytic efficiency (k_{cat}/K_m), of G48V/C95F mutant, for hydrolysis of the chromophoric substrate was about 17% as compared to wild type protease (Table 4.2).

4.3.2 Crystallization:

All our attempts to crystallize G48V/C95F HIV-1 protease de novo were unsuccessful. However, single crystals of G48V/C95F / SQV complex appeared under conditions similar to those used in our laboratory for growing crystals of unliganded wild-type HIV-1 protease. These crystals, which took comparatively longer to appear, belong to the same hexagonal space group as the crystals of the wild-type. Liu et al also have reported that despite extensive efforts, no crystals could be obtained for G48V single mutant in complex with SQV [22], while orthorhombic crystals were obtained when complexed with the drug molecule darunavir. Since the residue 48 is exposed on the surface of the molecule and is also involved in crystal contacts in hexagonal crystals, we believe that the mutation G48V affects the crystal packing leading to difficulty in crystallization.

4.3.3 Quality of the model:

Results of crystallographic refinement are summarized in Table 4.1. The residues in the first monomer are numbered as 1-99 and those in the second monomer are numbered

1001-1099. When 2mFo-DFc electron density map was contoured at 1 σ level, no breaks were observed in the main chain. There was no density, however, for the linker region of the tethered dimer under study, suggesting that linker region is not ordered. In the maps,

Table 4.1: Data collection and refinement statistics

Space group	P6 ₁
Unit cell parameters (Å)	a= b = 62.41, c= 83.26
Resolution (Å)	33.0-2.5 (3.1-2.5)*
Number of unique reflections	6307
R_{merge} (%)	8.9 (45.5)*
Completeness (%)	98.6 (98.4)*
I/σ(I)	14.2 (3.9)*
<u>Refinement statistics</u>	
R _{work} /R _{free} (%)	21.6/25.3 (25.5/34.8)*
<u>RMS deviations from ideal values</u>	
Bond lengths (Å)	0.007
Bond angles (°)	1.1
<u>Ramachandran plot:</u>	
<u>non glycine residues in</u>	
Most favoured region (%)	91.2
Additionally allowed region (%)	8.8
Generously allowed region (%)	0.0
Disallowed region (%)	0.0

*Data for highest resolution shell are given in the parenthesis.

$$R_{merge} = \frac{\sum_{hkl} \sum_j |I_{hkl,j} - \langle I_{hkl} \rangle|}{\sum_{hkl} \sum_j I_{hkl,j}}$$

Table 4.2: Kinetics data

Sample	K_m (μM)	k_{cat} (s^{-1})	k_{cat}/K_m ($\mu\text{M}^{-1} \text{s}^{-1}$)
Native	34.2 ± 6.4	4.1 ± 0.14	0.12 ± 0.02
G48V/C95F	114.1 ± 13.6	2.8 ± 0.1	0.02 ± 0.002

there was good electron density for the SQV molecule. When SQV was modeled in two orientations consistent with the two fold symmetry of functional HIV-1 protease dimer, the R factors were significantly higher. Further, the fit to electron density of SQV in the second orientation was not satisfactory. Therefore, in the present complex, SQV is binding only in one orientation. Although presence of G48V and C95F mutations in both the subunits was confirmed through DNA sequencing, electron density for VAL-48 was not very good, showing no density for atoms beyond the $C\beta$ atom. The probable side-chain conformation of VAL-48 was therefore probed by searching for a rotamer for which there would be no steric clash. In hexagonal crystals of HIV-1 protease, the residue 48 is in crystal contact with residue 1048 from a symmetry-related molecule. However, it was found that the separation between $C\gamma_2$ atoms was about 3.2 \AA , when the two residues were modeled to have same rotamer conformation. The steric clash could be relieved when the two residues were modeled to assume systematically different rotamer conformations. Thus, VAL-48/VAL-1048 appear to be statistically disordered because of their involvement in crystallographic contacts, and this could be the reason for the poor electron density. This could also be the reason for difficulties in the crystallization of this mutant. Electron density for PHE-95 was good.

The stereochemistry of the refined model is good with more than 90% of non-glycine residues occupying the most favoured regions of Ramachandran plot (Table 4.1). To study the structural effects of mutations, the present complex was compared with complexes of SQV with wild-type (PDB Id 1HXB) and G48V/L90M (PDB Id 1FB7) mutants of protease.

4.3.4 Mutation at position 48:

The overall structure of the present complex is very similar to that of the wild-type protease complex (PDB Id 1HXB), the rmsd for 198 C α atom pairs being 0.42 Å. Figure 4.2 shows a region around residue 48 for the superposed structures. One can see two differences directly related to the mutation G48V. The side chain of flap residue PHE-53 has assumed a different rotamer conformation and the quinoline ring in the P3 position of SQV has moved away from the flap. Both these effects are to avoid steric clash with VAL-48. Shifts in the positions of residues 1081, 1082, 1008 and 29 seen in Figure 4.2 are due to changes in the positions of SQV caused by G48V. The shift in the positions of the backbone atoms of residues 1079 to 1082 is in the range 0.3-0.8 Å. To investigate if the shift in residues 1079 – 1082 is solely due to G48V mutation, the present structure was compared with that of the double mutant G48V/L90M / SQV complex (Figure 4.3). The expectation was that perturbations to the protein structure arising largely out of the common mutation G48V would be similar in the two structures. It is interesting that the backbone for residues 1079 – 1082, in the G48V/L90M / SQV complex doesn't exactly superpose that of the present structure, suggesting that some contribution to movement of 80's loop (residues 1079 – 1082) also comes from the mutations C95F and L90M. Interestingly, theoretical calculations also reveal influence of L90M mutation on the

position of the 80's loop [22]. Due to altered position of 80's loop, the distance between PRO-81 and PRO-1081, which is a measure of the width of the active site cavity, is increased from 19.6 Å in wild-type complex to 20.7 Å in the present mutant complex, leading to an increase in the active site volume from 1094.7 Å³ in wild-type enzyme to 1280.3 Å³ in the present structure. The conformation of the drug molecule in the two complexes is almost identical, as expected.

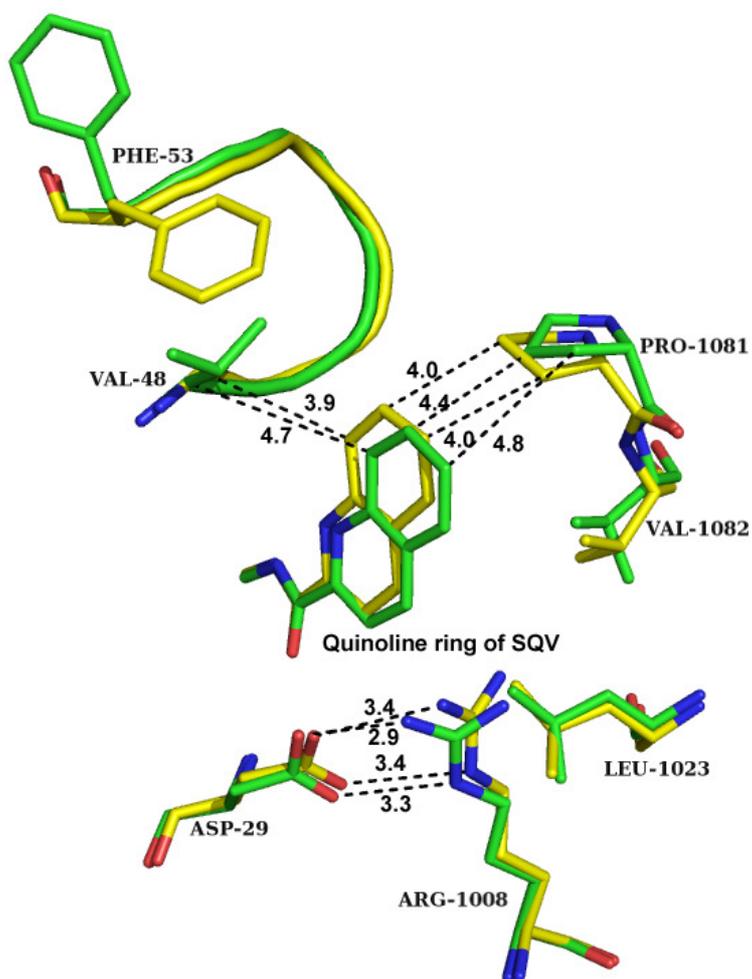


Figure 4.2: Structural effects around G48V mutation site: wild-type complex (yellow carbons, PDB Id 1HXB), present structure (green carbons). Distances are indicated in Å units.

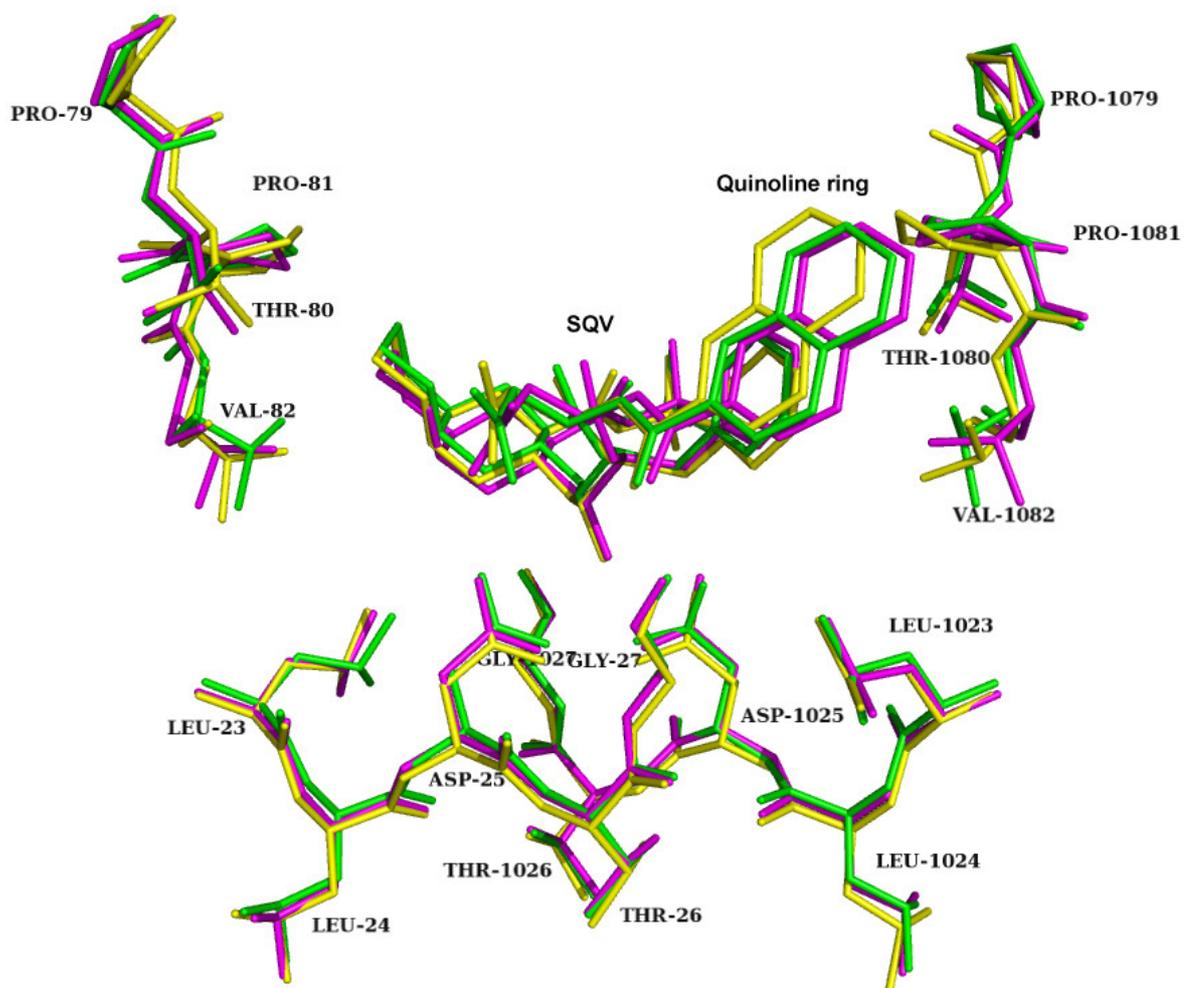


Figure 4.3: Structural comparison of present complex (green), wild-type complex (yellow) and G48V/L90M complex (purple) with the same drug SQV: Note changes in positions of quinoline ring and VAL-1082.

4.3.5 Mutation at position 95:

The residues that surround position 95 are: LEU-1097, LEU-1005, LEU-90, PHE-1099, ILE-93 and LEU-24. Similarly, the residues that surround position 1095 are: LEU-97, LEU-5, LEU-1090, PHE-99, ILE-1093 and LEU-1024. To avoid steric contact with LEU-1097, PHE-95 adopts a rotamer in which, the χ_1 conformational angle is altered by about 69° as compared to that of CYS-95 in the unmutated structure. Similarly, the change in χ_1 angle for PHE-1095 is about 104° . Again to avoid steric contacts, residue PHE-1099 has assumed altered conformation with changes in both main chain and side chain torsion angles (Table 4.3). On the other hand, only χ_2 angle of PHE-99 is changed from -76° to 94° . Another interesting adjustment is in the catalytic loop residues 23 – 27, which form the floor of the active site cavity. When compared with the wild-type complex, the backbone of this loop has shifted by about 0.3 \AA to 0.5 \AA , toward the bound inhibitor. Because of the rigid nature of this loop, the whole backbone has been moved to relieve short contacts between LEU-24 and the mutation residue PHE-95. The positions of the P1/P1' residues from SQV and the flap residues 49 – 52, that are part of the S1/S1' pocket, also have moved away in the same direction, as may be seen in Figures 4.4 & 4.5.

Table 4.3: Comparison of main chain and side chain torsion angles of Phe99 (Phe1099) in the wild-type complex (PDB Id 1HXB) and present structures

Structure	φ	ψ	χ_1	χ_2
1HXB	-165° (-159°)	12° (-4°)	70° (77°)	-76° (91°)
G48V/C95F	-168° (-102)	5° (153°)	73° (-79°)	94° (-91°)

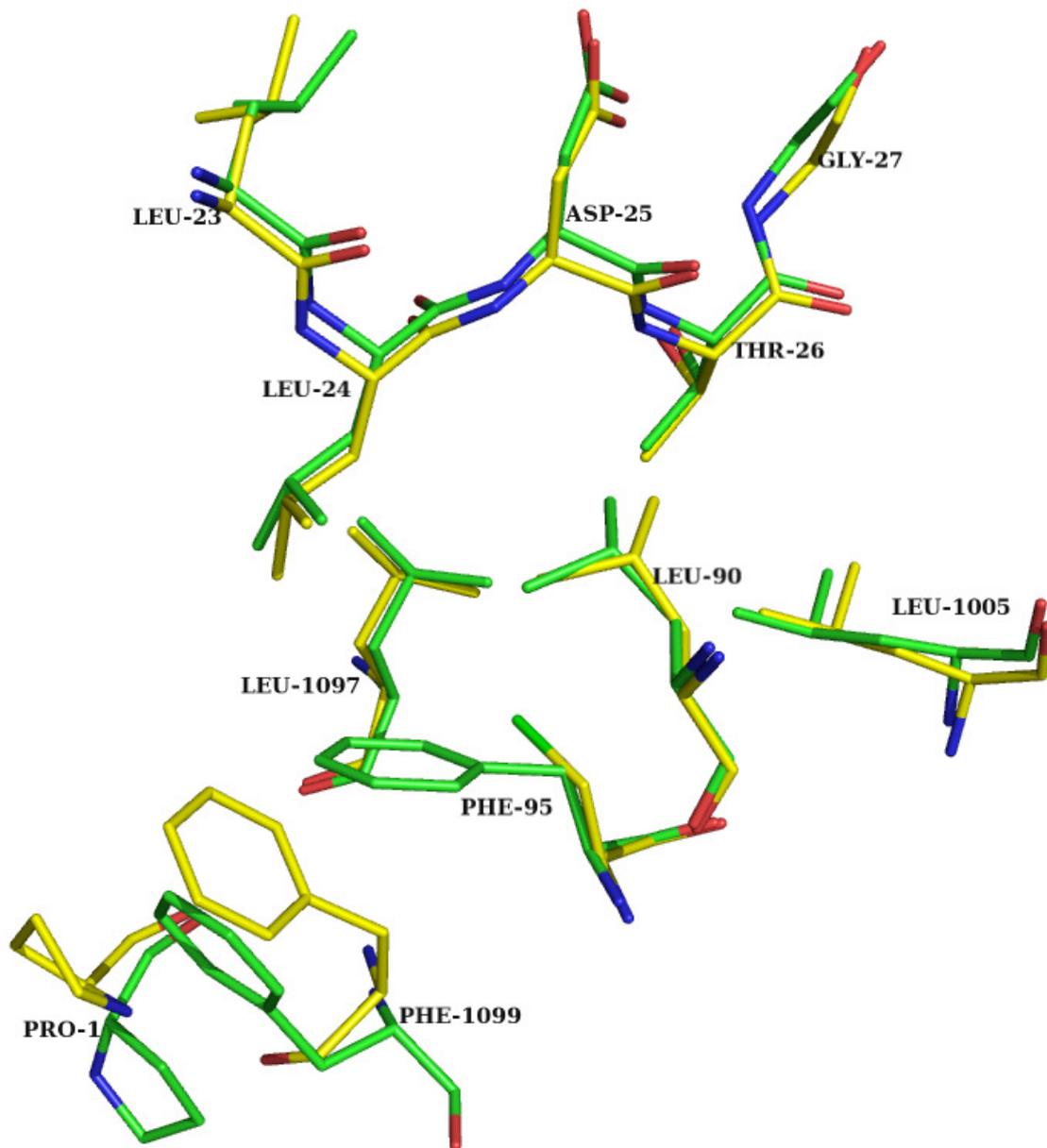


Figure 4.4: Structural effects around C95F mutation site: the catalytic loop has moved away from 95 and the N-H...O hydrogen bond between PRO-1 and PHE-1099 is lost in the present structure (green carbons). Wild-type complex is shown with yellow carbons. Note the stacking of aromatic rings of PHE-95 and PHE-1099 in the present structure.

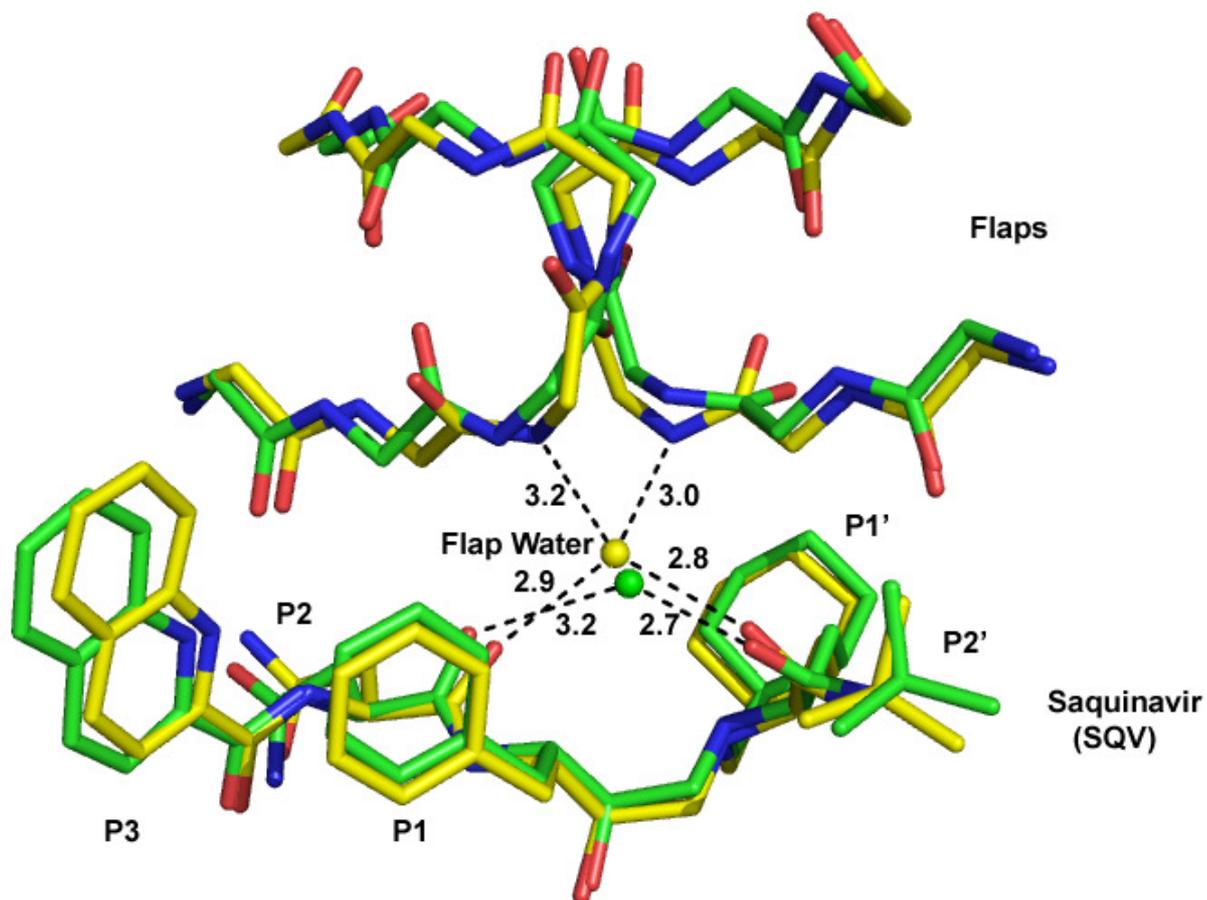


Figure 4.5: Hydrogen bonds from flap-water molecule: yellow sphere (1HXB) green sphere (present structure) Hydrogen bonds to flap amide N atoms 50 and 1050 are lost in the present structure. Distances are indicated in Å units.

4.3.6 Dimer stability:

Residues 95 and 99 are part of the inter-monomer four stranded beta sheet that contributes significantly to the dimeric stability in HIV-1 protease. The altered conformations of these residues as mentioned above, has had a domino-effect on other residues in the beta sheet. The C α atom of PRO-1 residue has been displaced by as much as 1.3 Å, with the result that there is no hydrogen bond between PRO-1 N and PHE-1099 CO, in the present structure (Figure 4.4), and therefore, the dimer is expected to be destabilised. On the other hand, the changes in conformations of ARG-8 and ARG-1008, have led to stronger hydrogen bonds with ASP-1029 and ASP-29 respectively (from 3.4 Å to 3 Å, Figure 4.2) thereby stabilizing the dimer. Further, the stacking interactions between PHE-95 and PHE-1099 are non-existent in the wild-type complex. Since changes have occurred in both dimer-stabilising and dimer-destabilising interactions, it is difficult to estimate their net influence on the overall stability of the dimer. We therefore decided to actually measure the melting temperatures of both the wild-type and present mutant using CD spectroscopy. It is interesting that the T_m's are lower for the mutant both in the presence and absence of the inhibitor SQV (Table 4.4). A lower T_m of the mutant in the absence of SQV indicates that the drug-resistant mutant is less stable compared to the wild-type enzyme. In the presence of SQV the dimer is stabilized and the T_m's are increased for both, as expected. However, the increase is by a smaller magnitude in the case of drug-resistant mutant, which suggests that the gain in free energy on drug binding is lower in the mutant G48V/C95F.

Table 4.4: T_m values for native and G48V/C95F mutant proteins in the presence and absence of saquinavir (SQV)

Molecular samples	T_m (°C)
Native	66
Native+SQV	80
G48V/C95F	60
G48V/C95F+SQV	68

4.3.7 Drug resistance:

Drug resistance is often associated with mutations of HIV-1 protease in both active site and non-active site regions, even though mutations at target cleavage sites have also been reported [8]. Effects of active site mutations can be explained on the basis of altered interactions with the inhibitors, but a variety of explanations have been advanced for non-active site mutations [24-27]. In the present system, which contains one active site (G48V) and one non-active site (C95F) mutation, the increase in the volume of the active site cavity implies non-optimal van der Waals interactions between the drug and the mutant enzyme. Due to altered positions of the quinoline ring of the drug at the P3 subsite and of PHE-53 of the enzyme, there is a loss in the van der Waals interactions (Figure 4.2). In most of the HIV-1 protease inhibitor complexes, a water molecule known as flap water, bridges the P2 and P1' carbonyl groups of the inhibitor molecule to the ILE-50 and ILE-1050 amide groups of the flaps. In the present mutant complex, these

hydrogen bonds are absent (Figure 4.5). The movement of the catalytic loop away from Phe-95 and toward the bound inhibitor adversely affects surface complementarity between SQV and the enzyme, thereby lowering the binding affinity [9]. Further, the structural flexibility of the catalytic loop needed to accomplish effective drug binding has been diminished by the mutation C95F, as this loop in the present structure is unable to match its position to that observed in structures of HIV-1 protease/ ligand complexes. The lower stability of the G48V/C95F dimer is another factor contributing to drug-resistance [26]. As mentioned before, C95F mutation is often found to be part of the mutation cluster C95F/G48V/V82A. The present structure provides a rationale for how V82A mutation can further enhance drug-resistance against SQV. The residue VAL-1082 adopts a different rotamer compared to wild-type complex, in order to avoid steric clashes with the P1 and P3 residues of SQV (Figure 4.2). As a result the number of favourable van der Waals interactions between C γ atoms of VAL-1082 and P1 residue from SQV has decreased from 10 to 6 contributing to a loss of affinity. If VAL-1082 were changed to ALA-1082, which does not have any gamma carbon atoms, these van der Waals interactions would be totally lost, thereby further decreasing the affinity for the drug.

Chapter 5

X-ray structure analysis of L90M/C95F and C95F mutants of tethered HIV-1 protease dimer in complex with saquinavir

5.1 Introduction:

As mentioned in chapter 4 of this thesis, Saquinavir (Ro31-8959) (SQV) was the first HIV-protease inhibitor (PI) to be released in the market for chemotherapy against HIV/AIDS. It is a peptidic inhibitor with a sub-nano molar inhibition constant (K_i). Mutants that display resistance against SQV are G48V and L90M. While G48V is an active site mutation, L90M is outside the active site. Even after initial onset of resistant mutations in HIV protease, virus undergoes additional mutations under continuous selection pressure of the drugs for longer periods of time [1]. C95F is one such mutation. From a statistical analysis of sequences of protease enzyme isolated from AIDS patients, both treated and untreated with PI's, it has been found that the mutation C95F occurs as a mutation cluster either with L90M and I93L or with G48V and V82A [2].

L90M is a major non-active site mutation against the drug SQV, while C95F is a major non-polymorphic mutation associated with L90M [3]. No SQV complexes have been reported for mutant L90M. A lower resolution (2.6 Å) crystal structure has been reported for the SQV complex with the double mutant G48V/L90M [4]. Crystal structures of L90M/C95F and C95F tethered HIV-1 protease, in complex with SQV are reported here. These structures were compared among themselves and with the wild-type complex

(PDB Id 1HXB) to reveal the effects of individual mutations on protein structure and inhibitor binding.

5.2 Materials and methods:

5.2.1 Site directed mutagenesis:

The L90M and C95F mutations were carried out using Quick change site directed mutagenesis kit (Stratagene, La Jolla, CA) following protocol explained in chapter 4 of this thesis. Oligonucleotide primers 5'-TTATCGGCCGTAACCTGATGACTCAGATCGGTAAAACC-3' and 5'-CCCGCCAAAGTTCAGAGTAAAACCGATCTGAGTCAGCAGG-3' were used for making L90M and C95F mutations, respectively. Presence of mutations was confirmed through DNA sequencing.

5.2.2 Protein expression and purification:

HIV-1 protease tethered dimer used in the present study contains a five residue linker, GGSSG, covalently linking the two monomers [5]. Expression, and purification of HIV-1 protease tethered dimer followed the procedures reported earlier [6] and explained in chapter 3 of this thesis.

5.2.3 Circular dichroism spectroscopy:

CD spectra on wild-type and mutant proteins were recorded using a Jasco J-815 CD spectrometer. Protein solutions were used at a concentration of 10 μ M in 50 mM sodium acetate buffer, pH 4.5. Thermal denaturation experiments were carried out by monitoring the CD signal at 215 nm over a temperature range of 20–80° C with a scan rate of

2°C/min. Temperatures of midpoint unfolding transitions (T_m) were estimated by evaluating the maximum of the first derivative of $[\Theta]_{215}$ in relation to the temperature data [7].

5.2.4 Protease assay:

Assay with chromogenic substrate of amino acid sequence His-Lys-Ala-Arg-Val-Leu*NPhe-Glu-Ala-Nle-Ser (where * denotes the cleavage site, and NPhe and Nle denote *p*-nitrophenylalanine and norleucine residues respectively) was carried out as explained in chapter 2 of this thesis. Values for k_{cat} and K_m were obtained by fitting the curves to the Michaelis-Menten equation using Graphpad Prism software (Figure 5.1).

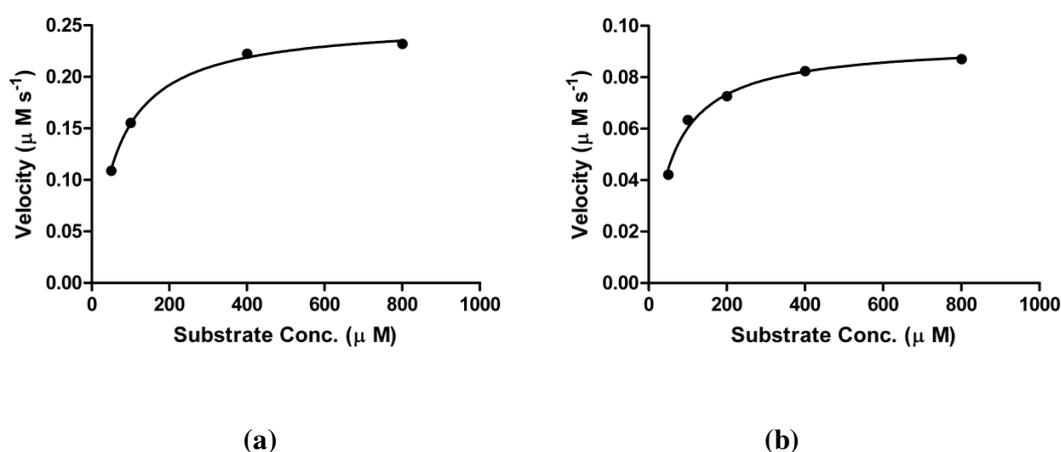


Figure 5.1: Determination of kinetic constants: Velocity vs. Substrate concentration curves for (a) L90M/C95F (b) C95F mutants

5.2.5 Crystallization:

Single crystals were obtained by the hanging drop vapour diffusion method. Protein (5 mg/ml in 50mM sodium acetate, pH 4.5, containing 1 mM dithiothreitol) was reacted at

room temperature for 30 minutes with ten-fold molar excess of SQV dissolved in dimethyl sulfoxide. For crystallization, equal volumes of reaction mixture and reservoir solution (1% saturated ammonium sulfate, 200 mM sodium phosphate, and 100 mM sodium citrate at pH 6.2) were mixed on a cover slip and sealed over the reservoir well at room temperature.

5.2.6 X-ray data collection and refinement:

The crystals were equilibrated in the cryo-protectant (25% glycerol and 75% reservoir buffer) before flash freezing, for exposure to X-rays on the FIP-BM30A beam line [8]. The diffraction data were indexed, integrated, and scaled by using the computer program XDS [9].

Crystal structure was solved by difference Fourier method using native coordinates (PDB Id 1LV1). The structure was refined in computer programme REFMAC 5.5.0072 by using the amplitude-based maximum likelihood target function [10]. A test set containing 5.0% of randomly chosen reflections were reserved for determination of R_{free} [11]. All interactive model building and molecular superpositions were carried out using the molecular modeling software O [12]. Structural comparisons were based on superpositions of protein C α atoms. The quality of model was checked using programme PROCHECK [13]. All figures were drawn using computer program PYMOL [14]. Active site volumes were calculated using CASTp server with a probe sphere radius of 1.4 Å [15].

5.3 Results:

5.3.1 Kinetics and Stability of mutants:

Purified L90M/C95F and C95F mutants exhibited a K_m of $64.7 \pm 4.1 \mu\text{M}$ and $55.3 \pm 5.8 \mu\text{M}$ respectively and a k_{cat} of $5.1 \pm 0.08 \text{s}^{-1}$ and $3.1 \pm 0.1 \text{s}^{-1}$, respectively. The catalytic efficiencies (k_{cat}/K_m), of L90M/C95F and C95F mutants, for hydrolysis of the chromophoric substrate, were about 67% and 50% respectively, w.r.t. wild type protease (Table 5.1). Mid transition temperatures (T_m) as measured using CD spectrometry for L90M/C95F and C95F mutants were 60°C and 58°C , respectively as against 66°C for wild type protease.

Table 5.1: Kinetics data

Sample	K_m (μM)	k_{cat} (s^{-1})	k_{cat}/K_m ($\mu\text{M}^{-1} \text{s}^{-1}$)
Native	34.2 ± 6.4	4.1 ± 0.14	0.12 ± 0.02
L90M/C95F	64.7 ± 4.1	5.1 ± 0.08	0.08 ± 0.005
C95F	55.3 ± 5.8	3.1 ± 0.1	0.06 ± 0.006

5.3.2 Crystal structures:

Results of data collection and crystallographic refinement are summarized in Table 5.2. The crystal structures of the L90M/C95F_SQV and C95F_SQV complexes were refined to $R_{\text{work}}/R_{\text{free}}$ values of 20.4/24.3% and 20.6/24.5% at resolutions of 2.04 and 1.75 Å, respectively. The stereochemistry of both the refined models was good with more than

90% of non-glycine residues occupying the most favoured regions of Ramachandran plot (Table 5.2). The crystallographic asymmetric unit contains a tethered dimer of HIV-1 protease, in which residues in the first monomer are numbered 1–99 and those in the second monomer 1001–1099. Residues of the drug SQV are labeled P3-P1 /P1'-P2' (and corresponding binding sites of protein as S1-S3/S1'-S2') as per Schechter and Berger nomenclature [16]. Electron densities for mutated residues were unambiguous (Figure 5.2). Residues Gln1007, Thr1012 and Glu1035 were present in alternate conformations in the C95F_SQV complex. There was good difference electron density for SQV in the active sites of both the mutant complexes and indicated that drug was bound only in one orientation in both the mutants.

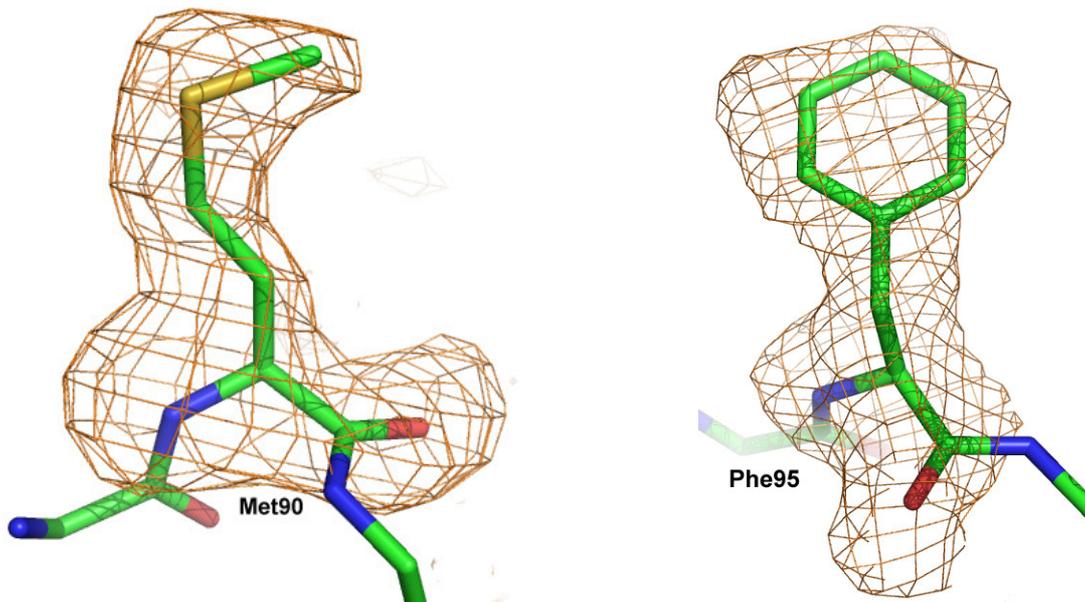


Figure 5.2: F_0-F_C omit maps for Methionine 90 and Phenylalanine 95: Maps are contoured at 2σ level, showing electron densities for mutated residues Methionine 90 and Phenylalanine 95.

Table 5.2: Data collection and refinement statistics

Structure	L90M/C95F	C95F
Space group	P6 ₁	P6 ₁
Unit cell parameters (Å)	a= b = 62.68,c= 82.62	a= b = 62.19,c= 81.84
Resolution (Å)	54.29-2.04 (2.09-2.04)*	53.86-1.75 (1.79-1.75)*
Number of unique reflections	11476	17195
R _{merge} (%)	6.6 (52.8)*	3.7(49.8)*
Completeness (%)	97.5 (98.6)*	99.87 (99.85)*
I/σ(I)	18.3 (3.5)*	23.8(3.2)*
<u>Refinement statistics</u>		
R _{work} /R _{free} (%)	20.4/24.3 (25.4/35.6)*	20.6/24.5 (19.3/28.2)*
<u>RMS deviations from ideal values</u>		
Bond lengths (Å)	0.008	0.01
Bond angles (°)	1.12	1.3
<u>Ramachandran plot:</u>		
<u>non glycine residues in</u>		
Most favoured region (%)	91.2	93.7
Additionally allowed region (%)	8.8	6.3
Generously allowed region (%)	0.0	0.0
Disallowed region (%)	0.0	0.0

*Data for highest resolution shell are given in the parenthesis.

$$R_{merge} = \frac{\sum_{hkl} \sum_j |I_{hkl,j} - \langle I_{hkl} \rangle|}{\sum_{hkl} \sum_j I_{hkl,j}}$$

5.3.3 Effects of L90M mutation:

In HIV-1 protease, residue 90 is part of a short helix (residues 86-94) near one end of which is the loop containing active site aspartates and near the other end is the four stranded inter-subunit beta sheet (Figure 1.4, chapter 1). Side chain of mutated residue Met90/1090 extends into a small pocket formed by residues Asp25/1025 and Ile85/1085-Arg87/1087 (Figure 5.3).

To dissect out the effect of L90M mutation, L90M/C95F_SQV complex was compared to the C95F_SQV complex. The two structures superposed to an rmsd of 0.4 Å over 198 C α atom pairs. Introduction of bulkier methionine side chain at position 90/1090 leads to some additional van der Waals contacts. S δ and C γ atoms of Met90/1090 side chain form van der Waals contacts with carbonyl oxygen and C γ 2 atoms of Ile85/1085. Methionine side chain also makes van der Waals interactions with main chain and side chain atoms of residues Leu24/1024, Asp25/1025 and Thr26/1026 (Figs. 5.3 and 5.4). In subunit 1, C ϵ methyl of methionine interacts with C β of Leu24 while in subunit 2, it interacts with C γ 2 of Thr1026. This results in the shift in the backbone atoms of Ile85/1085 with its C α shifted by 0.3Å (Figure 5.3). Since, Ile85/1085 carbonyl oxygen and amide nitrogen are hydrogen bonded to amide nitrogen and carbonyl oxygen of residues Asp25/1025 and Leu24/1024, respectively, a shift in the backbone atoms of the residues Asp25/1025 and Leu24/1024 was also seen (Figure 5.4). S δ of Met90/1090 can also accept a hydrogen bond from main chain amide of Asp25/1025. In subunit 2, C α of Asp1025 has shifted by 0.3Å while for Asp25 in subunit 1 this shift was 0.2Å. But χ 2 torsion angle of Asp25 has changed by 18° (Table 5.3). A slight dislocation of the Asp25/1025 residue causes rotation around C α -C β bond of the side chain at the 84/1084th residue (Figure 5.5). O δ 1

atom of Asp25/1025 is hydrogen bonded to the amide nitrogen of Gly27/1027. Therefore shift in the main chain and side chain of Asp25/1025 also results in the backbone shift of Gly27/1027.

Table 5.3: Comparison of conformation of Asp25/1025 in L90M/C95F_SQV and C95F_SQV complex structures

Torsion angles	Asp25		Asp1025	
	L90M/C95F	C95F	L90M/C95F	C95F
χ_1 (°)	-166	-166	-168	-168
χ_2 (°)	-25	-7	5	-6
ϕ (°)	-117	-123	-127	-130
ψ (°)	-80	-89	-94	-186

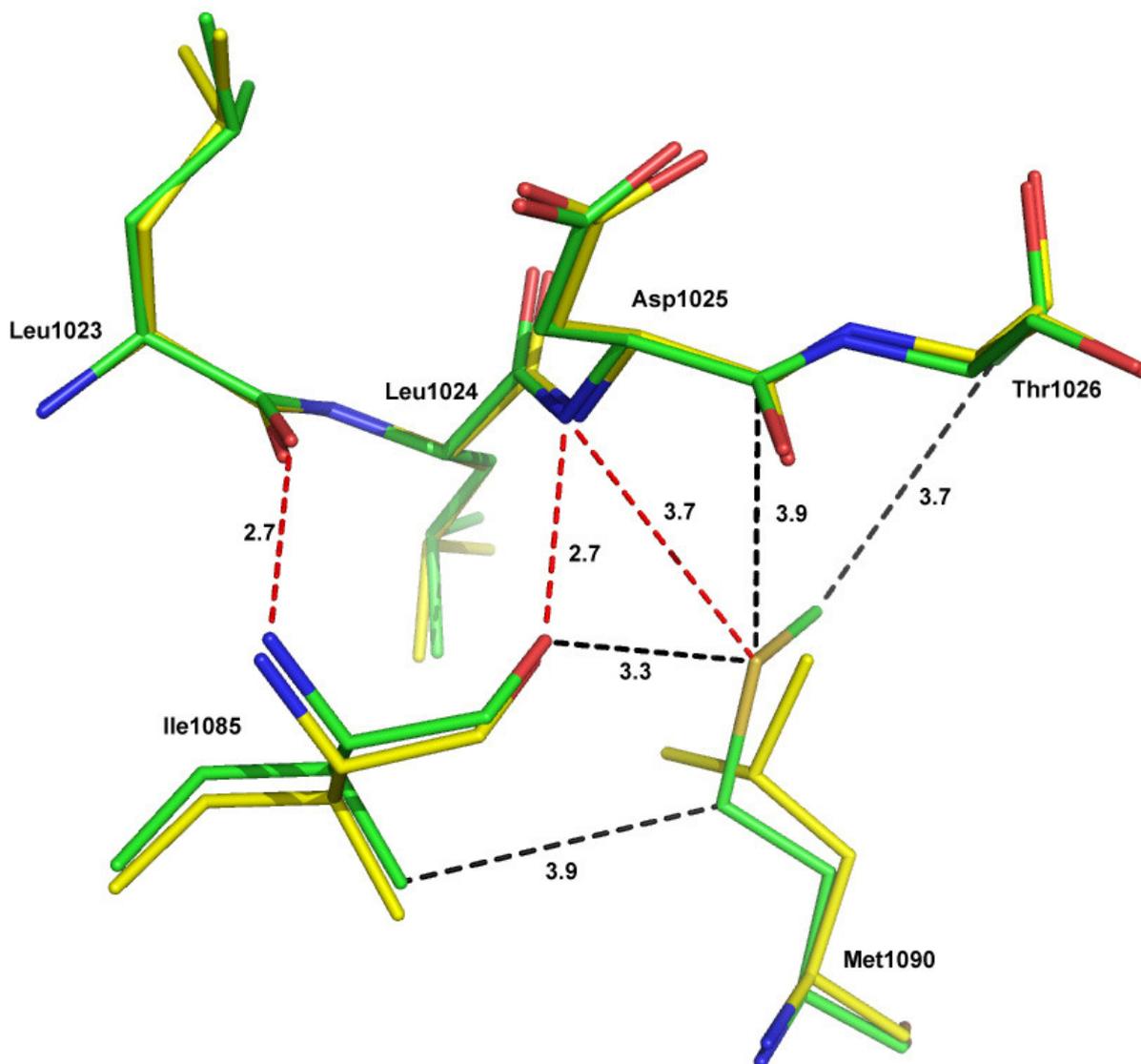


Figure 5.3: Structural comparison of L90M/C95F_SQV (green carbons) complex with C95F_SQV complex (yellow carbons). Black dashes refer to van der Waals interactions, while red dashes correspond to hydrogen bonds. Distances are indicated in Å units.

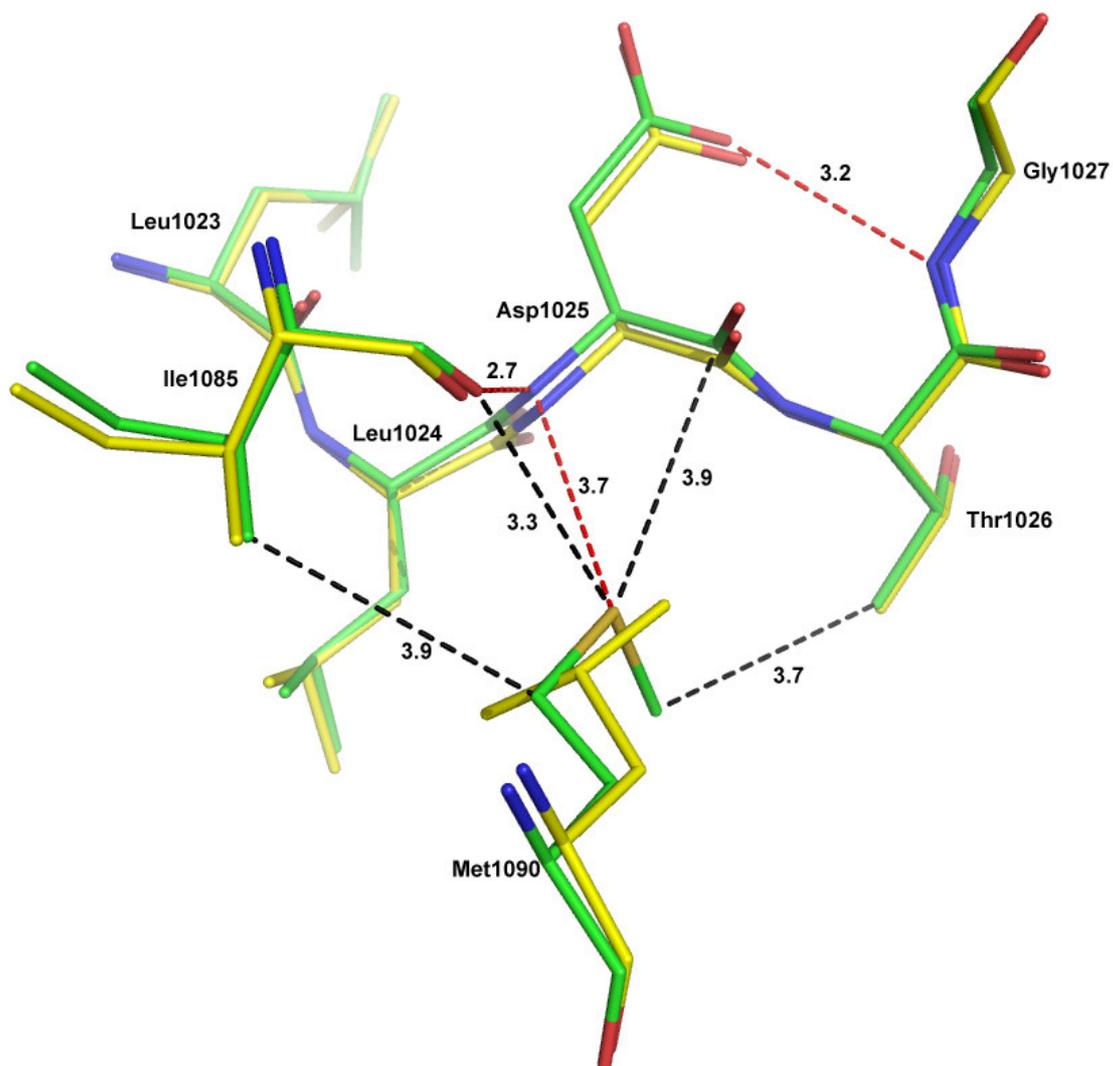


Figure 5.4: Structural effects on the active site floor: comparison of L90M/C95F_SQV complex (green carbons) with C95F_SQV complex (yellow carbons). Black dashes refer to van der Waals interactions, while red dashes correspond to hydrogen bonds. Distances are indicated in Å units.

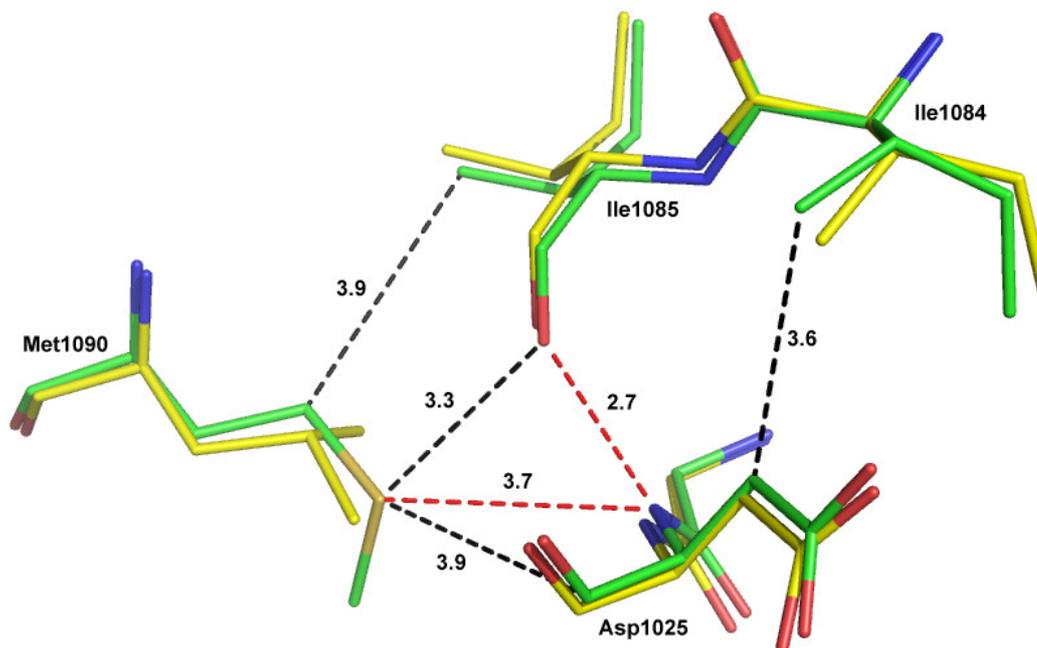


Figure 5.5: Rotation of Ile1084 side chain around χ_1 torsion angle: Structural comparison of L90M/C95F_SQV (green carbons) complex with C95F_SQV complex (yellow carbons). Black dashes refer to van der Waals interactions, while red dashes correspond to hydrogen bonds. Distances are indicated in Å units.

As residue 90/1090 is part of the small helix 86-1086/94-1094, introduction of methionine at position 90/1090 results in the shift in the backbone atoms of residues 87/1087 to 89/1089 (Figure 5.6). Backbone amide nitrogens of residues 87/1087 and 88/1088 are hydrogen bonded to the carbonyl oxygens of residues 29/1029 and 28/1028, respectively. Therefore, shift in the main chain atoms of these residues was also noticed. Overall effect is the shift in the main chain of residues 24/1024 through 29/1029 which include active site triad. Thus, the L90M mutation results in perturbations at the catalytic

centre via residues 85-1085/88-1088, leading to the reshaping of the floor of the inhibitor binding cavity.

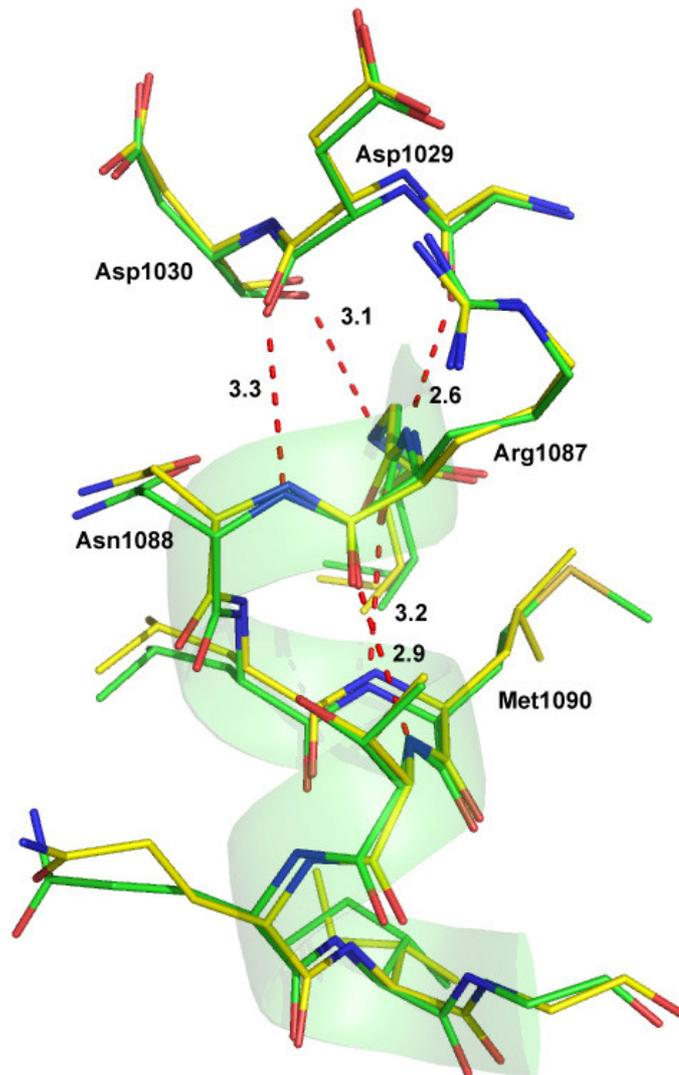


Figure 5.6: Shift of residues in the helix due to L90M mutation: L90M/C95F_SQV complex (green carbons), C95F_SQV complex (yellow carbons). Note concerted shift of Asp1029. Distances are indicated in Å units.

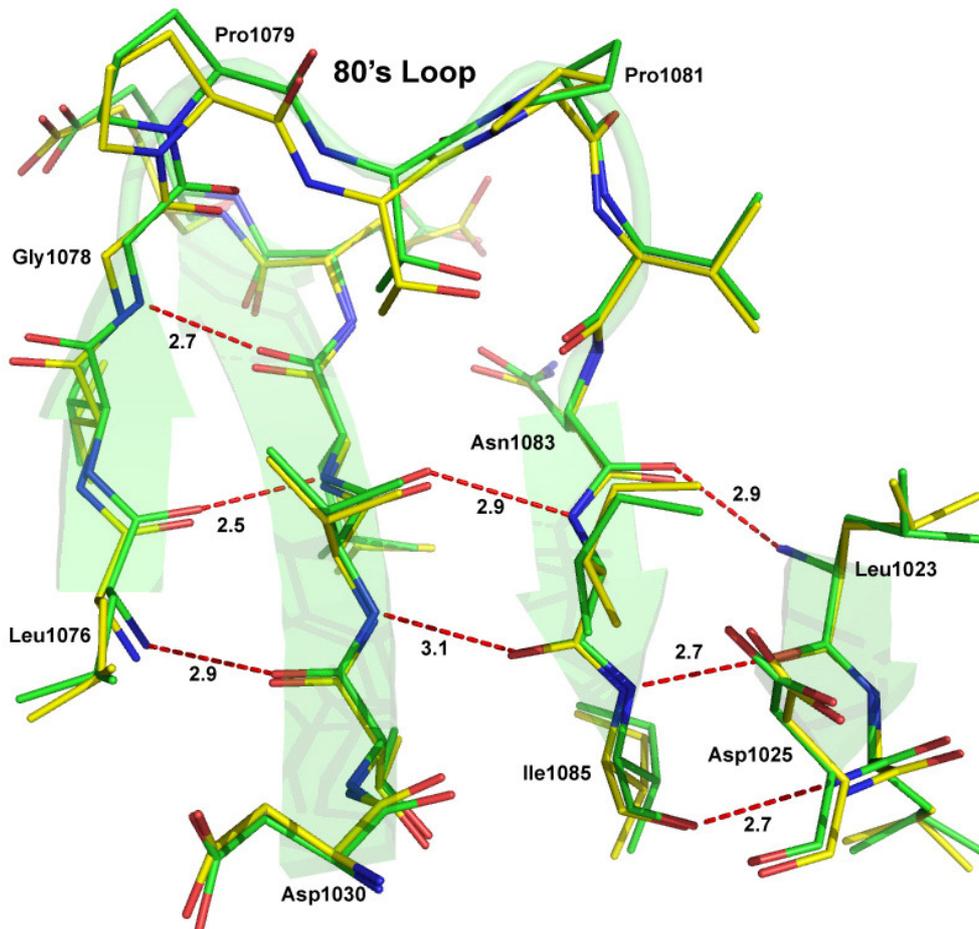


Figure 5.7: Shift in the 80's loop: L90M/C95F_SQV complex (green carbons), C95F_SQV complex (yellow carbons). Distances are indicated in Å units.

The molecular core of HIV-protease is characterized by the presence of a Ψ -shaped sheet made up by four strands comprising of residues 23-25/1023-25, 30-35/1030-35, 76-78/1076-78 and 83-85/1083-85, respectively (Figure 5.7). Movement of the residues 23-25/1023-25 and 83-85/1083-85 described before is accompanied by slight shifts in the other two strands. This is accompanied by shift of 80's loop (residues 79-82/1079-1082) which joins two of these strands (Figure 5.7). In subunit 2, 80's loop has bulged out while in subunit 1, it has shifted laterally. The 80's loop is highly flexible and adaptable region

of the enzyme and the conformation of the loop can also be influenced by mutations in the enzyme [17]. When compared to the apo-enzyme, there is an outward movement of 80's loop upon SQV binding but this movement is even more pronounced in the mutant structures. C ϵ methyl group of Met90/1090 also forms additional van der Waals contacts with mutated Phe95/1095 side chain from the same subunit and Leu 1097/97 and Leu1005/5 side chains from the other subunit (Figure 5.8)

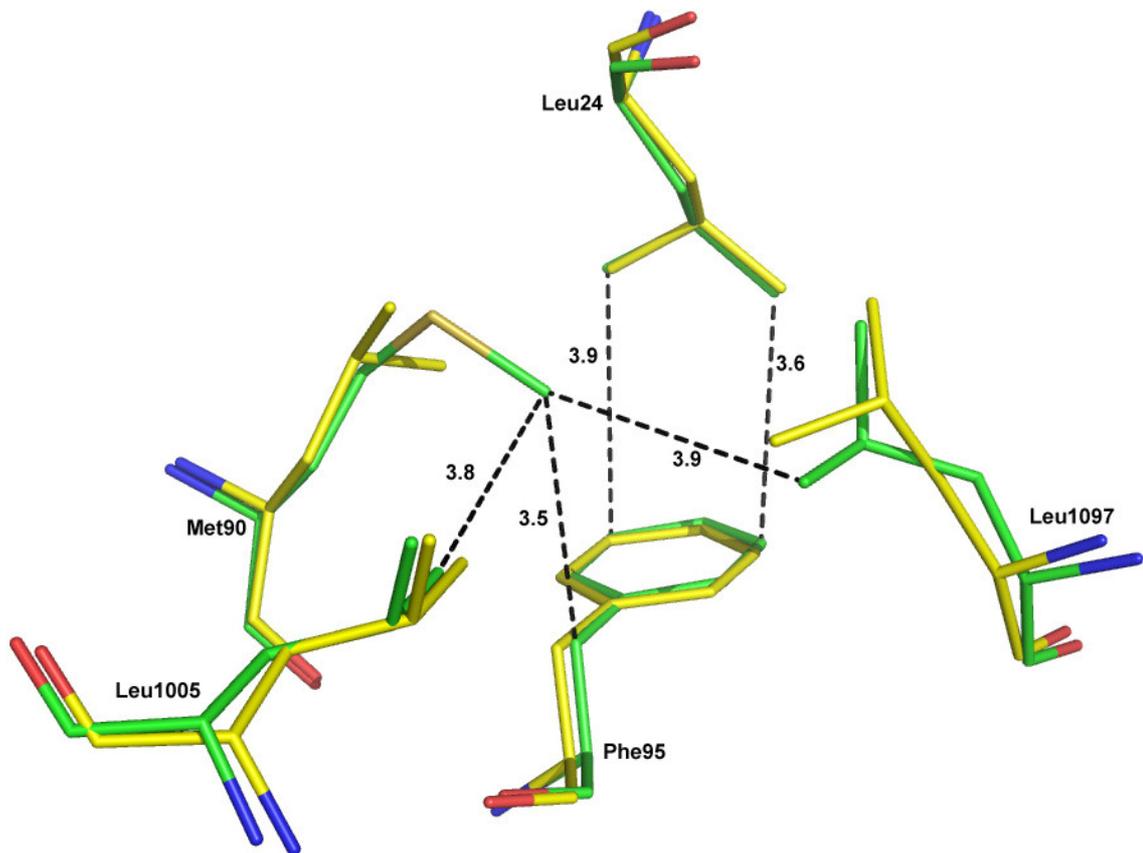


Figure 5.8: Additional interactions of C ϵ methyl group of Methionine90: L90M/C95F_SQV complex (green carbons), C95F_SQV complex (yellow carbons). Distances are indicated in Å units. Black dashes refer to van der Waals interactions.

5.3.4 L90M/C95F mutant-SQV interactions:

Figure 5.9 shows the structural superposition of L90M/C95F_SQV complex with native HIV-1 protease-SQV complex (PDB Id 1HXB). The two structures superpose to an rmsd of 0.43 Å over 198 C α atom pairs. When compared to the native HIV-1 protease- SQV complex, interesting changes were observed in the conformation of the drug as well as of the protein. To avoid short contacts and adjust to the reshaping of active site cavity, there were many concerted shifts both in the inhibitor as well as the binding pockets of the enzyme. Phenylalanine ring at P1 site of SQV has rotated with χ_2 and χ_1 torsion angles changing by about 9° and 13° respectively, when compared to the native-SQV complex. Since in the 3-D structures of SQV, P1 Phenylalanine ring is in close proximity with the Quinoline ring at P3 position; rotation of P1 ring has caused rotation of P3 ring around χ_2 by about 11°. To avoid steric contacts with the P3 quinoline, Arg1008 side chain moves away. Slight shift in P1' iso-quinoline ring was also seen (Figure 5.9). These changes in the conformations of both the enzyme and the inhibitor lead to loss of van der Waals interactions between P1 Phenylalanine of the drug and Ile1023 and Val1082 of the enzyme. Similarly van der Waals contacts between P3 quinoline and Pro1081 are lost.

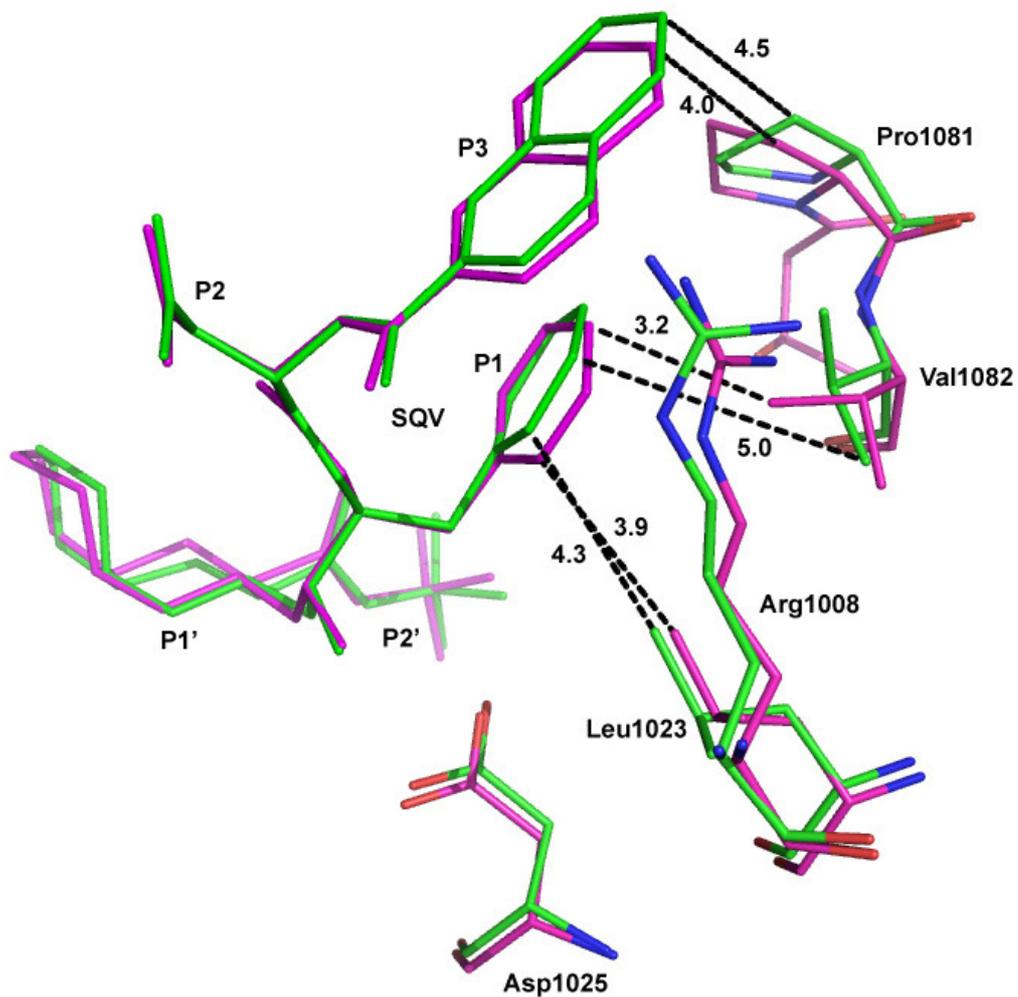


Figure 5.9: Loss of interactions of SQV in the active site: L90M/C95F_SQV complex (green carbons), wild type complex (magenta carbons). Distances are indicated in Å units. Van der Waals cut off is 4 Å.

5.3.5 Effects of C95F mutation:

To see the effects of C95F mutation, native HIV protease-SQV complex (PDB Id 1HXB) structure was superposed onto the C95F_SQV complex structure using only C α atoms for the superposition. The overall structure of the mutant is very similar to that of the native protease, the rmsd for 198 C α atom pairs being 0.42 Å. Residue 95 is part of the four stranded inter subunit beta sheet located at the dimer interface. The residues that make contact with Phe95/1095 are Leu1097/97, Leu1005/5, Leu90/1090, Phe1099/99, and Leu24/1024. At the mutation site, there are two changes compared to the wild type complex. One, the C α of Phe1095 has shifted by around 0.8 Å and, two, the χ 1 angle in Phe is -83.6°, while it is -158.5° for Cys. These two changes lead to avoidance of steric contacts with Ile97 (Figure 5.10). As a consequence of altered conformation of Phe1095, the rotamer Ile1093 adopts is also different in the mutant. This leads to a cascade of effects with Ile1066 adopting a different rotameric conformation to avoid short contact with Ile1093 and Ile1064 in turn adopts a different rotamer to avoid short contact with Ile1066. In the first subunit, C α of Phe95 has not shifted much. Instead C α of Leu93 has shifted by around 0.5 Å to avoid steric contacts. Here, change in the side chain conformations of Leu93 and corresponding changes in the side chain conformations of Ile66 and Ile64 were absent. Phe1095/95 ring is stacked upon Phe99/1099. Mutated side chain of Phe95/1095 can now form van der Waals interactions with the side chain of Leu 24/1024. Main chain of Phe99/1099 is pushed down in the process which further leads to the downwards movement of Pro1001/1 by as much as 1.9 Å. The hydrogen bonds between Pro1001/1 and Phe99/1099 are broken in the process (Figure 5.11). All these movements also result in the loss of van der Waals interactions between Leu1024/24,

Ile1093/93 and Phe99/1099. These interactions are important forces, amongst others, in maintaining the dimeric state. To avoid steric contacts, residue Phe99 has assumed slightly altered conformation, with the χ_2 angle changed from 91° to 114° . His 1069/69 side chain also adopts a different conformation to avoid short contacts with Phe99/ 1099 ring (Figure 5.10).

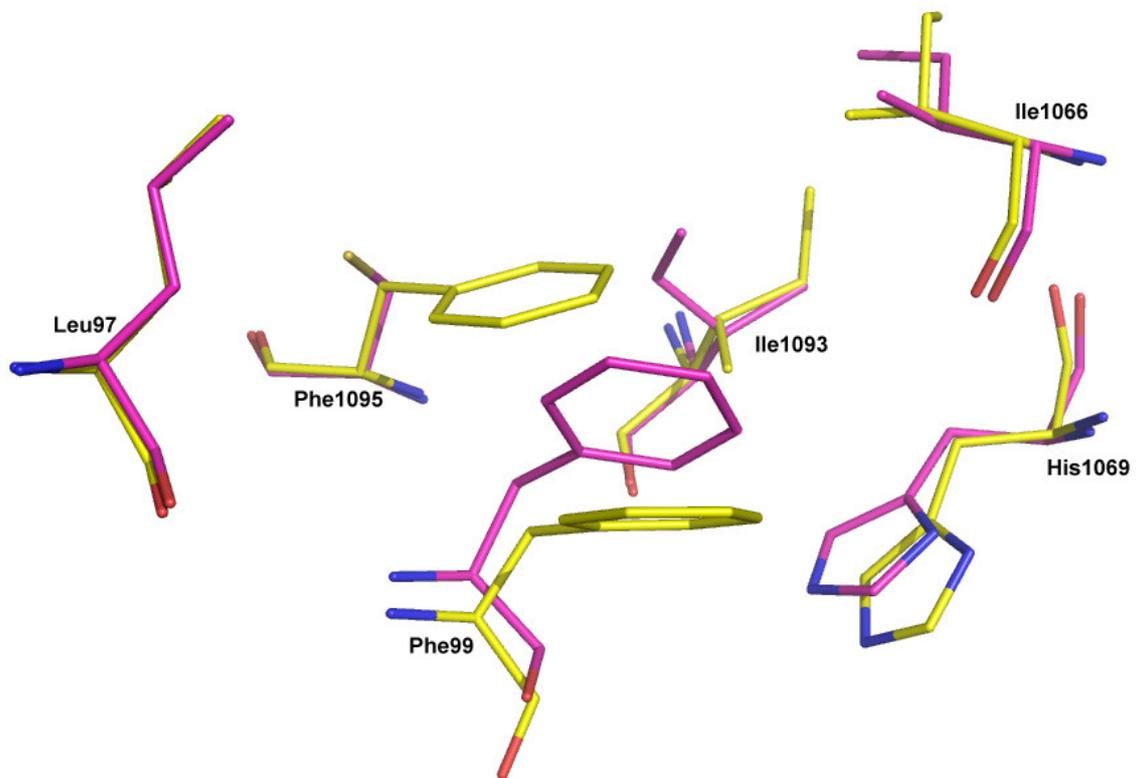


Figure 5.10: Local effects of C95F mutation in the second subunit: C95F_SQV complex (yellow carbons), wild type complex (magenta carbons)

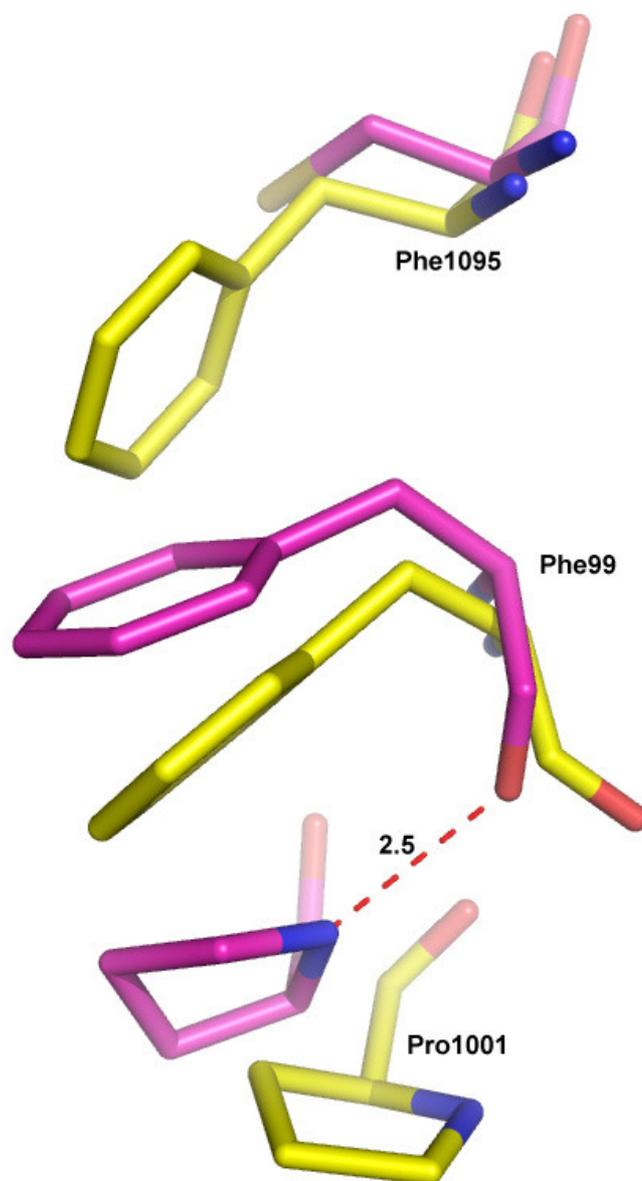


Figure 5.11: Loss of hydrogen bond in the four stranded inter subunit beta sheet: C95F_SQV complex (yellow carbons), wild type complex (magenta carbons). Distances are indicated in Å units.

5.4 Discussion:

Structurally, drug resistant mutations of HIV-1 protease can be categorized into active and non-active site mutations. Many of the active site residues are in direct contact with the bound inhibitor, thus active site mutants may directly affect the drug binding. But the effect of non-active site mutants is hard to comprehend. L90M is a major non-active site mutation of HIV-1 protease against the drug SQV. It also provides cross-resistance to some other drugs targeting HIV-1 protease. Therefore, it could provide resistance to these drugs in some general way. Comparisons of SQV bound mutant structures in the present study with that of native enzyme have shown reshaping of the active site cavity in the mutant structures. The active site volume has increased by 203 \AA^3 in L90M/C95F mutant and by 58 \AA^3 in C95F mutant (Table 5.4). This is the result of altered packing in the core of the enzyme. SQV and in fact most of the other anti HIV-1 protease drugs are designed as competitive inhibitors so that they bind to the enzyme more tightly than its natural substrate. Hence, any change in the shape of active site cavity may affect the binding of these drugs more than that of natural substrates. Small changes in the conformation of SQV were also seen in response to those in the active site. Conversely, flexible substrates could adapt to these changes comparatively easily. Catalytic loop residues 23–27 form the floor of the active site cavity. This loop is one of the most rigid, but adaptable, regions in the structure of HIV-1 protease, and is observed to be pushed down towards the four stranded β -sheet interface when a substrate/inhibitor is bound in the active site. Presence of these mutations may limit the structural flexibility of the active site loop which is required to accommodate the incoming inhibitor.

Table 5.4: Comparison of active site volumes and T_m 's

Structure	Active site volume (\AA^3)	T_m ($^{\circ}\text{C}$)
Wild Type_SQV	1094.7	66
L90M/C95F_SQV	1298.2	60
C95F_SQV	1153.4	58

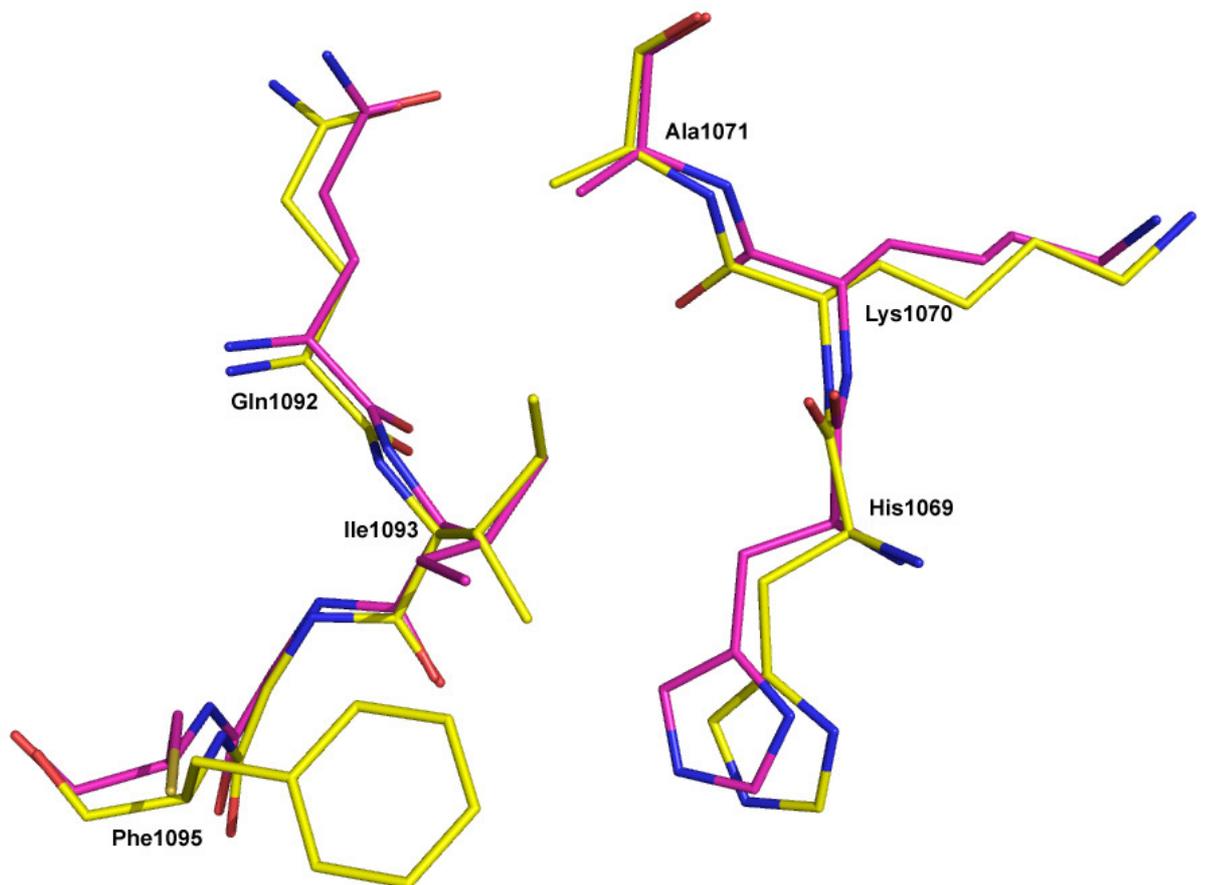


Figure 5.12: Repacking at the interface of core and terminal domains: C95F mutant (yellow carbons), wild type complex (magenta carbons)

C95F is a major non-active site polymorphism associated with L90M mutation. The structural changes in the C95F mutant-SQV complex suggest the molecular basis for reduced stability of the HIV protease variants carrying this mutation. Mid transition temperatures (T_m) as measured using CD spectrometry for G48V/C95F, L90M/C95F and C95F mutants were 60°C, 60°C and 58°C, respectively as against 66°C for wild type protease (Table 5.4). This indicates that major destabilization of dimer comes from C95F mutation. Xie et al had suggested reduction in dimer stability as a mechanism of drug resistance, and results presented here support that [18].

Rose et al [19] have divided HIV-protease into five rigid domains, which move relative to one another on ligand binding. They then suggest that mutations at these domain interfaces could influence affinity towards ligand binding by affecting inter-domain movement. Figure 5.11 shows that C95F mutation has altered the position of residues at the interface of terminal and core domains. And hence, C95F mutation may play a role in reducing the affinity towards inhibitors through the mechanism suggested by Rose et al.

Chapter 6

X-ray structure analysis of D30N tethered HIV-1 protease dimer/saquinavir complex

6.1 Introduction:

The arrival of HIV protease inhibitors (PIs) in late 1995 marked the beginning of an important era of AIDS chemotherapy [1]. The development of these PIs and their introduction into Highly Active Anti-Retroviral Therapy (HAART) along with reverse transcriptase inhibitors was a major step forward in the battle against HIV/AIDS [2]. HAART treatment regimens have significantly reduced the morbidity and mortality caused by HIV/AIDS [3]. Despite these important advances, effective long-term antiretroviral therapy has been plagued by various issues. The most concerning of all is the emergence of drug resistant viral strains against approved PIs. In addition, Patient tolerance and compliance continue to be critical problems. Daily therapeutic doses of these drugs are high because of low oral bioavailability. As a consequence, side effects such as peripheral lipodystrophy, hyperlipidemia, and insulin resistance have been of concern [4]. Because of these multiple problems, there is a critical need for the development of a new generation of PIs with improved pharmacokinetic properties and drug-resistance profiles. In this regard, comparing the resistance profiles of the present anti HIV protease drugs and correlating them with their chemical structures is important.

Since the Food and Drug Administration (FDA) approval of the first protease inhibitor (PI) in 1995, several other PIs quickly followed [5]. With the exception of tipranavir, all

commercial inhibitors are more or less substrate analogues where the scissile P1-P1' amide bond has been replaced with a hydroxyethylamine (HEA) or a hydroxyethylene (HE) transition state isostere. However, peptidic inhibitors lack appropriate physical chemical properties and metabolic stability to be ideally suited as therapeutics. They display poor bioavailability and are rapidly metabolized *in vivo*. Systematic modification of the substrate-based peptides using rational design techniques led to peptidomimetic compounds which abandon, as much as possible, the peptidic backbone while retaining the essential functionalities for HIV protease active site inhibition. The X-ray structures of various peptidomimetic inhibitors bound to HIV-1 protease have provided a wealth of information regarding ligand-binding site interactions. Saquinavir (SQV) was the first drug to be released in market by Hoffman-La-Roche [6]. Its initial pentapeptide lead was based on the HIV-1 pol substrate sequence containing the unusual Phe-Pro amide bond at the cleavage site. Lead optimization, including replacement of the P1-P1' amide bond with non-cleavable hydroxyethylamine based dipeptide isostere, replacement of the P1' proline with a bicyclic decahydroisoquinoline, and introduction of a quinoline moiety at P3 led to the discovery of SQV. Saquinavir contains four amide bonds and has a molecular size of 679 Da. It is an exceedingly potent inhibitor ($K_i = 0.12$ nM); however, its oral bioavailability is rather poor, possibly because of the presence of multiple amide/peptide like bonds. Nelfinavir (NFV) was developed by truncating the N-terminal moiety in SQV and replacing the P2 asparagine with 3-hydroxy-2-methylbenzamide fragment (Figure 6.1). These changes in combination with a novel S-phenyl moiety at P1 in the hydroxyethylamine isostere led to NFV ($K_i = 2$ nM) with significantly reduced

molecular weight and improvement in bioavailability, though NFV is less potent than SQV [7].

D30N is a major drug resistant mutation against NFV. But this mutation does not affect SQV even though SQV and NFV have common functional groups (Figure 6.1). In fact, D30N mutation does not exhibit cross-resistance with other PIs [8, 9]. To explore the reason for this, structure of HIV-1 protease D30N mutant in complex with the drug SQV was solved. This structure was then compared with the D30N mutant- NFV complex.

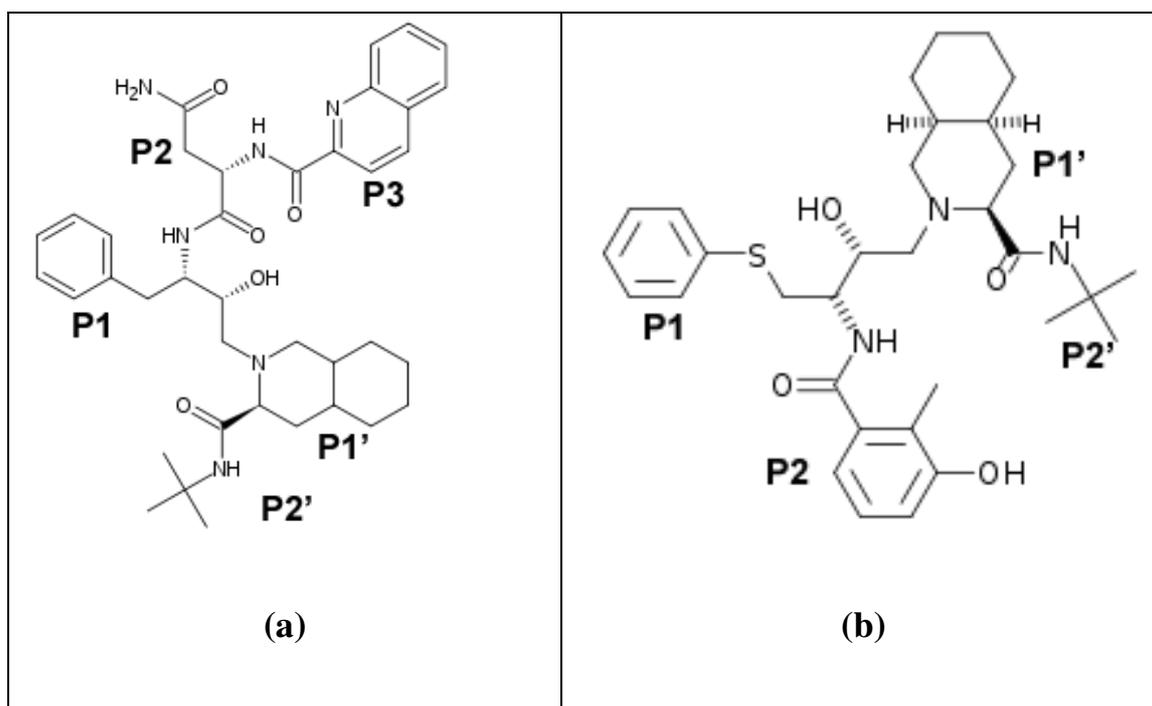


Figure 6.1: Chemical structures of (a) Saquinavir (SQV) (b) Nelfinavir (NFV)

6.2 Materials and methods:

6.2.1 Site directed mutagenesis:

Protocol explained in chapter 4 of this thesis with Oligonucleotide primer 5'-CTCCTCCAGTACAGTATTATCAGCACCGGTATCCAG-3' was used for making D30N mutation. Presence of mutation was confirmed through DNA sequencing.

6.2.2 Protein expression and purification:

HIV-1 protease tethered dimer used in the present study contains a five residue linker, GGSSG, covalently linking the two monomers. Expression, and purification of HIV-1 protease tethered dimer followed the procedures reported earlier [10] and explained in chapter 3 of this thesis.

6.2.3 Crystallization:

Single crystals were obtained by the hanging drop vapour diffusion method. Protein (5 mg/ml in 50mM sodium acetate, pH 4.5, containing 1 mM dithiothreitol) was reacted at room temperature for 30 minutes with ten-fold molar excess of SQV dissolved in dimethyl sulfoxide. For crystallization, equal volumes of reaction mixture and reservoir solution (1% saturated ammonium sulfate, 200 mM sodium phosphate, and 100 mM sodium citrate at pH 6.2) were mixed on a cover slip and sealed over the reservoir well at room temperature.

6.2.4 X-ray data collection and refinement:

The crystals were equilibrated in the cryo-protectant (25% glycerol and 75% reservoir buffer) before flash freezing, for exposure to X-rays on the FIP-BM30A beam line [11]. The diffraction data were indexed, integrated, and scaled by using the computer program XDS [12]. Crystal and intensity data statistics are given in Table 6.1.

Crystal structure was solved by difference Fourier method using native coordinates (PDB Id 1LV1). The structure was refined in computer programme PHENIX by using the amplitude-based maximum likelihood target function [13]. A test set containing 5.0% of randomly chosen reflections were reserved for determination of R_{free} [14]. All interactive model building and molecular superpositions were carried out using the molecular modelling software O [15]. Structural comparisons are based on superpositions of protein $C\alpha$ atoms. The quality of model was checked using programme PROCHECK [16]. All figures were drawn using computer programme PYMOL [17].

6.3 Results:

The crystal structure of D30N mutant of HIV-1 protease complexed with SQV was determined to a resolution of 1.8Å in the space group $P6_1$. Results of crystallographic refinement are summarized in Table 6.1. The crystallographic asymmetric unit contains a tethered dimer of HIV-1 protease. The residues in the first monomer are numbered as 1–99 and those in the second monomer are numbered 1001–1099. When $2mF_o - DF_c$ electron density map was contoured at 1σ level, no breaks were observed in the main chain. There was no density, however, for the linker region of the tethered dimer under study, suggesting that linker region is not ordered. The stereochemistry of the model was

quite good with more than 90% of non-glycine residues occupying the most favored regions of Ramachandran plot. In the difference electron density maps, there was good electron density for the SQV molecule. This difference density could be rationalized by SQV modeled in two orientations consistent with the 2-fold symmetry of functional HIV-1 protease dimer (Figure 6.2). Difference maps also indicated that residues 84 and 1084 were present in two alternate conformations (Figure 6.3).

Table 6.1: Data collection and refinement statistics

Space group	P6 ₁
Unit cell parameters (Å)	a= b = 62.84,c= 82.49
Resolution (Å)	32.8-1.8 (1.9-1.8)*
Number of unique reflections	17145 (2819)*
I/σ(I)	16.6 (2.2)*
R_{merge} (%)	4.1 (49.7)*
Completeness (%)	99.13 (98)*
<u>Refinement statistics</u>	
R _{work} / R _{free} (%)	20.2 / 23.4 (27.7/33.4)*
<u>RMS deviations from ideal values</u>	
Bond lengths (Å)	0.01
Bond angles (°)	1.1
<u>Ramachandran plot: non glycine residues in</u>	
Most favoured region (%)	96.2
Additionally allowed region (%)	3.8
Generously allowed region (%)	0.0
Disallowed region (%)	0.0

*Data for highest resolution shell are given in the parenthesis.

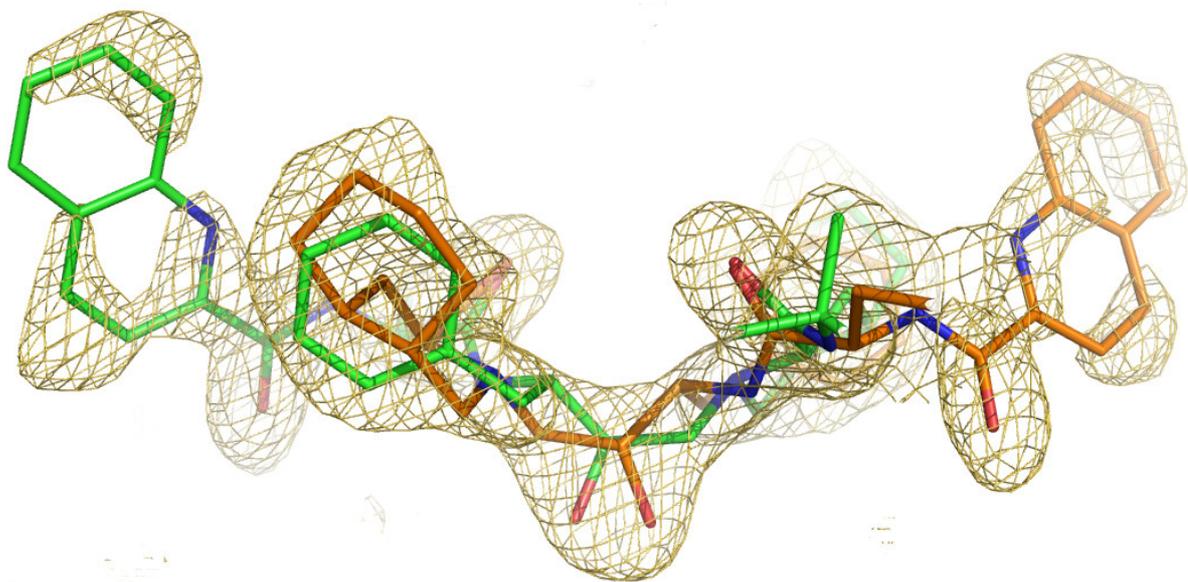


Figure 6.2: Fit of SQV in the F_O-F_C omit map: Simulated annealed F_O-F_C omit electron density map for SQV contoured at 2.0σ level.

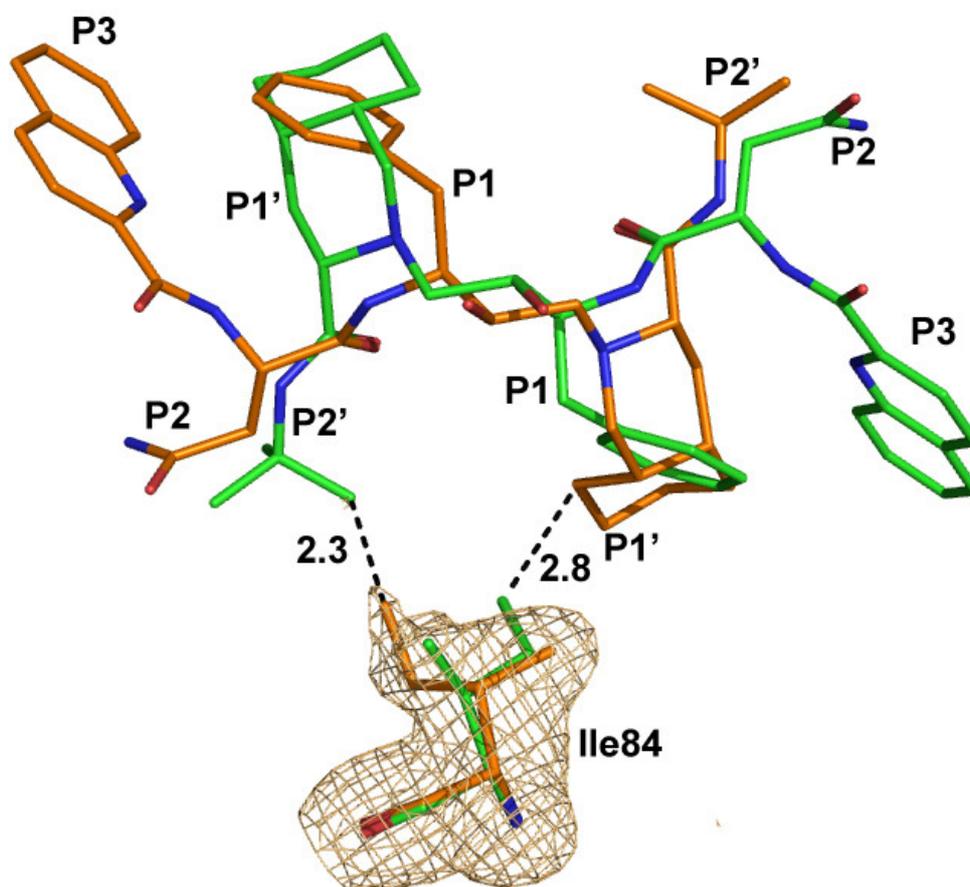


Figure 6.3: Alternate conformations of SQV and Ile84: Simulated annealed F_O-F_C omit map, contoured at 2.0σ level, showing alternate conformations of Ile84 corresponding with alternate conformations of SQV to avoid steric clashes. Distances are indicated in Å units.

6.3.1 Drug-Mutant interactions:

SQV binds to the mutant enzyme in an extended conformation through hydrogen bonds between backbone atoms of the drug and the protein. The side chain groups P3-P2' of the drug interact with residues in corresponding sub-sites S3-S2' mainly through hydrophobic interactions. The tert-butylcarboxamide moiety occupies the S2' subsite of

HIV-1 protease, the decahydroisoquinoline ring system fits into the hydrophobic S1' pocket, while the central hydroxyl group binds to the catalytic aspartates of the enzyme through hydrogen bonds. The P1 phenyl group resides in the S1 site. The polar asparagine side chain of the inhibitor occupies the S2 pocket. A tightly bound water molecule serves to relay hydrogen bonds from the two carbonyls of the inhibitor to Ile50/1050 main chain amides in the flap region of the enzyme in a manner analogous to that observed in other HIV-1 protease inhibitor complexes. The Ile84 side chain was located opposite to the inhibitor, and 84 CD1 forms van der Waals contacts with P1 Phe and P2' tert-butylcarboxamide residue of SQV. In the second orientation of the inhibitor, P1' decahydroisoquinoline ring is located opposite to the Ile84, and to optimize interactions with this, Ile84 adopts an alternate conformation (Figure 6.3). Steric contacts with the two fold related conformations of the inhibitor are avoided by Ile 84/1084 adopting the alternate conformations as shown in Figure 6.3. Residues 84/1084 showed alternate positions with the occupancy distribution of 0.5 each consistent with the two fold related conformations of the inhibitor.

6.3.2 Comparison with D30N-NFV complex:

The present complex was compared with complex of NFV with D30N mutant of HIV-1 protease (PDB Id 2Q64). The two structures are very similar with an rmsd for 198 C α atom pairs of 0.2 Å. Relative positions of SQV and NFV in the active site are shown in Figure 6.4. Chemically SQV and NFV are different on the non-primed side chain groups (Figure 6.1). At P2 site, there is an asparagine residue in SQV while NFV has 2-methyl-3-hydroxybenzamide moiety at this position. P2-asparagine of SQV forms a hydrogen bond with the Asp30 NH of the protease backbone. In case of NFV, 2-methyl-3-

hydroxybenzamide portion of the inhibitor occupies the S2 pocket, with the o-methyl substituent making hydrophobic interactions with Val32 and Ile84, and the m-phenol group hydrogen bonding to the Asp30 side chain carboxylate (Figure 6.5). At P1, slightly longer *S*-phenyl side chain of NFV partially extends into the S3 region and forms van der Waals contacts with Arg8. But in case of SQV, the bulky quinoline ring at P3 position would be in steric contact with Arg8/1008, the inter-atomic separation being only 2Å (Figure 6.4). This possibly has been avoided by a change in the conformation of the side chain of Arg8/1008. Comparison of torsion angles of Arg8 in the two complexes is given in table 6.2. The corresponding separation is now 3.4 Å. Carbonyl oxygen of the P3-quinaldic amide forms a hydrogen bond with backbone nitrogen of Asp29. There is a change in the main chain torsion angles of residue Gly48 of the protein to avoid steric contacts with the P3 quinoline ring of SQV (Figure 6.4).

Table6.2: Comparison of conformation of Arg8 in D30N-NFV and D30N-SQV Complexes

Conformation of Arg8	D30N-NFV Complex	D30N-SQV Complex
χ_1	-70°	-69°
χ_2	-177°	179°
χ_3	174°	-170°
χ_4	-144°	160°

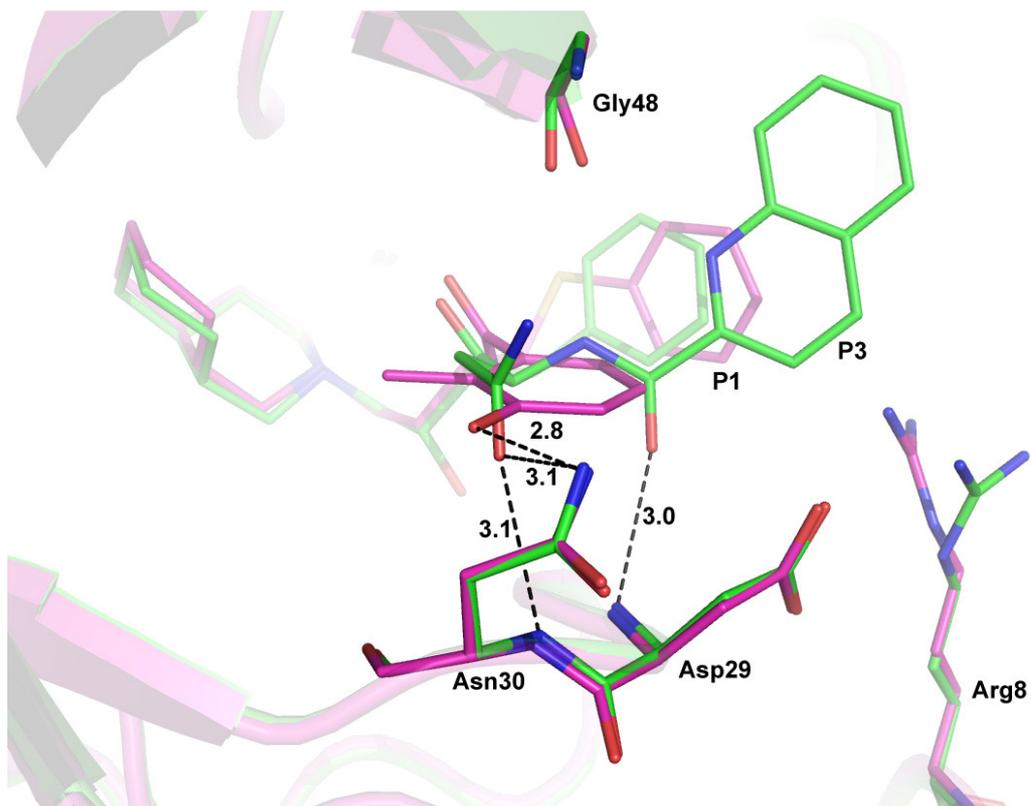


Figure 6.4: Structural comparison of the present complex with the D30N-NFV complex in the active site: Present complex (green), D30N-NFV complex (magenta). Note the change in conformation of Arg8 to avoid steric clash. Distances are indicated in Å units.

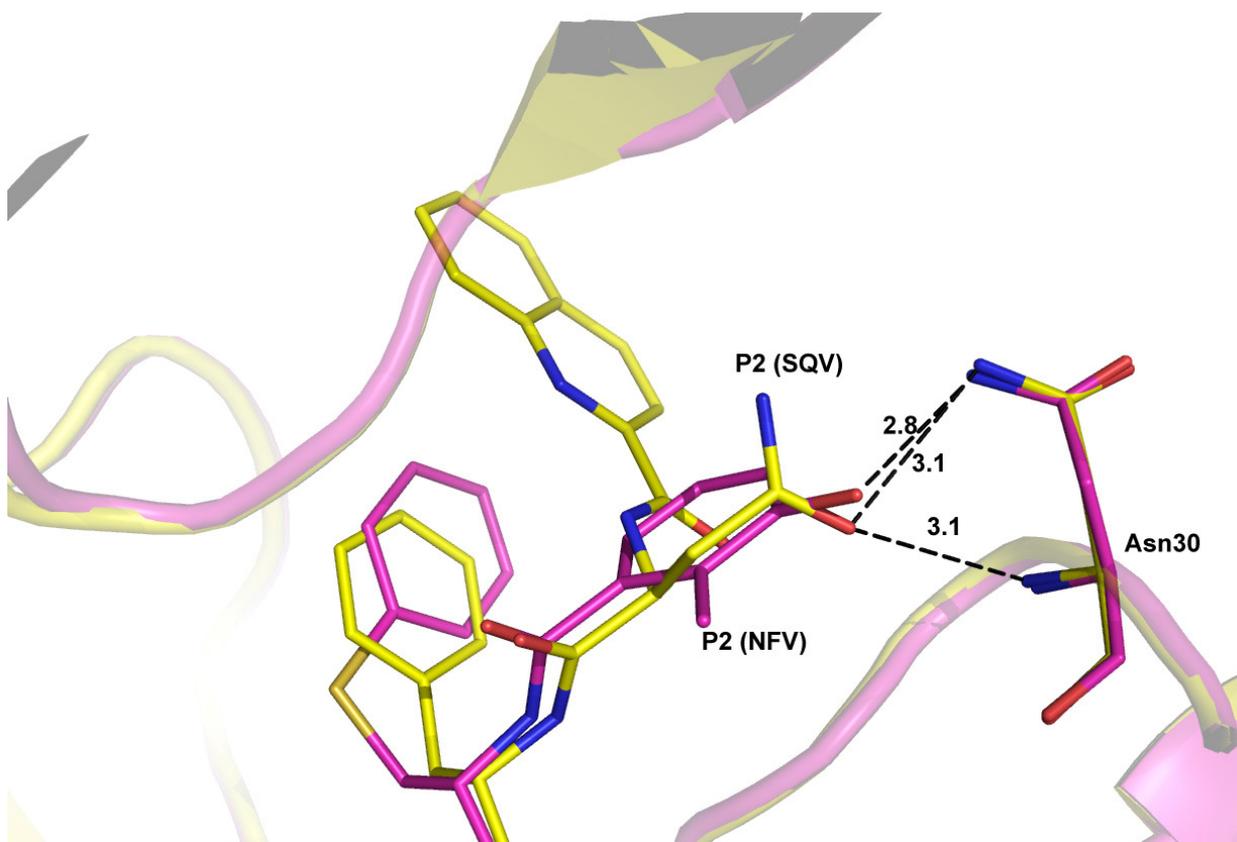


Figure 6.5: Comparison of hydrogen bonding interactions of P2 residues in SQV and NFV in D30N mutants: D30N-SQV complex (yellow) and D30N-NFV complex (magenta). Distances are indicated in Å units.

6.3.3 Comparison with native HIV-1 protease-SQV complex:

Figure 6.6 shows the structural superposition of the present SQV complex with D30N mutant and that with wild type HIV-1 protease-SQV complex (PDB Id 1HXB). The two structures superpose to an rmsd of 0.2 Å over 198 C α atom pairs. When compared to the native HIV-1 protease-SQV complex, interesting changes were observed in the conformation of the drug as well as of the protein. There is a slight rotation around χ_2 from -49° to -67° in Asn30 to form a hydrogen bond with the side chain of Asn residue at P2 position of SQV. Side chain of P2 Asn of SQV also adjusts to form hydrogen bonds with both main chain and side chain of Asn30 (Figure 6.6). In this position there are two hydrogen bonds which are not observed in the native complex.

6.4 Discussion:

Generally, the design of first generation HIV PIs included efforts to minimize the inhibitor molecular weight, and reduce its peptidic character while still imitating the interactions natural substrate made with the enzyme. NFV resulted from the optimization of SQV which suffers from poor oral bioavailability, presumably due in part to its retained peptide character. However, resistance to NFV arises with the selection of D30N mutation in HIV-1 protease. The substitution of asparagine for aspartic acid at position 30 is very specific to NFV resistant protease. Thus, the evolutionary pathway leading to NFV resistance has a rather low genetic barrier. The crystal structures of HIV-1 protease complexed with peptide substrates showed that the Asp30 residue main chain is involved in hydrogen bonding to the P2/P2' side chains of various substrates [18, 19]. This interaction is conserved whenever a polar side chain is present at the P2/P2' position. This is the only direct hydrogen bonding interaction seen between Asp30 side

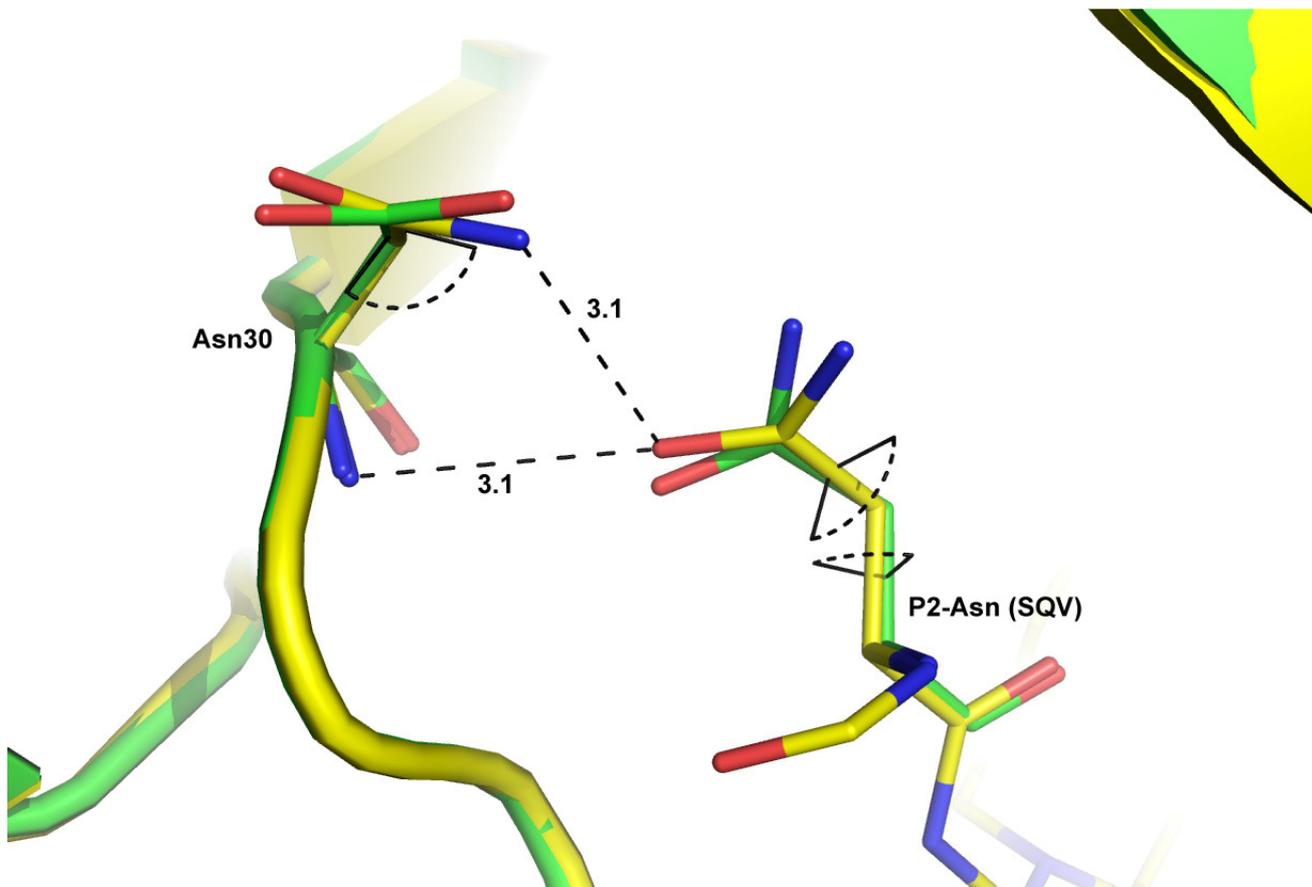


Figure 6.6: Structural effects around D30N mutation site: Wild type complex is shown in green carbons while D30N-SQV complex is shown in yellow carbons. Distances are indicated in Å units.

chain of the enzyme and P2 side chain of NFV, and therefore D30N mutation in the protease likely results in a weaker hydrogen bond that destabilizes NFV binding [20]. On the contrary, SQV forms two direct hydrogen bonds with the main chain atoms of the residues Asp29 (Asp29-NH...P3-CO= 3.0 Å) and Asp30 (Asp30 NH...P2-Asn Oδ1= 3.1 Å) (Figure 6.4). When Asp30 is mutated to Asn30, side chain Asn residue at P2 position can accept additional hydrogen bond (Asn30 Nδ1....P2-Asn Oδ1 = 3.1 Å) from the mutated side chain of the enzyme (Figure 6.6).

The significant residues at the floor of the active site for binding to the inhibitors are Asp25, Gly27, Ala28, Asp29, and Asp30. Among these residues, Asp25, Gly27, Ala28, and Asp29 are well conserved. The position of the backbone of the protease is less sensitive to mutations, and interactions between a potential inhibitor and backbone atoms of the enzyme is therefore of importance to maintain potency against multi-drug-resistant HIV-1 protease. The X-ray structure of saquinavir bound HIV-1 protease revealed that the P2-asparagine carbonyl forms a hydrogen bond with the Asp30 NH of the protease backbone. Furthermore, the carbonyl oxygen of the P3-quinaldic amide forms a hydrogen bond with main chain nitrogen of Asp29 (Figure 6.4). Importance of this information was realized in the development amprenavir and darunavir [21-23] and this should also be taken under consideration for designing new drugs.

Chapter 7

Summary and future directions

Human Immunodeficiency Virus (HIV) is the causative agent of Acquired Immunodeficiency Syndrome (AIDS). One of the enzymes of this virus, known as HIV protease, cleaves the viral gag and gag-pol polypeptide precursors to release the structural and functional proteins. Blocking the activity of HIV protease leads to immature and non-infectious viral particles. HIV protease is thus target of intensive efforts to design inhibitors against it. Till now, US Food and Drug Administration (FDA) have approved the release of 9 drugs for the treatment of HIV/AIDS which target HIV protease. HIV protease inhibitors are in fact one of the important components of Highly Active Anti Retroviral Therapy (HAART) which is a standard treatment regimen followed for HIV/AIDS. However, emergence of resistant mutations in HIV protease under the selection pressure of these drugs limits long term clinical efficacy of the drugs. This necessitates the need for continuous improvement of existing drugs and design of new HIV protease inhibitors based on the understanding of the interactions of HIV protease with its inhibitors as well as with substrates/products.

HIV protease is a homodimeric aspartyl protease with each monomer consisting of 99 amino acid residues. The enzyme active site is situated at the dimer interface. While the two catalytic aspartates are located on the floor of the active site cavity, two double stranded beta-ribbon structures, one coming from each subunit, form roof of the cavity. The flaps are conformationally very flexible segments of the protein and become ordered into a closed conformation whenever a substrate or inhibitor is bound in the active site.

HIV-1 protease tethered dimer used in the present study contains a five residue linker of sequence Gly-Gly-Ser-Ser-Gly, covalently linking the two monomers. Residues in the first monomer are numbered 1-99 while those in the second monomer, 1001-1099. The cloned insert also contains 57 extra codons in the beginning, which is a part of N-terminal polyprotein of pol gene. Therefore, the insert codes for a 29 kDa precursor protein, containing natural cleavage site for HIV-1 protease, which after self-cleavage results in a mature protein of 22 kDa. The drug resistant mutations were carried out using Quick change site directed mutagenesis kit (Stratagene, La Jolla, CA). Presence of mutations was confirmed through DNA sequencing. Mid transition temperatures (T_m) of native and mutant proteins were determined using Circular Dichroism (CD) spectrometry. Kinetic constants of native and mutant enzymes were determined by spectrophotometric enzyme assay. Single crystals of tethered HIV-1 protease and mutant-drug complexes were obtained by hanging drop vapour diffusion method. The crystals were equilibrated in a cryoprotectant solution before flash freezing, for exposure to X-rays on beamlines at European Synchrotron Radiation Facility (ESRF) and Swiss Light Source (SLS). Diffraction data were collected as a number of oscillation frames. The diffraction data were processed and scaled by using the computer programmes XDS, MOSFLM and SCALA. Crystal structures were solved by difference Fourier method using native coordinates. Graphics Programmes O and COOT were used for electron density interpretation and model building. The structures were refined using computer programmes CNS, REFMAC and PHENIX. The quality of models was checked using computer programme PROCHECK.

Discovery in our laboratory of closed-flap conformation of the enzyme in hexagonal crystals of HIV-1 protease even when the enzyme is unliganded, made possible the crystallographic studies on active HIV-1 protease complexed with different substrate peptides. In the present work, native crystals of active tethered HIV-1 protease were soaked in an undecapeptide substrate solution at pH 7.0. Three dimensional crystal structure, determined to 1.69 Å resolution shows that the carboxyl terminal product (Q product) peptide generated within the crystal is still bound in the active site of HIV-1 protease along with a set of water molecules. This is the first observation of the coexistence in the active site, of a water molecule WAT1, along with the carboxyl terminal product peptide. WAT1, which is activated through hydrogen bonds to catalytic aspartates, is located at a distance of 2.7 Å along the direction perpendicular to the scissile peptide bond. Assuming the present structure to be a close approximation to Michaelis complex, we propose that the incoming substrate pushes the nucleophilic water from the position observed in the unliganded protease to the WAT1 position from where it attacks the scissile peptide bond. Once the Q product also diffuses out, the catalytic water molecule can move back to the position observed in unliganded structures of HIV-1 protease. Comparison of geometries at the catalytic centre shows systematic changes in the conformation and interactions of catalytic aspartates at pHs 2.0 and 7.0. The structure reported here also suggests that in the design of effective inhibitors of HIV-1 protease, it is important to duplicate the hydrogen-bonding pattern of WAT1 with catalytic aspartates.

Drug resistance is often associated with mutations of HIV protease in both active site and non-active site regions. Effects of active site mutations can be explained on the basis of

altered interactions with the inhibitors but an explanation for non-active site mutations has largely been elusive. Saquinavir (SQV) was the first HIV protease inhibitor to be released in the market for chemotherapy against HIV/AIDS. It is a potent inhibitor of HIV protease and used as a first line drug. G48V is a major active site mutation against the drug SQV which occurs in combination with the mutation C95F. Crystal structure of the complex between the drug SQV and the corresponding drug-resistant G48V/C95F mutant tethered HIV-1 protease was solved and refined to 2.5 Å resolution, and the refined molecular structure allows inferences to be drawn about the mechanisms of drug-resistance by these mutations. The active site mutation G48V causes loss in attractive van der Waals interactions between the drug and the enzyme. The non-active site C95F mutation reduces structural flexibility of the catalytic loop and also the dimer stability by disrupting inter-monomer hydrogen bonds, and this could be another source of resistance. The crystal structure can rationalize the clinical observation that resistance mutations occur as a mutation-cluster C95F/G48V/V82A in AIDS patients treated with SQV.

L90M is a major non-active site mutation against the drug SQV which occurs in combination with the mutation C95F. The structures of L90M/C95F and C95F mutants of tethered HIV-1 protease complexed with the drug SQV are described. The crystal structures of L90M/C95F_SQV and C95F_SQV complexes were solved to resolutions of 2.04 and 1.75 Å, respectively. The catalytic efficiencies (k_{cat}/K_m), of L90M/C95F and C95F mutants were about 67% and 50% respectively, w.r.t. wild type protease. Comparisons of SQV bound L90M/C95F mutant structure with that of C95F_SQV complex and native enzyme-SQV complex have shown that L90M mutation can perturb packing interactions within the core of enzyme and these perturbations are carried to the

active site leading to its reshaping. Since all anti HIV protease drugs including SQV are competitive inhibitors, any change in the shape of active site cavity may affect the binding of these drugs more than that of natural substrates. Thus non active site mutation L90M can affect the inhibitor binding through global, long range perturbations of the active site. C95F mutation leads to repacking in the inter subunit four stranded beta sheet, leading to loss of the inter-subunit hydrogen bonds between Pro1/1001 and Phe99/1099 in the mutant. All the mutants with C95F mutation have lower dimer stability as compared to wild type enzyme as shown by reduction in the mid transition temperatures (T_m) by about 6-8°C and C95F mutation seems to be major contributor. Changes in the packing at sub domain interfaces could also affect the energetics of sub domain movements associated with inhibitor-binding.

Although drug Nelfinavir (NFV) and SQV share common functional groups, D30N is a major mutation against the drug nelfinavir but does not provide resistance against the drug SQV. The structure of D30N-SQV complex was solved to 1.8 Å. Structural analysis showed that P2' Asn of SQV forms an additional hydrogen bond to the mutated side chain of residue 30 apart from a main chain hydrogen bond and this could be the reason why D30N is not a drug resistant mutation against SQV.

To conclude, structure of HIV-1 protease-product peptide complex reported in the present thesis gives an insight into the molecular mechanism of HIV-1 protease. The structure suggests repositioning of catalytic water, WAT1, during substrate binding for activation and subsequent nucleophilic attack. Structures of mutants in complex with SQV give insight into the mechanisms of drug resistance. Residues that undergo conformational changes leading to loss of van der Waals interactions, reshaping of active site cavity and

reduction in dimer stability through repacking at the inter domain regions have been identified. The reason why D30N mutation is not effective against SQV has been found out.

With respect to the current set of problems associated with the use of protease inhibitors (PIs), there is a critical need for the development of a new generation of PIs that exhibit improved pharmacokinetic properties and drug-resistance profiles. Four suggestions, which are listed below, can be made based on results reported in this thesis.

(1) Replicate interactions made by the water molecule, WAT1, in the structure reported in chapter 3: Mimicking the transition state geometry will remain crucial to the design of potent inhibitors. Position of WAT1 represents the location of one of the oxygens of tetrahedral transition state intermediate formed during substrate hydrolysis. Therefore, transition state mimics having a functional group at the position of WAT1 and interacting in a way similar to WAT1, will prove to be very potent for the reason that they will resemble the transition state more closely. KNI-272 is one such compound, the hydroxyl group of which superposes WAT1 perfectly, and it is also one of the most potent inhibitors. However, all hydrogen bonding interactions of WAT1 are not made by the hydroxyl group of KNI-272. There is thus scope of designing modified KNI-272 to add additional hydrogen bonds to further enhance its potency.

(2) Part of the inhibitor should mimic flap water: A common feature in the complexes between peptidomimetic inhibitors and HIV-1 protease is a conserved water molecule known as flap water, which is involved in four hydrogen bonds, bridging P2/P1' carbonyl oxygen atoms of the inhibitors to amide groups of Ile50/Ile1050 of the enzyme. In G48V/C95F SQV resistant mutant structure reported in chapter 4, due to

changes in flap conformation, the flap water is unable to make all the four hydrogen bonds. Hydrogen bonds to the flap amide groups have been lost while hydrogen bonds to the rigid carbonyl groups from the inhibitor are retained. This loss could be the major reason for decrease in SQV affinity. If this loss could be prevented by making flap water as part of a more flexible inhibitor, then the mutant protein will not be resistant to this modified-SQV inhibitor. Therefore, design of inhibitors which can bind directly with the flaps should be considered. Displacing this water may also lead to increased affinity from entropic point of view.

(3) Enhance interactions with the protein backbone atoms: Mutation D30N confers high degree of drug resistance against NFV but is ineffective against SQV. On comparison of complexes of these two drugs with wild type HIV-1 protease and D30N mutant, only one hydrogen bonding interaction between Asp30 side chain of the enzyme and P2 side chain of NFV was seen, and therefore D30N mutation in the protease likely results in a weaker hydrogen bond that destabilizes NFV binding. On the contrary as shown by the work reported in chapter 6, SQV forms two direct hydrogen bonds with the main chain atoms of the residues Asp29 and Asp30. These hydrogen bonds are present even when Asp30 is mutated to Asn30, as shown in chapter 6 of the thesis. The inference therefore is that inhibitors that interact with main chain atoms are less prone to be escaped by mutations in the enzyme, and interactions between a potential inhibitor and backbone atoms of the enzyme is therefore of importance to maintain potency against multi-drug-resistant HIV-1 protease. Main chain interactions between inhibitors and the active site residues of the enzyme should be maximized.

(4) Inclusion of flexible pharmacophores in the inhibitor design: In case of L90M/C95F mutant structure reported in chapter 5 of this thesis, the non-active site mutations have caused a change in the shape of active site cavity. Although, there were concerted shifts in the conformation of drug SQV to cope with the altered active site, all van der Waals interactions between the drug and the mutant enzyme could not be maintained. Therefore, flexible pharmacophores should be included in the inhibitor design so that they permit retention of all van der Waals interactions through appropriate change of conformation.

References

Chapter 1

1. Weiss A, Hollander H, Stobo J (1985) Acquired immunodeficiency syndrome: epidemiology, virology, and immunology. *Annu. Rev. Med.* 36: 545-562.
2. Gallo RC, Montagnier L (1988) AIDS in 1988. *Sci. Am.* 259:41-25948.
3. The Joint United Nations Program on HIV/AIDS (UNAIDS). Global report: UNAIDS Report on the Global AIDS Epidemic 2010; UNAIDS/10.11E/JC1958E; UNAIDS: Geneva, Switzerland, 2010.
4. Gao F, Bailes E, Robertson DL, Chen Y, Rodenburg CM, Michael SF, Cummins LB, Arthur LO, Peeters M, Shaw GM, Sharp PM, Hahn BH (1999) Origin of HIV-1 in the chimpanzee *Pan troglodytes troglodytes*. *Nature* 397: 436–441.
5. Van Heuverswyn F, Li Y, Neel C, Bailes E, Keele BF, Liu W, Loul S, Butel C, Liegeois F, Bienvenue Y, Ngolle EM, Sharp PM, Shaw GM, Delaporte E, Hahn BH, Peeters M (2006) Human immunodeficiency viruses: SIV infection in wild gorillas. *Nature* 444: 164.
6. Takehisa J, Kraus MH, Ayouba A, Bailes E, Van Heuverswyn F, Decker JM, Li Y, Rudicell RS, Learn GH, Neel C, Ngole EM, Shaw GM, Peeters M, Sharp PM, Hahn BH (2009) Origin and biology of simian immunodeficiency virus in wild-living western gorillas. *J. Virol.* 183: 1635–1648.
7. Ayouba A, Maucière P, Martin PM, Cunin P, Mfoupouendoun J, Njinku B, Souquières S, Simon F (2001) HIV-1 group O infection in Cameroon, 1986 to 1998. *Emerg. Infect. Dis.* 7: 466–467.

8. Brennan CA, Bodelle P, Coffey R, Devare SG, Golden A, Hackett J Jr., Harris B, Holzmayer V, Luk KC, Schochetman G, Swanson P, Yamaguchi J, Vallari A, Ndembu N, Ngansop C, Makamche F, Mbanya D, Gürtler LG, Zekeng L, Kaptué L (2008) The prevalence of diverse HIV-1 strains was stable in Cameroonian blood donors from 1996 to 2004. *J. Acquir. Immune Defic. Syndr.* 49: 432–439.
9. Simon F, Maucière P, Roques P, Loussert-Ajaka I, Müller-Trutwin MC, Saragosti S, Georges-Courbot MC, Barré-Sinoussi F, Brun-Vénizet F (1998) Identification of a new human immunodeficiency virus type 1 distinct from group M and group O. *Nat. Med.* 4:1032–1037.
10. Roques P, Robertson DL, Souquière S, Apetrei C, Nerrienet E, Barré-Sinoussi F, Müller-Trutwin M, Simon F (2004) Phylogenetic characteristics of three new HIV-1 N strains and implications for the origin of group N. *AIDS* 18: 1371–1381.
11. Ayouba A, Souquière S, Njinku B, Martin PM, Müller-Trutwin MC, Roques P, Barré-Sinoussi F, Maucière P, Simon F, Nerrienet E (2000) HIV-1 group N among HIV-1-seropositive individuals in Cameroon. *AIDS* 14: 2623–2625.
12. Bodelle P, Vallari A, Coffey R, McArthur CP, Beyeme M, Devare SG, Schochetman G, Brennan CA (2004) Identification and genomic sequence of an HIV type 1 group N isolate from Cameroon. *AIDS Res. Hum. Retroviruses* 20: 902–908.
13. Plantier JC, Leoz M, Dickerson JE, De Oliveira F, Cordonnier F, Lemée V, Damond F, Robertson DL, Simon FA (2009) New human immunodeficiency virus derived from gorillas. *Nat. Med.* 15: 871–872.
14. HIV sequence database. *The Human Retroviruses and AIDS 1994 Compendium.*

III23-III33.

Available

online:

[Http://www.hiv.lanl.gov/content/sequence/HIV/COMPENDIUM/](http://www.hiv.lanl.gov/content/sequence/HIV/COMPENDIUM/)

94compendium.html.

15. Salminen MO, Koch C, Sanders-Buell E, Ehrenberg PK, Michael NL, Carr JK, Burke DS, McCutchan FE (1995) Recovery of virtually full-length HIV-1 provirus of diverse subtypes from primary virus cultures using polymerase chain reaction. *Virology* 213:80–6.
16. Carr JK, Salminen MO, Albert J, Sanders-Buell E, Gotte D, Birx DL, McCutchan FE (1998) Full-length genome sequences of human immunodeficiency virus type 1 subtypes G and A/G intersubtype recombinants. *Virology* 247:22–31.
17. McCutchan FE (2000) Understanding the genetic diversity of HIV-1. *AIDS* 14: (Suppl 3): S31–44.
18. Robertson DL, Anderson JP, Bradac JA, Carr JK, Foley B, Funkhouser RK, Gao F, Hahn BH, Kalish ML, Kuiken C, Learn GH, Leitner T, McCutchan F, Osmanov S, Peeters M, Pieniazek D, Salminen M, Sharp PM, Wolinsky S, Korber B (2000) HIV-1 nomenclature proposal. *Science* 288: 55–56.
19. Gao F, Vidal N, Li Y, Trask SA, Chen Y, Kostrikis LG, Ho DD, Kim J, Oh MD, Choe K, Salminen M, Robertson DL, Shaw GM, Hahn BH, Peeters M (2001) Evidence of two distinct subsubtypes within the HIV-1 subtype A radiation. *AIDS Res. Hum. Retroviruses* 17: 675–688.
20. Meloni ST, Kim B, Sankalé JL, Hamel DJ, Tovanabutra S, Mboup S, McCutchan FE, Kanki PJ (2004) Distinct human immunodeficiency virus type 1 subtype A virus circulating in West Africa: sub-subtype A3. *J. Virol.* 78: 12438–12445.

21. Triques K, Bourgeois A, Saragosti S, Vidal N, Mpoudi-Ngole E, Nzilambi N, Apetrei C, Ekwilanga M, Delaporte E, Peeters M (1999) High diversity of HIV-1 subtype F strains in Central Africa. *Virology* 259: 99–109.
22. Greene WC (1993) AIDS and the immune-system. *Scientific American* 269: 98-105.
23. Navia MA, Fitzgerald PM, McKeever BM, Leu CT, Heimbach JC, Herber WK, Sigal IS, Darke PL, Springer JP (1989) Three-dimensional structure of aspartyl protease from human immunodeficiency virus HIV-1. *Nature* 337: 615-620.
24. Stebbins J, Debouck C (1994) Expression systems for retroviral proteases. *Methods Enzymol.* 241: 3-15.
25. Cheng Y-SE, Yin FH, Foundling S, Blomstrom D, Kettner CA (1990) Stability and activity of human immunodeficiency virus protease: comparison of the natural dimer with a homologous, single chain tethered dimer. *Proc. Natl. Acad. Sci. USA* 87: 9660–9664.
26. Miller M, Schneider J, Sathyanarayana BK, Toth MV, Marshall GR, Clawson L, Selk L, Kent SBH, Wlodawer A (1989b) Structure of complex of synthetic HIV-1 protease with a substrate-based inhibitor at 2.3 Å resolution. *Science* 246: 1149-1152.
27. Gustchina A, Weber IT (1990) Comparison of inhibitor binding in HIV-1 protease and in non-viral aspartic proteases: the role of the flap. *FEBS Lett.* 269: 269-272.
28. Pettit SC, Moody MD, Wehbie RS, Kaplan AH, Nantermet PV, Klein CA, Swanstrom R (1993) The p2 domain of human immunodeficiency virus type 1 Gag regulates sequential proteolytic processing and is required to produce fully

- infectious virions. *J. Virol.* 68: 8017-8027.
29. Schechter I, Berger A (1967) On the size of the active site in proteases. I. Papain. *Biochem. Biophys. Res. Commun.* 27: 157-162.
30. Tözsér J, Weber IT, Gustchina A, Blaha I, Copeland TD, Louis JM, Oroszlan S (1992) Kinetic and modeling studies of S3-S3' subsites of HIV proteinases. *Biochemistry* 31: 4793-4800.
31. Dunn BM, Gustchina A, Wlodawer A, Kay J (1994) Subsite preferences of retroviral proteinases. *Methods Enzymol.* 241: 254-278.
32. Wlodawer A, Vondrasek J (1998) Inhibitors of HIV-1 protease: A major success of structure-assisted drug design. *Annu. Rev. Biophys. Biomol. Struct.* 27: 249–284.
33. Tözsér J, Gustchina A, Weber IT, Blaha I, Wondrak EM, Oroszlan S (1991) Studies on the role of the S4 substrate binding site of HIV proteinases. *FEBS Lett.* 279: 356-360.
34. Vondrasek J, van Buskirk CP, Wlodawer A (1997) Database of three dimensional structures of HIV proteinases. *Nat. Struct. Biol.* 4: 8.
35. Fitzgerald PM, Springer JP (1991) Structure and function of retroviral proteases. *Annu. Rev. Biophys. Biophys. Chem.* 20: 299-320.
36. Moore ML, Dreyer GB (1993) Substrate-based inhibitors of HIV-1 protease. *Prospect. Drug Discovery Des.* 1: 85-108.
37. Henderson LE, Benveniste RE, Sowder R, Copeland TD, Schultz AM, Oroszlan S (1988) Molecular characterization of gag proteins from simian immunodeficiency virus (SIVMne). *J. Virol.* 62: 2587-2595.

38. Pettit SC, Simsic J, Loeb DD, Everitt L, Hutchison CA, Swanstrom R (1991) Analysis of retroviral protease cleavage sites reveals two types of cleavage sites and the structural requirements of the P1 amino acid. *J. Biol. Chem.* 266: 14539-14547.
39. Griffiths JT, Phylip LH, Konvalinka J, Strop P, Gustchina A, Wlodawer A, Davenport RJ, Briggs R, Dunn BM, Kay J (1992) Different requirements for productive interaction between the active site of HIV-1 proteinase and substrates containing -hydrophobic*hydrophobic- or - aromatic*pro- cleavage sites. *Biochemistry* 31: 5193-51200.
40. Tözsér J, Bagossi P, Weber IT, Louis JM, Copeland TD, Oroszlan S (1997) Studies on the symmetry and sequence context dependence of the HIV-1 proteinase specificity. *J. Biol. Chem.* 272: 16807-16814.
41. Oroszlan S, Henderson LE, Stephenson JR, Copeland TD, Long CW, Ihle JN, Gilden RV (1978) Amino- and carboxyl-terminal amino acid sequences of proteins coded by gag gene of murine leukemia virus. *Proc. Natl. Acad. Sci. USA* 75: 1404-1408.
42. (a) Cairns J, Overbaugh J, Miller S (1988) The origin of mutants. *Nature* 335:142-145
- (b) Stahl FW (1988) Bacterial genetics. A unicorn in the garden. *Nature* 335: 112-113.
43. Darke PL, Leu CT, Davis LJ, Heimbach JC, Diehl RE, Hill WS, Dixon RA, Sigal IS (1989) Human immunodeficiency virus protease. Bacterial expression and characterization of the purified aspartic protease. *J. Biol. Chem.* 264: 2307-2312.

44. Seelmeir S, Schmidt H, Turk V, Von Der Helm K (1988) Human Immunodeficiency virus has an aspartic-type protease that can be inhibited by pepstatin A. *Proc. Natl. Acad. Sci. USA* 85: 6612-6616.
45. Mous J, Heimer EP, Le Grice SJJ (1988) Processing protease and reverse transcriptase from human immunodeficiency virus type I polyprotein in *Escherichia coli*. *J. Virol.* 62: 1433-1436.
46. Hansen J, Billich S, Schulze T, Sukrow S, Moelling K (1998) Partial purification and substrate analysis of bacterially expressed HIV protease by means of monoclonal antibody. *EMBO J.* 7: 1785-1791.
47. Wlodawer A, Miller M, Jaskolski M, Sathyanarayana BK, Baldwin E, Weber IT, Selk LM, Clawson L, Schneider J, Kent SBH (1989) Conserved folding in retroviral proteases: crystal structure of a synthetic HIV-1 protease. *Science* 245: 616-621.
48. Lapatto R, Blundell T, Hemmings A, Overington J, Wilderspin A, Wood S, Merson JR, Whittle PJ, Danley DE, Geoghegan KF, Hawrylik SJ, Lee SE, Scheld KG, Hobart PM (1989) X-ray analysis of HIV-1 proteinase at 2.7 Å resolution confirms structural homology among retroviral enzymes. *Nature* 342: 299-302.
49. Rao JKM, Erickson JW, Wlodawer A (1991) Structural and evolutionary relationships between retroviral and eucaryotic aspartic proteinases. *Biochemistry* 30: 4663-4671.
50. Chatfield DC, Brooks BR (1995) HIV-1 protease cleavage mechanism elucidated with molecular dynamics simulation. *J. Am. Chem. Soc.* 117: 5561-5572.
51. Okimoto N, Tsukui T, Hata M, Hoshino T, Tsuda M (1999) Hydrolysis

- mechanism of the phenylalanine-proline peptide bond specific to HIV-1 protease: Investigation by the ab initio Molecular Orbital Method. *J. Am. Chem. Soc.* 121: 7349–7354.
52. Piana S, Bucher D, Carloni P, Rothlisberger U (2004) Reaction mechanism of HIV-1 protease by hybrid car-parrinello/classical md simulations. *J. Phys. Chem.* 108: 11139–11149.
53. Trylska J, Grochowski P, McCammon JA (2004) The role of hydrogen bonding in the enzymatic reaction catalyzed by HIV-1 protease. *Protein Sci.* 13: 513–528.
54. Hyland LJ, Tomaszek TA Jr, Meek TD (1991) Human immunodeficiency virus-1 protease. 2. Use of pH rate studies and solvent kinetic isotope effects to elucidate details of chemical mechanism. *Biochemistry* 30: 8454–8463.
55. Hyland LJ, Tomaszek TA Jr, Roberts GD, Carr SA, Magaard VW, Bryan HL, Fakhoury SA, Moore ML, Minnich MD, Culp JS, Desjarlais RL, Meek TD (1991) Human Immunodeficiency Virus-1 Protease. 1. Initial velocity studies and kinetic characterization of reaction intermediates by ^{18}O isotope exchange. *Biochemistry* 30: 8441–8453.
56. Northrop DB (2001) Follow the protons: a low-barrier hydrogen bond unifies the mechanisms of the aspartic proteases. *Acc. Chem. Res.* 34: 790–797.
57. Jaskolski M, Tomasselli AG, Sawyer TK, Staples DG, Heinrikson RL, Schneider J, Kent SB, Wlodawer A (1991) Structure at 2.5 Å resolution of chemically synthesized human immunodeficiency virus type 1 protease complexed with a hydroxyethylene based inhibitor. *Biochemistry* 30: 1600–1609.
58. Rose RB, Craik CS, Douglas NL, Stroud RM (1996) Three-dimensional

- structures of HIV-1 and SIV protease product complexes. *Biochemistry* 35: 12933–12944.
59. Silva AM, Cachau RE, Sham HL, Erickson JW (1996) Inhibition and catalytic mechanism of HIV-1 aspartic protease. *J. Mol. Biol.* 255: 321–346.
60. Kovalevsky AY, Chumanevich AA, Liu F, Louis JM, Weber IT (2007) Caught in the Act: The 1.5 Å Resolution Crystal Structures of the HIV-1 protease and the I54V Mutant Reveal a Tetrahedral Reaction Intermediate. *Biochemistry* 46: 14854–14864.
61. Johnson EB, Malito E, Shen Y, Pentelute B, Rich D, Florian J, Tang WJ, Kent SBH (2007) Insights from atomic-resolution X-ray structures of chemically-synthesized HIV-1 protease in complex with inhibitors. *J. Mol. Biol.* 373: 573–586.
62. Brik A, Wong CH (2003) HIV-1 protease: mechanism and drug discovery. *Org. Biomol. Chem.* 1: 5–14.
63. Dunn BM (2002) Structure and mechanism of the pepsin-like family of aspartic peptidases. *Chem. Rev.* 102: 4431–4458.
64. International AIDS Vaccine Initiative (IAVI). AIDS vaccine blueprint 2008: A challenge to the field, a roadmap for progress. *IAVI* (August 2008). (www.iavi.org/Lists/IAVIPublications/attachments/750bc28e-0e47-4a1e-91bf-eeea7c8bb98e/IAVI_AIDS_Vaccine_Blueprint_2008_ENG.pdf)
65. Kramer RA, Schaber MD, Skalka AM, Ganguly K, Wong-Staal F, Reddy EP (1986) HTLV-III *gag* protein is processed in yeast cells by the virus *pol* protein. *Science* 231: 1580-1585.

66. Kohl NE, Emini EA, Schleif WA, Davis LJ, Heimbach JC, Dixon RA, Scolnick EM, Sigal IS (1988) Active human immunodeficiency virus protease is required for viral infectivity. *Proc. Natl. Acad. Sci. U.S.A.* 85: 4686-4690.
67. Schatz H, Gelderbloom HR, Nitschko N, von der Helm K (1991) Analysis of non-infectious HIV-1 particles in the presence of HIV proteinase inhibitor. *Arch.Virol.* 120: 71-81.
68. Lambert DM, Petteway Jr. SR, McDanal CE, Hart TK, Leary JJ, Dreyer GB, Meek TD, Buigelski PJ, Bolognesi DP, Metcalf BW, Matthews TJ (1992) Human immunodeficiency virus type 1 protease inhibitors irreversibly block infectivity of purified virions from chronically infected cells. *Antimicrob.Agents Chemother.* 36: 982-988.
69. Miller M, Schneider J, Sathyanarayana BK, Toth MV, Marshall GR, Clawson L, Selk L, Kent SB, Wlodawer A (1989) Structure of complex of synthetic HIV-1 protease with a substrate-based inhibitor at 2.3 Å resolution. *Science* 246: 1149–1152.
70. Roberts NA, Martin JA, Kinchington D, Broadhurst AV, Craig JC, Duncan IB, Galpin SA, Handa BK, Kay J, Krohn A (1990) Rational design of peptide-based HIV proteinase inhibitors. *Science* 248: 358-361.
71. Mehandru S, Markowitz M (2003) Tipranavir: a novel non-peptidic protease inhibitor for the treatment of HIV infection. *Expert. Opin. Investig. Drugs* 12: 1821-1828.
72. Koh Y, Nakata H, Maeda K, Ogata H, Bilcer G, Devasamudram T, Kincaid JF, Boross P, Wang YF, Tie Y, Volarath P, Gaddis L, Harrison RW, Weber IT,

- Ghosh AK, Mitsuya H (2003) Novel bis-tetrahydrofuranylurethane-containing nonpeptidic protease inhibitor (PI) UIC-94017 (TMC114) with potent activity against multi-PI-resistant human immunodeficiency virus *in vitro*. *Antimicrob. Agents Chemother.* 47: 3123-3129.
73. Kempf DJ, Marsh KC, Denissen JF, McDonald E, Vasavanonda S, Flentge CA, Green BE, Fino L, Park CH, Kong XP, Wideburg NE, Saldivar A, Ruiz L, Kati WM, Sham HL, Robins T, Stewart KD, Hsu A, Plattner JJ, Leonard JM, Norbeck DW (1995) ABT-538 is a potent inhibitor of human immunodeficiency virus protease and has high oral bioavailability in humans. *Proc. Natl. Acad. Sci. U. S. A.* 92: 2484–2488.
74. Coffin, J.M. (1995) HIV population dynamics in vivo: implications for genetic variation, pathogenesis, and therapy. *Science* 267: 483-489.
75. Stanford HIV Drug Resistance Database. Available online: <http://hivdb.Stanford.edu>.
76. Wu TD, Schiffer CA, Gonzales MJ, Taylor J, Kantor R, Chou S, Israelski D, Zolopa AR, Fessel WJ, Shafer RW (2003) Mutation patterns and structural correlates in human immunodeficiency virus type 1 protease following different protease inhibitor treatments. *J. Virol.* 77: 4836–4847.
77. Rhee SY, Taylor J, Fessel WJ, Kaufman D, Towner W, Troia P, Ruane P, Hellinger J, Shirvani V, Zolopa A, Shafer RW (2010) HIV-1 protease mutations and protease inhibitor cross resistance. *Antimicrob. Agents Chemother.* 54: 4253–4261.
78. Velazquez-Campoy A, Vega S, Freire E (2002) Amplification of the effects of

- drug resistance mutations by background polymorphisms in HIV-1 protease from African subtypes. *Biochemistry* 41: 8613–8619.
79. Shafer RW, Schapiro JM (2008) HIV-1 drug resistance mutations: an updated framework for the second decade of HAART. *AIDS Rev.* 10: 67-84.
80. Nijhuis M, van Maarseveen NM, Lastere S, Schipper P, Coakley E, Glass B, Rovenska M, de Jong D, Chappey C, Goedegebuure IW, Heilek-Snyder G, Dulude D, Cammack N, Brakier-Gingras L, Konvalinka J, Parkin N, Krausslich HG, Brun-Vezinet F, Boucher CA (2007) A novel substrate-based HIV-1 protease inhibitor drug resistance mechanism. *PLoS. Med.* 4: e36.
81. Pereira-Vaz J, Duque V, Trindade L, Saraiva-da-Cunha J, Melico-Silvestre A (2009) Detection of the protease codon 35 amino acid insertion in sequences from treatment-naïve HIV-1 subtype C infected individuals in the Central Region of Portugal. *J. Clin. Virol.* 46: 169-172.
82. Kozisek M, Saskova KG, Rezacova P, Brynda J, van Maarseveen NM, de Jong D, Boucher CA, Kagan RM, Nijhuis M, Konvalinka J (2008) Ninety-nine is not enough: molecular characterization of inhibitor-resistant human immunodeficiency virus type 1 protease mutants with insertions in the flap region. *J. Virol.* 82: 5869-5878.
83. Tramuto F, Bonura F, Mancuso S, Romano N, Vitale F (2005) Detection of a new 3-base pair insertion mutation in the protease gene of human immunodeficiency virus type 1 during highly active antiretroviral therapy (HAART). *AIDS Res. Hum. Retroviruses* 21: 420-423.

Chapter 2

1. Bergfors, TM (1999). Protein crystallization: techniques, strategies and tips. La Jolla: International University Line.
2. Jancarik J, Kim S-H (1991) Sparse matrix sampling: a screening method for crystallization of proteins. *J. Appl. Cryst.* 24: 409-411.
3. Arndt UW (2003) Personal X-ray reflections. *Methods Enzymol.* 368: 21–42.
4. Yang C, Pflugrath JW, Courville DA, Stence CN, Ferrara JD (2003) Away from the edge: SAD phasing from the sulfur anomalous signal measured in-house with chromium radiation. *Acta Cryst.* D59: 1943–1957.
5. Arndt UW, Long JVP, Duncumb P (1998) A microfocus X-ray tube used with focusing collimators. *J. Appl. Cryst.* 31: 936–944.
6. Kabsch W (2010) Integration, scaling, space-group assignment and post-refinement. *Acta Cryst.* D66: 133–144.
7. Leslie AGW (1999) Integration of macromolecular diffraction data. *Acta Cryst.* D55: 1696–1702.
8. Otwinowski Z, Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* 276: 307–326.
9. Pflugrath JW (1999) The finer things in X-ray diffraction data collection. *Acta Cryst.* D55: 1718–1725.
10. Evans P (2006) Scaling and assessment of data quality. *Acta Cryst.* D62: 72–82.
11. Otwinowski Z, Borek D, Majewski W, Minor W (2003) Multiparametric scaling of diffraction intensities. *Acta Cryst.* A59: 228–234.

12. Kabsch W (1988b) Evaluation of single crystal X-ray diffraction data from a position-sensitive detector. *J. Appl. Cryst.* 21: 916-924.
13. Miller R, Gallo SM, Khalak HG, Weeks CM (1994) *SnB*: crystal structure determination via *Shake-and-Bake*. *J. Appl. Cryst.* 27: 613-621.
14. Schneider TR, Sheldrick GM (2002) Substructure solution with SHELXD. *Acta Cryst. D58*: 1772- 1779.
15. La Fortelle E de, Bricogne G (1997) Maximum-likelihood heavy-Atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Methods Enzymol.* 276: 472-494.
16. Debreczeni JE, Bunkoczi G, Ma Q, Blaser H, Sheldrick GM (2003) In-house measurement of the sulfur anomalous signal and its use for phasing. *Acta Cryst. D59*: 688-696.
17. Foadi J, Woolfson MM, Dodson EJ, Wilson KS, Jia-xing Y, Chao-de Z (2000) A flexible and efficient procedure for the solution and phase refinement of protein structures. *Acta Cryst. D56*: 1137-1147.
18. Jones TA, Zou JY, Cowan SW, Kjeldgaard M (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Cryst. A47*: 110–119.
19. McRee D (1999) XtalView/Xfit—A versatile program for manipulating atomic coordinates and electron density. *J. Struct. Biol.* 125: 156–165.
20. Oldfield TJ (2000) From maps to molecules in minutes. *Acta Cryst. A56*: s27.
21. Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Cryst. D60*: 2126–2132.

22. Emsley P, Lohkamp B, Scott WG and Cowtan K (2010) Features and Development of Coot. *Acta Cryst.* D66: 486-501.
23. Turk D (2000) From modeling and refinement to refinement with modelling. *Acta Cryst.* A56: s27.
24. Turk D (2004) MAIN in 2004: model building at 100 residues per minute. *Acta Cryst.* A60: s16.
25. Perrakis A, Morris R, Lamzin VS (1999) Automated protein model building combined with iterative structure refinement. *Nature Struct. Biol.* 6: 458–463.
26. Terwilliger TC (2000a) Maximum likelihood density modification. *Acta Cryst.* D56: 965–972.
27. Levitt DG (2001) A new software routine that automates the fitting of protein X-ray crystallographic electron-density maps. *Acta Cryst.* D57: 1013–1019.
28. Cowtan K (2006) The buccaneer software for automated model building. *Acta Cryst.* D62: 1002-1011.
29. Schröder GF, Levitt M, Brunger AT (2010) Super-resolution biomolecular crystallography with low-resolution data. *Nature* 464: 1218-1224.
30. Murshudov GN, Skubařk P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, Winn MD, Longa F, Vagin AA (2011) REFMAC5 for the refinement of macromolecular crystal structures. *Acta Cryst.* D67: 355–367.
31. Adams PA, Afonine PV, Bunkóczy G, Chen VB, Davis IW, Echols N, Headd JJ, Hung L-W, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC, Zwart PH (2010)

- PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Cryst.* D66: 213-221.
32. Brunger AT, Adams PD, Clore GM, Gros P, Grosse-Kunstleve RW, Jiang J -S, Kuszewski J, Nilges N, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL (1998) Crystallography & NMR System (CNS), A new software suite for macromolecular structure determination. *Acta Cryst.* D54: 905-921.
33. Blanc E, Roversi P, Vonrhein C, Flensburg C, Lea SM, Bricogne G (2004) Refinement of severely incomplete structures with maximum likelihood in BUSTER-TNT. *Acta. Cryst.* D60: 2210-2221.
34. Brünger AT (1992). The free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature (London)* 355: 472–474.
35. Kleywegt GJ, Jones TA (1995) Where freedom is given, liberties are taken. *Structure* 3: 535–540.
36. Ramakrishnan C, Ramachandran GN (1965) Stereochemical criteria for polypeptide and protein chain conformations. II. Allowed conformations for a pair of peptide units. *Biophys. J.* 5: 909–933.
37. Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK: A program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* 26: 283–291.
38. Hooft RWW, Vriend G, Sander C, Abola EE (1996). Errors in protein structures. *Nature (London)* 381: 272.

39. Vaguine AA, Richelle J, Wodak SJ (1999) SFCHECK: A unified set of procedures for evaluating the quality of macromolecular structure-factor data and their agreement with the atomic model. *Acta Cryst. D55*: 191–205.
40. Davis IW, Murray LW, Richardson JS, Richardson DC (2004) MolProbity: structure validation and all-atom contact analysis for nucleic acids and their complexes. *Nucleic Acids Res.* 32: W615–W619.

Chapter 3

1. Chatfield DC, Brooks BR (1995) HIV-1 protease cleavage mechanism elucidated with molecular dynamics simulation. *J. Am. Chem. Soc.* 117: 5561-5572.
2. Okimoto N, Tsukui T, Hata M, Hoshino T, Tsuda M (1999) Hydrolysis mechanism of the phenylalanine-proline peptide bond specific to HIV-1 protease: Investigation by the ab initio molecular orbital method. *J. Am. Chem. Soc.* 121: 7349- 7354.
3. Piana S, Bucher D, Carloni P, Rothlisberger U (2004) Reaction mechanism of HIV-1 protease by hybrid car-parrinello/ classical md simulations. *J. Phys. Chem.* 108: 11139-11149.
4. Trylska J, Grochowski P, McCammon JA (2004) The role of hydrogen bonding in the enzymatic reaction catalyzed by HIV-1 protease. *Protein Sci* 13: 513-528.
5. Hyland LJ, Tomaszek TA Jr, Meek TD (1991) Human immunodeficiency virus-1 protease. 2. Use of pH rate studies and solvent kinetic isotope effects to elucidate details of chemical mechanism. *Biochemistry* 30: 8454–8463.

6. Hyland LJ, Tomaszek TA Jr, Roberts GD, Carr SA, Magaard VW, Bryan HL, Fakhoury SA, Moore ML, Minnich MD, Culp JS, Desjarlais RL, Meek TD (1991) Human immunodeficiency virus-1 protease. 1. Initial velocity studies and kinetic characterization of reaction intermediates by ^{18}O isotope exchange. *Biochemistry* 30: 8441-8453.
7. Northrop DB (2001) Follow the protons: a low-barrier hydrogen bond unifies the mechanisms of the aspartic proteases. *Acc. Chem. Res.* 34: 790–797.
8. Jaskolski M, Tomasselli AG, Sawyer TK, Staples DG, Heinrikson RL, Schneider J, Kent SB, Wlodawer A (1991) Structure at 2.5-Å resolution of chemically synthesized human immunodeficiency virus type 1 protease complexed with a hydroxyethylene-based inhibitor. *Biochemistry* 30: 1600–1609.
9. Rose RB, Craik CS, Douglas NL, Stroud RM (1996) Three-dimensional structures of HIV-1 and SIV protease product complexes. *Biochemistry* 35: 12933–12944.
10. Silva AM, Cachau RE, Sham HL, Erickson JW (1996) Inhibition and catalytic mechanism of HIV-1 aspartic protease. *J. Mol. Biol.* 255: 321–346.
11. Kovalevsky AY, Chumanevich AA, Liu F, Louis JM, Weber IT (2007) Caught in the act: The 1.5 Å resolution crystal structures of the HIV-1 protease and the I54V mutant reveal a tetrahedral reaction intermediate. *Biochemistry* 46: 14854-14864.
12. Johnson EB, Malito E, Shen Y, Pentelute B, Rich D, Florian J, Tang WJ, Kent SBH (2007) Insights from atomic-resolution X-ray structures of chemically-

- synthesized HIV-1 protease in complex with inhibitors. *J. Mol. Biol.* 373: 573-586.
13. Brik A, Wong CH (2003) HIV-1 protease: mechanism and drug discovery. *Org. Biomol. Chem.* 1: 5-14.
 14. Dunn BM (2002) Structure and mechanism of the pepsin-like family of aspartic peptidases. *Chem. Rev.* 102: 4431–4458.
 15. Torbeev VY, Mandal K, Terechko VA, Kent BH (2008) Crystal structure of chemically synthesized HIV-1 protease and a ketomethylene isostere inhibitor based on the p2/NC cleavage site. *Bioorg. Med. Chem. Lett.* 18:4554-4557.
 16. Yunfeng Tie, Boross PI, Wang YF, Gaddis L, Hussain AK, Leshchenko S, Ghosh AK, Louis JM, Harrison RW, Weber IT (2004) High resolution crystal structures of HIV-1 protease with a potent non-peptide inhibitor (UIC-94017) active against multi-drug-resistant clinical strains. *J. Mol. Biol.* 338: 341-352.
 17. Kumar M, Prashar V, Mahale S, Hosur MV (2005) Observation of a tetrahedral reaction intermediate in the HIV-1 protease-substrate complex. *Biochem. J.* 389: 365-371.
 18. Das A, Prashar V, Mahale S, Serre L, Ferrer J-L, Hosur MV (2006) Crystal structure of HIV-1 protease in situ product complex and observation of a low-barrier hydrogen bond between catalytic aspartates. *Proc. Natl. Acad. Sci. U S A* 103: 18464-18469.

19. Bihani S, Das A, Prashar V, Ferrer J-L, Hosur MV (2009) X-ray structure of HIV-1 protease in situ product complex. *PROTEINS: Structure, Function and Bioinformatics* 74: 594 – 602.
20. Pillai B, Kannan KK, Hosur MV (2001) 1.9 Å X-ray study shows closed flap conformation in crystals of tethered HIV-1 PR. *PROTEINS: Struct. Funct. Genet.* 43: 57-64.
21. Kumar M, Kannan KK, Hosur MV, Bhavesh NS, Chatterjee A, Mittal R, Hosur RV (2002) Effects of remote mutation on the autolysis of HIV-1 PR: X-ray and NMR investigations. *Biochem. Biophys. Res. Commun.* 294:395-401.
22. Anderson VE, Ruszczycky MW, Harris ME (2006) Activation of oxygen nucleophiles in enzyme catalysis. *Chem. Rev.* 106: 3236-3251.
23. Adachi M, Ohhara T, Kurihara K, Tamada T, Honjo E, et al. (2009) Structure of HIV-1 protease in complex with potent inhibitor KNI-272 determined by high-resolution X-ray and neutron crystallography. *Proc. Natl. Acad. Sci. U S A* 106: 4641-4646.
24. Cheng Y-SE, Yin FH, Foundling S, Blomstrom D, Kettner CA (1990) Stability and activity of human immunodeficiency virus protease: Comparison of the natural dimer with a homologous, single chain tethered dimer. *Proc. Natl. Acad. Sci. U S A* 87: 9660-9664.
25. Roth M, Carpentier P, Kaikati O, Joly J, Charrault P, Pirocchi M, Kahn R, Fanchon E, Jacquamet L, Borel F, Bertoni A, Israel-Gouy P, Ferrer JL (2002) FIP: A highly automated beamline for multiwavelength anomalous diffraction

- experiments. *Acta Cryst.* D58: 805-814.
26. Kabsch W (1993) Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Cryst.* 26:795–800.
 27. McCoy AJ, Grosse-Kunstleve RW, Sotoroni LC, Read RJ (2005) Likelihood-enhanced fast translation functions. *Acta Cryst.* D61:458-464.
 28. Sotoroni LC, McCoy AJ, Read RJ (2004) Likelihood-enhanced fast rotation functions. *Acta Cryst.* D60: 432-438.
 29. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissiq H, Shindyalov IN, Bourne PE (2000) The protein data bank. *Nucleic Acids Res.* 28: 235-242.
 30. Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges N, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL (1998) Crystallography and NMR system (CNS): A new software system for macromolecular structure determination. *Acta Cryst.* D54: 905–921.
 31. Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges N, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL (2001) *Crystallography & NMR System*, Yale University.
 32. Brunger AT (1992) The free R value: A novel statistical quantity for assessing the accuracy of crystal structures. *Nature* 355: 472-474.
 33. Jones TA, Zou JY, Cowan SW, Kjeldgaard M (1991) Improved methods for building protein models in electron density maps and the location of errors in

- these models. *Acta Cryst.* A47: 110-119.
34. DeLano WL (2002) The PYMOL molecular graphics system, DeLano Scientific, San Carlos, USA.
 35. Hine J (1977) The principle of least nuclear motion. *Adv. Phys. Org. Chem.* 15: 1-61.
 36. Prabu-Jeyabalan M, Nalivaika E, Schiffer CA (2002) Substrate shape determines specificity of recognition for HIV-1 protease: Analysis of crystal structures of six substrate complexes structure. *Structure* 10: 369–381.
 37. Tyndall JDA, Pattenden LK, Reid RC, Hu SH, Alewood D, Alewood PF, Walsh T, Fairlie DP, Martin JL (2008) Crystal structures of highly constrained substrate and hydrolysis products bound to HIV-1 protease: Implications for the catalytic mechanism. *Biochemistry* 47: 3736–3744.
 38. James MNG, Sielecki AR, Hayakawa K, Gelb MH (1992) Crystallographic analysis of transition state mimics bound to penicillopepsin: difluorostatine- and difluorostatone-containing peptides. *Biochemistry* 31: 3872-3886.

Chapter 4

1. Montagnier L (2010) 25 years after HIV discovery: Prospects for cure and vaccines. *Virology* 397: 248–254.
2. Barré-Sinoussi F, Chermann JC, Rey F, Nugeyre MT, Chamaret S, Gruest J, Dauguet C, Axler-Blin C, Vezinet-Brun F, Rouzioux C, Rozenbaum W,

- Montagnier L (1983) Isolation of a T-lymphotropic retrovirus from a patient at risk for Acquired Immunodeficiency Syndrome (AIDS). *Science* 220: 868-871.
3. Gallo RC, Salahuddin SZ, Popovic M, Shearer GM, Kaplan M, Haynes BF, Palker TJ, Redfield R, Oleske J, Safai B, White G, Foster P, Markham PD (1984) Frequent detection and isolation of cytopathic retroviruses (HTLV-III) from patients with AIDS and at risk for AIDS. *Science* 224: 500-503.
 4. Clercq ED (2002) Strategies in the design of antiviral drugs. *Nat. Rev. Drug Discovery* 1:13–25.
 5. Kohl NE, Emini EA, Schleif WA, Davis LJ, Heimbach JC, Dixon RAF, Scolnick EM, Sigal IE(1988) Active human immunodeficiency virus protease is required for viral infectivity. *Proc. Natl. Acad. Sci. USA* 85: 4686-4690.
 6. Roberts NA, Martin JA, Kinchington D, Broadhurst AV, Craig JC, Duncan IB, Galpin SA, Handa BK, Kay J, Krohn A, Lambert RW, Merrett JH, Mills JS, Parkes KEB, Redshaw S, Ritchie AJ, Taylor DL, Thomas GJ, Machin PJ (1990) Rational design of peptide-based HIV Proteinase inhibitors. *Science* 248: 358-361.
 7. Svicher V, Ceccherini-Silberstein F, Erba F, Santoro M, Gori C, Concetta Bellocchi M, Giannella S, Trotta MP, Monforte A, Antinori A, Perno CF (2005) Novel human immunodeficiency virus Type 1 protease mutations potentially involved in resistance to protease inhibitors. *Antimicrob. Agents Chemother.* 49: 2015-2025.
 8. Shafer RW, Schapiro JM (2008) HIV-1 drug resistance mutations: an updated framework for the second decade of HAART. *AIDS Rev.* 10: 67-84.

9. Hong L, Zhang XC, Hartsuck JA, Tang J (2000) Crystal structure of an in vivo HIV-1 protease mutant in complex with saquinavir: Insights into the mechanisms of drug resistance. *Protein Sci.* 9: 1898–1904.
10. Wittayanarakul K, Aruksakunwong O, Saen-oon S, Chantratita W, Parasuk V, Sompornpisut P, Hannongbua S (2005) Insights into saquinavir resistance in the G48V HIV-1 protease: Quantum calculations and molecular dynamic simulations. *Biophys. J.* 88: 867–879.
11. Stoica I, Sadiq SK, Coveney PV (2008) Rapid and accurate prediction of binding free energies for saquinavir-bound HIV-1 proteases. *J. Am. Chem. Soc.* 130: 2639–2648.
12. Cheng Y-SE, Yin FH, Foundling S, Blomstrom D, Kettner CA (1990) Stability and activity of human immunodeficiency virus protease: Comparison of the natural dimer with a homologous, single chain tethered dimer. *Proc. Natl. Acad. Sci. USA* 87: 9660-9664.
13. Pillai B, Kannan KK, Hosur MV (2001) 1.9 Å X-ray study shows closed flap conformation in crystals of tethered HIV-1 PR. *PROTEINS: Struct. Funct. Genet.* 43: 57-64.
14. Leslie AGW (1999) Integration of macromolecular diffraction data. *Acta Cryst.* D55: 1696-1702.
15. Evans PR (2005) Scaling and assessment of data quality. *Acta Cryst.* D62: 72-82.
16. Adams PA, Afonine PV, Bunkóczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung L-W, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC, Zwart PH (2010)

- PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Cryst. D*66: 213-221.
17. Brunger AT (1992) The free R value: A novel statistical quantity for assessing the accuracy of crystal structures. *Nature* 355: 472-474.
 18. Jones TA, Zou JY, Cowan SW, Kjeldgaard M (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Cryst. A*47: 110-119.
 19. Laskowski RA, Mac Arthur MW, Moss DS, Thornton JM (1993) PROCHECK: A program to check the stereochemical quality of protein structure. *J. Appl. Cryst.* 26: 283-291.
 20. DeLano WL (2002) The PyMOL molecular graphics system, DeLano Scientific, San Carlos, CA.
 21. Greenfield NJ (2006) Using circular dichroism collected as a function of temperature to determine the thermodynamics of protein unfolding and binding interactions. *Nat. Protoc.* 16: 2527–2535.
 22. Liu F, Kovalevsky AY, Tie Y, Ghosh AK, Harrison RW, Weber IT (2008) Effect of flap mutations on structure of HIV-1 protease and inhibition by saquinavir and darunavir. *J. Mol. Biol.* 381: 102–115.
 23. Ode H, Neya S, Hata M, Sugiura W, Hoshino T (2006) Computational simulations of HIV-1 proteases-multi-drug resistance due to nonactive site mutation L90M. *J. Am. Chem. Soc.* 128: 7887–7895.

24. Muzammil S, Ross P, Freire E (2003) A major role for a set of non-active site mutations in the development of HIV-1 protease drug resistance. *Biochemistry* 42: 631-638.
25. Xie D, Gulnik S, Gustchina E, Yu B, Shao W, Qoronfleh W, Nathan A, Erickson JW (1999) Drug resistance mutations can affect dimer stability of HIV-1 protease at neutral pH. *Protein Sci.* 8: 1702-1707.
26. Noel AF, Bilsel O, Kundu A, Wu Y, Zitzewitz JA, Matthews R (2009) The folding free-energy surface of HIV-1 protease: Insights into the thermodynamic basis for resistance to inhibitors. *J. Mol. Biol.* 387: 1002 – 1016.
27. Rose RB, Craik CS, Stroud RM (1998) Domain flexibility in retroviral proteases: Structural implications for drug resistant mutations. *Biochemistry* 37: 2607-2621.

Chapter 5

1. Muzammil S, Ross P, Freire E (2003) A major role for a set of non-active site mutations in the development of HIV-1 protease drug resistance. *Biochemistry* 42: 631-638.
2. Svicher V, Ceccherini-Silberstein F, Erba F, Santoro M, Gori C, Concetta Bellocchi M, Giannella S, Trotta MP, Monforte A, Antinori A, Perno CF (2005) Novel human immunodeficiency virus Type 1 protease mutations potentially involved in resistance to protease inhibitors. *Antimicrob. Agents Chemother.* 49: 2015-2025.
3. Shafer RW, Schapiro JM (2008) HIV-1 drug resistance mutations: an updated framework for the second decade of HAART. *AIDS Rev.* 10: 67-84.

4. Hong L, Zhang XC, Hartsuck JA, Tang J (2000) Crystal structure of an in vivo HIV-1 protease mutant in complex with saquinavir: Insights into the mechanisms of drug resistance. *Protein Sci.* 9: 1898–1904.
5. Cheng Y-SE, Yin FH, Foundling S, Blomstrom D, Kettner CA (1990) Stability and activity of human immunodeficiency virus protease: Comparison of the natural dimer with a homologous, single chain tethered dimer. *Proc. Natl. Acad. Sci. USA* 87: 9660-9664.
6. Pillai B, Kannan KK, Hosur MV (2001) 1.9 Å X-ray study shows closed flap conformation in crystals of tethered HIV-1 PR. *PROTEINS: Struct. Funct. Genet.* 43: 57-64.
7. Greenfield NJ (2006) Using circular dichroism collected as a function of temperature to determine the thermodynamics of protein unfolding and binding interactions. *Nat. Protoc.* 16: 2527–2535.
8. Roth M, Carpentier P, Kaikati O, Joly J, Charrault P, Pirocchi M, Kahn R, Fanchon E, Jacquamet L, Borel F, Bertoni A, Israel-Gouy P, Ferrer JL (2002) FIP: A highly automated beamline for multiwavelength anomalous diffraction experiments. *Acta Cryst.* D58: 805-14.
9. Kabsch W (2010) XDS. *Acta Cryst.* D66: 125–132.
10. Murshudov GN, Vagin AA, Dodson, EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Cryst.* D53: 240-255.
11. Brunger AT (1992) The free R value: A novel statistical quantity for assessing the accuracy of crystal structures. *Nature* 355: 472-474.

12. Jones TA, Zou JY, Cowan SW, Kjeldgaard M (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Cryst.* A47: 110-119.
13. Laskowski RA, Mac Arthur MW, Moss DS, Thornton JM (1993) PROCHECK: A program to check the stereochemical quality of protein structure. *J. Appl. Cryst.* 26: 283-291.
14. DeLano WL (2002) The PyMOL molecular graphics system, DeLano Scientific, San Carlos, CA.
15. Dundas J, Ouyang Z, Tseng J, Binkowski A, Turpaz Y, Liang J (2006) CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. *Nucl. Acids Res.* 34:W116-W118.
16. Schechter I, Berger A (1967) On the size of the active site in proteases. I. Papain. *Biochem. Biophys. Res. Commun.* 27: 157-162.
17. Munshi S, Chen ZG, Yan YW, Li Y, Olsen DB, Schock, HB, Galvin BB, Dorsey B, Kuo LC (2000) An alternate binding site for the P1-P3 group of a class of potent HIV-1 protease inhibitors as a result of concerted structural change in the 80s loop of the protease. *Acta Cryst.* D56: 381-388.
18. Xie D, Gulnik S, Gustchina E, Yu B, Shao W, Qoronfleh W, Nathan A, Erickson JW (1999) Drug resistance mutations can affect dimer stability of HIV-1 protease at neutral pH. *Protein Sci.* 8: 1702-1707.

19. Rose RB, Craik CS, Stroud RM (1998) Domain flexibility in retroviral proteases: Structural implications for drug resistant mutations. *Biochemistry* 37: 2607-2621.

Chapter 6

1. Flexner, C (1998) HIV-protease inhibitors. *N. Engl. J. Med.* 338: 1281–1292.
2. Cihlar T, Bischofberger N (2000) Recent developments in antiretroviral therapies. In *Annual reports in medicinal chemistry*; Academic Press: San Diego, CA, 2000; Vol. 35, Chapter 16, pp 177-189.
3. Sepkowitz KA (2001) AIDS- the first 20 years. *N. Engl. J. Med.* 344: 1764–1772.
4. Waters L, Nelson M (2007) Why do patients fail HIV therapy? *Int. J. Clin.Pract.* 61: 983–990.
5. The First HIV protease inhibitors approved by FDA. *Antiviral Agents Bull.* 1995, 8, 353–355.
6. Roberts NA, Martin JA, Kinchington D, Broadhurst AV, Craig JC, Duncan IB, Galpin SA, Handa BK, Kay J, Krohn A, Lambert RW, Merrett JH, Mills JS, Parkes KEB, Redshaw S, Ritchie AJ, Taylor DL, Thomas GJ, Machin PJ (1990) Rational design of peptide-based HIV proteinase inhibitors. *Science* 248: 358–361.
7. Kaldor SW, Kalish VJ, Davies JF, Shetty BV, Fritz JE, Appelt K, Burgess JA, Campanale KM, Chirgadze NY, Clawson DK, Dressman BA, Hatch SD, Khalil DA, Kosa MB, Lubbehusen PP, Muesing MA, Patick AK, Reich SH, Su KS, Tatlock JH (1997) Viracept (Nelfinavir Mesylate, AG1343): A potent, orally bioavailable inhibitor of HIV-1 protease. *J. Med. Chem.* 40: 3979-3985.

8. Rhee SY, Taylor J, Fessel WJ, Kaufman D, Towner W, Troia P, Ruane P, Hellinger J, Shirvani V, Zolopa A, Shafer RW (2010) HIV-1 protease mutations and protease inhibitor cross resistance. *Antimicrob. Agents Chemother.* 54: 4253–4261.
9. Patick AK, Duran M, Cao Y, Shugarts D, Keller MR, Mazabel E, Knowles M, Chapman S, Kuritzkes DR, Markowitz M (1998) Genotypic and phenotypic characterization of human immunodeficiency virus type 1 variants isolated from patients treated with the protease inhibitor nelfinavir. *Antimicrob. Agents Chemother.* 42: 2637-2644.
10. Pillai B, Kannan KK, Hosur MV (2001) 1.9 Å X-ray study shows closed flap conformation in crystals of tethered HIV-1 PR. *PROTEINS: Struct. Funct. Genet.* 43: 57-64.
11. Roth M, Carpentier P, Kaikati O, Joly J, Charrault P, Pirocchi M, Kahn R, Fanchon E, Jacquamet L, Borel F, Bertoni A, Israel-Gouy P, Ferrer JL (2002) FIP: A highly automated beamline for multiwavelength anomalous diffraction experiments. *Acta Cryst. D*58: 805-14.
12. Kabsch W (2010) XDS. *Acta Cryst. D*66: 125–132.
13. Adams PA, Afonine PV, Bunkóczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung L-W, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC, Zwart PH (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Cryst. D*66: 213-221.
14. Brunger AT (1992) The free R value: A novel statistical quantity for assessing the accuracy of crystal structures. *Nature* 355: 472-474.

15. Jones TA, Zou JY, Cowan SW, Kjeldgaard M (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Cryst. A*47: 110-119.
16. Laskowski RA, Mac Arthur MW, Moss DS, Thornton JM (1993) PROCHECK: A program to check the stereochemical quality of protein structure. *J. Appl. Cryst.* 26: 283-291.
17. DeLano WL (2002) The PyMOL molecular graphics system, DeLano Scientific, San Carlos, CA.
18. Prabu-Jeyabalan M, Nalivaika E, Schiffer CA (2000) How does a symmetric dimer recognize an asymmetric substrate? A substrate complex of HIV-1 protease. *J. Mol. Biol.* 301: 1207-1220.
19. Prabu-Jeyabalan M, Nalivaika E, Schiffer CA (2002) Substrate shape determines specificity of recognition for HIV-1 protease: Analysis of crystal structures of six substrate complexes. *Structure* 10: 369–381
20. Kolli M, Lastere S, Schiffer CA (2006). Co-evolution of nelfinavir-resistant HIV-1 protease and the p1-p6 substrate. *Virology* 347: 405-409.
21. Ghosh AK (2009) Harnessing Nature's Insight: Design of aspartyl protease Inhibitors from treatment of drug-Resistant HIV to alzheimer's disease. *J. Med. Chem.* 52: 2163-2175.
22. Tie Y, Boross PI, Wang YF, Gaddis L, Hussain AK, Leshchenko S, Ghosh AK, Louis JM, Harrison RW, Weber IT (2004) High resolution crystal structures of HIV-1 protease with a potent non-peptide inhibitor (UIC-94017) active against multi-drug-resistant clinical strains. *J. Mol. Biol.* 338: 341-352.

23. Koh Y, Nakata H, Maeda K, Ogata H, Bilcer G, Devasamudram T, Kincaid JF, Boross P, Wang Y-F, Tie Y, Volarath P, Gaddis L, Harrison RW, Weber IT, Ghosh A K, Mitsuya H (2003) Novel *bis*-tetrahydrofuranylurethane-containing nonpeptidic protease inhibitor (PI) UIC-94017 (TMC114) with potent activity against multi-PI-resistant human immunodeficiency virus in vitro. *Antimicrob. Agents Chemother.* 47: 3123-3129.