Understanding Drug Resistance in HIV-1 Protease through X-ray Crystallography

By

Subhash Chandra Bihani
(LIFE01200604004)
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 **Subhash Chandra Bihani** entitled "**Understanding drug resistance in HIV-1 protease through X-ray crystallography**" and recommend that it may be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy.

Chairman- Dr. S.F. D'Souza **Date:** 01/10/2012

External examiner- Prof. M. Vijayan **Date:** 01/10/2012

Guide / Convener- Dr. M.V. Hosur **Date:** 01/10/2012

Member 1- Dr. J. K. Sainis **Date:** 01/10/2012

Member 2- Dr. Vinay Kumar **Date:** 01/10/2012

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: 01/10/2012

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.

Subhash

(Subhash Chandra Bihani)

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.

Subhash

(Subhash Chandra Bihani)

Dedicated to my family

ACKNOWLEDGEMENT

I would like to express my deepest sense of gratitude to my advisor Dr. M. V. Hosur for providing continuous support and guidance during the course of my Ph.D study. I wish to express my sincere gratitude to the rest of my doctoral committee: Dr, J. K. Dattgupta, Dr. S. F. D'Souza, Dr. J. K. Sainis, and Dr. V. Kumar, for their invaluable suggestions.

I would like to show my greatest appreciation to Dr. K. K. Kannan for his suggestions and motivation. I would like to take this opportunity to thank my colleagues at Bhabha Atomic Research Centre: Dr. Lata Paniker, Dr. Mukesh Kumar, Mr. Vishal Prashar, and Dr. Amit Das. I also thank Mr. S. R. Jadhav for his help in carrying out the laboratory experiments. My thanks are also due to Jean Luc Ferrer, France for providing facility for X-ray diffraction data collection. I would like to acknowledge the Bhabha Atomic Research Centre, Mumbai for the financial support and providing facilities to carry out the experiments.

I take immense pleasure in thanking my family. Words cannot express how grateful I am to my parents and others in the family for their love and support. I would also like to thank to my wife, Neetu for supporting me for everything and great patience. To my lovely son Meet, for his selfless love and doing his best not to disturb me while I was writing my thesis.

Finally, special thanks to all those who have helped me directly or indirectly and whom I have forgotten to mention here.

CONTENTS

Understanding drug resistance in

HIV-1 protease through X-ray crystallography

A synopsis

Subhash Chandra Bihani

Homi Bhabha National Institute,

Mumbai – 400094, India.

Acquired Immuno Deficiency Syndrome (AIDS) is among one of the most devastating diseases in the world. Since its discovery in 1981, it has killed more than 30 million people worldwide and even now approximately 33 million people are living with the virus. The causal organism for AIDS is identified as a retrovirus, Human Immunodeficiency Virus (HIV) and worldwide, two serotypes were identified for HIV: HIV type 1 (HIV-1) and HIV type 2 (HIV-2). HIV-1 is the most common and more virulent between the two subtypes. Although, there is no permanent cure or a vaccine available for HIV/AIDS, several drugs have been developed which stop or slow down the progression of HIV/AIDS by targeting viral replication and keeping the virus in the patient's blood below pathological level. Currently, a cocktail of many of these drugs is used in a therapy known as Highly Active Anti-Retroviral Therapy (HAART). These drugs employ different mechanisms to fight the virus. Important ingredients of HAART therapy are inhibitors of HIV-1 protease. These drugs have reduced the problem of HIV/AIDS from the once deadly disease into a treatable medical condition. However, emergence of drug resistance mutations within HIV-1 protease is now limiting the effectiveness of these antiretroviral drugs, thereby necessitating continuous development of novel and better drugs that will remain effective against the resistant virus. This requires a clear understanding of HIV-1 protease mechanism through structural studies of HIV-1 protease in complex with true substrates. A complimentary approach is to study the molecular basis of drug resistance through crystal structures of various drug resistant HIV-1 proteases. The aim and objective of the present study are along these two approaches and the details have been arranged into four chapters as described below:

Chapter I briefly introduces HIV and describes the scientific literature related to the present work. Human Immunodeficiency Virus type-1 produces its proteins as polyproteins, which need to be cleaved, in a regulated fashion by HIV-1 protease to produce structural proteins and active enzymes. This step is called maturation, and inhibition of this step results into immature & non-infectious virus particles. Therefore, HIV-1 protease is a prime target for development of drugs against HIV/AIDS. Functionally, HIV-1 protease is an aspartyl protease comprised of two non-covalently associated, chemically identical monomers. In three-dimension, the C-terminus of first monomer is about 5 Å away from the N-terminus of the second monomer. Single chain tethered HIV-1 protease dimer have been prepared by covalently joining these two ends of the homo-dimer. The catalytic properties of tethered HIV-1 protease dimer are very similar to those of the wild type native dimer. HIV-1 protease has wide substrate specificity as it can cut at nine different cleavage sites of very different amino acid sequences. One unique feature of HIV-1 protease is that some of the cleavage sites contain a proline residue in P1' position. However, the cleavage mechanism proposals in the literature make no distinction between proline containing substrates and non-proline substrates. To gain further insight into the cleavage mechanism, structural study of HIV-1 protease in complex with proline containing oligo-peptide substrate has been undertaken.

As of now, nine inhibitors of HIV-1 protease, (Saquinavir, Indinavir, Ritonavir, Nelfinavir, Amprenavir, Lopinavir, Atazanavir, Tipranavir, and Darunavir) are approved by the Food and Drug Administration, USA, for the treatment of HIV/AIDS patients. Drug resistance mutations in HIV-1 protease are observed against all these drugs. The drug resistance mutations are classified as major and minor mutations based on the extent

of their contributions towards drug resistance. These mutations are located both inside and outside the active site region of the enzyme. Active site mutations were shown to affect by directly changing the chemical interactions between the drug and HIV-1 protease, whereas mechanism of resistance due to non-active site mutations is poorly understood. Three-dimensional structures of resistant proteases and of their drug complexes will help in elucidating the molecular mechanism of drug resistance. For the current study, I have chosen drug resistance mutations against the drug nelfinavir, which elicits following unique mutations: D30N, L90M, N88D, and N88S. D30 is located in the active site whereas L90 and N88 are located outside the active site.

Chapter II describes the materials and experimental methods. A brief description of the principles behind structure solution by X-ray diffraction is also given in this chapter. A HIV-1 protease tethered dimer gene construct in which the two protease coding regions are linked in frame by codons for Gly-Gly-Ser-Ser-Gly, was used in the present study. Site directed mutagenesis was done using Stratagene's quick-change site-directed mutagenesis protocol. Both native and mutant HIV-1 proteases were expressed and purified using the same protocol. Briefly, *Escherichia coli* (*E.coli*) BL21 (DE3) cells with plasmids carrying appropriate HIV-1 protease gene insert were grown at 37°C to mid-log phase. At this stage, the protease expression was induced by adding 1mM IPTG. Three – four hours after induction, cells were harvested by centrifugation and lysed using ultra sonication. HIV-1 protease is accumulated as inclusion bodies. Inclusion bodies were thoroughly washed and protein was extracted in denatured form with 67% ice-cooled acetic acid. Extract was diluted 33 fold in cold water and dialyzed overnight against ice-

cooled water. This was followed by dialysis against refolding buffer at pH 6.5, containing 20 mM PIPES, 100 mM NaCl, 1 mM DTT and 10% Glycerol. The protein was concentrated using 10 kD MWCO centrifugal concentrator to 1-3 mg/ml by measuring OD at 280 nm and exchanged into 50 mM sodium acetate buffer (pH 4.5). Purity was checked by running a SDS-PAGE gel.

HIV-1 protease was crystallized by hanging-drop vapor-diffusion method. Chemical conditions for crystallization of native and mutant HIV-1 proteases were explored by systematically varying: (1) pH, (2) type and concentration of buffer, (3) type and concentration of precipitant, and (4) additives. Hexagonal rod shaped crystals were obtained with ammonium sulfate (1-6% of saturated solution) as precipitant and sodium citrate/ sodium dihydrogen phosphate buffer (200/100 mM) at pH 6.2. Crystals of nelfinavir and drug resistant HIV-1 protease complexes were obtained by the method of co-crystallization. Crystals of drug complexes could be obtained in two different conditions. One of the conditions is similar to that for unliganded HIV-1 proteases. In the second condition, NaCl (0.5-1.25 M) is used as a precipitant with sodium acetate buffer (50 mM, pH 4.5) as reservoir, and DMSO (1-10 %) as additive. Diffraction data as 1º oscillation frames were collected at European Synchrotron Radiation Facility (ESRF), Grenoble, France by our collaborator Prof. Jean-Luc Ferrer. In most of the cases, 60-120 frames were indexed, integrated, and scaled using the software package XDS. All the structures were solved by difference-Fourier method using native HIV-1 protease structure (PDB ID 1LV1) as template. Molecular models were refined using the computer programs Crystallography and NMR system (CNS), Refmac and Phenix. Electron density

map interpretation and model building was done using software packages O and Coot. Production of molecular figures was done using the software Pymol.

Chapter III describes the structure of HIV-1 protease in complex with a proline containing oligo-peptide of sequence SQNYPIV, which represents the MA-CA cleavage site in the polyprotein substrate. Single crystals of unliganded native HIV-1 protease were soaked in solutions containing excess amount of the hepta-peptide. The soaking period was systematically varied, and the structure reported here is for a soak period of 24 hours. At the end of soak period, the crystal was flushed out and mounted in a Cryo-loop for the diffraction data collection. The structure was refined in CNS, using standard simulated annealing protocols and maximum likelihood target function for structure factor amplitudes. The crystallographic R-factors are: R_{work} = 20.54% and R_{free} = 23.75% (Table 1). The electron density for the protein part is of excellent quality and a positive difference density in the active site was clearly visible. This electron density was interpreted as a cleaved substrate based on electron density fit and R-factors. Present structure was compared with the structure of another enzyme-product complex, in which the substrate used did not contain a proline residue at the cleavage site. This structural superposition shows that both the products have shifted laterally in the same direction across the catalytic aspartates in the present complex. These observations suggest that positioning of the substrate in the active site may be different whenever there is a proline ring in the P1' position. Because of this lateral shift, the distance between scissile nitrogen and aspartate outer oxygen is 3.5 Å, which is too long to be a hydrogen bond. In contrast, this hydrogen bond is present in the non-proline substrate complex. This

hydrogen bond has, in fact been suggested to act as a conduit for proton transfer during the cleavage reaction. The absence of this hydrogen bond may be suggesting that the proton transfer to scissile nitrogen atom required for bond cleavage, does not happen through catalytic aspartate, either in all substrates or in substrates containing proline at P1' position.

Chapter IV gives details about the molecular basis of drug resistance caused by the mutations associated with the drug nelfinavir. This chapter is divided into three parts:

Chapter IV Part A: This part describes the effect of mutation D30N on the structure of HIV-1 protease and on its interactions with the drug nelfinavir. D30N is a major mutation against the drug nelfinavir and gives 39-fold resistance to it but also reduces cleavage efficiency of HIV-1 protease to 10–50% of wild type HIV-1 protease. This mutation is also exclusive to nelfinavir and does not give cross-resistance to any other drug. Structures of both D30N unliganded and D30N in complex with nelfinavir were solved to resolution of 1.65Å and 1.91Å respectively (Table 1). Nelfinavir phenolic-OH group forms a strong hydrogen bond ($d_{O\ldots O}$ = 2.9Å) with side chain of D30 in wild type HIV-1 protease. It was proposed by molecular dynamics studies that this hydrogen bond is lost in D30N HIV-1 protease, and this loss probably resulted into the reduction in the binding affinity of nelfinavir. Present structure of D30N HIV-1 protease in complex with nelfinavir shows that this hydrogen bond is maintained ($d_{N\ldots O}$ = 2.9Å) and therefore loss of hydrogen bond is not the real cause of decreased affinity of nelfinavir. However, D30N mutation causes loss of a charge on the side chain of 30th residue and that changes

the nature of this crucial hydrogen bond. So, we propose that change in the nature of hydrogen bond with the drug in D30N HIV-1 protease reduces the binding affinity of nelfinavir to the enzyme.

In addition to D30N, L90M is another major mutation against the drug nelfinavir, but interestingly these mutations are mutually exclusive *in-vivo*. It was shown by *in-vitro* experiments that the combination of D30N and L90M mutations is lethal to the virus. However, these *in-vitro* experiments have also shown that this combination reduces the affinity of nelfinavir to a great extent. In order to investigate further we have solved the structures of D30N/L90M double mutant HIV-1 protease in both unliganded form and as a complex with nelfinavir. **Chapter IV (B)** describes these structures. The unliganded structure of D30N/L90M HIV-1 protease shows that due to its larger size as compared to leucine, methionine has some close Van der Waals interactions with catalytic aspartates. C^{δ} atom of methionine is 3.2 and 3.5Å away from main chain O of Asp25 and Asp1025 respectively while in wild type HIV-1 protease these distances are 3.7Å. This causes structural perturbation at active site, which is one of the most rigid regions of HIV-1 protease. It is likely that these perturbations reduce the catalytic activity of the enzyme, as residues, which bind to the substrates, are affected. L to M mutation in the core region of the protein makes protein inherently less stable as immobilization of methionine incurs a greater entropy cost than leucine due to the presence of an extra bond in methionine. The stability was probed by measuring the melting temperature of D30N/L90M by Circular Dichroism Spectropolarimetry. The melting temperature was found to be lower by 8 ºC compared to the wild type HIV-1 protease suggesting lower stability of the double mutant. 90^{th} residue is located close to the dimer interface so structural perturbation

around residue 90 may also affect dimer interface. All these effects may affect HIV-1 protease activity *in-vivo* to such an extent that virus is not able to survive. To understand the reduced affinity caused by D30N/L90M, nelfinavir complexed structure was superposed onto native and D30N nelfinavir complex structures. The comparisons show that there are large deviations in the binding of nelfinavir in case of D30N/L90M. All the four subsites are affected by these mutations resulting into sub-optimal binding, and these changes should be the reason behind reduced affinity of nelfinavir.

Chapter IV (C) describes the effect of N88D and N88S mutations on the structures of HIV-1 protease. N88D alone reduces the catalytic efficiency of HIV-1 protease by 20–40% and is shown to be deleterious to virus. However, when present along with the D30N mutation, it acts as a compensatory mutation restoring loss of cleavage efficiency caused by D30N. N88S alone provides some resistance to nelfinavir. To better understand the mechanism of resistance by these non-active site mutations, we have solved the crystal structures of unliganded N88D and N88S mutant HIV-1 proteases, and have refined them to resolutions of 1.65 Å and 1.8 Å, respectively (Table 1). Interestingly, in both structures, the residue most affected by mutation is not N88, but it is D30. N88D mutation resulted in higher RMSD (0.28–0.33 Å) at catalytic loop, which may reduce cleavage efficiency of N88D protease. In N88D PR, the active site volume is increased to 1118.6 $Å^3$ from 996.6 $Å^3$ in wild type HIV-1 protease. The increased size of the active site may result in sub-optimal binding of substrates reducing cleavage efficiency of N88D HIV-1 protease. N88S mutation on the other hand has caused large structural change in the conformation of D30. The oxygen atom D30 OD2 has moved by

about 2.8 Å to form a hydrogen bond with S88 OH. This large shift in the position of D30 will lead to a loss of direct hydrogen bond between nelfinavir m-phenol group and side chain of D30 thereby reducing binding affinity toward nelfinavir.

Summary

The major results described in the present thesis are summarized:

Table 1: New HIV-1 Protease structures solved:

Sr. No.	HIV-1 protease	Ligand	Resolution (Å)	R/R_{free} (%)
1	Native	Hepta	2.0	20.54/23.75
2	D30N	-	1.65	18.87/22.43
3	D30N	NFV	1.91	21.87/24.66
4	D30N+L90M	-	1.8	18.09/22.21
5	D30N+L90M	NFV	2.0	16.61/21.67
6	N88S	-	1.8	17.80/22.52
7	N88D		1.65	18.59/20.76

2. First structure of HIV-1 protease in complex with substrate having proline at P1' was determined, which shows that these substrates bind differently in the active site.

3. Hydrogen bond interactions suggest a different method of protonation of scissile nitrogen atom in proline containing substrates.

4. D30N resistance is attributed to the weakening of crucial hydrogen bond between drug and protein due to the change in nature of hydrogen bond.

5. Plausible reasons for mutual exclusivity of D30N and L90M have been suggested.

6. Molecular basis for reduced nelfinavir affinity by D30N/L90M combination have been explained.

7. Effect of N88D mutation on HIV-1 protease structure and its catalytic activity are explained.

8. Loss of a direct hydrogen bond is probably the reason for nelfinavir resistance shown by N88S.

Future Work

Present study shows that HIV-1 protease may be using a different mechanism to cleave proline-containing substrate. Further work with other proline containing substrates is required to confirm the same and elucidate detailed molecular mechanism for cleavage. The present study also describes the molecular basis of resistance caused by some of the drug resistance mutations against the drug nelfinavir. This information can be used to suggest modifications of the drug nelfinavir in order to keep it effective against these drug resistant variants. The information can also be used for development of new drugs. One key observation, which can be useful for designing new drugs, is that the drug, which minimizes its interactions with side chain of D30, is less likely to be affected by these mutations. For N88D and N88S mutations, crystal structures in complex with nelfinavir need to be solved. These structures can provide some more information about these non-active site mutations and can validate some of the proposals made in present study.

Publications from thesis work

1. *Bihani S C, Das A, Prashar V, Ferrer J-L and Hosur M V (2009). Resistance Mechanism revealed by Crystal Structures of unliganded nelfinavir-resistant HIV-1 protease non-active site mutants N88D and N88S. Biochemical and Biophysical Research Communications 389: 295-300.*

2. *Bihani S C, Das A, Prashar V, Ferrer J-L and Hosur M V (2008). X-ray Structures of HIV-1 Protease in-situ product complex. PROTEINS, Structure, Function and Bioinformatics, 74: 594-602.*

3. *Hosur M V, Prashar V, Das A, and Bihani S (2008). Structural investigations of HIV-1 Protease complex with substrates and inhibitors; In 37th National Seminar on Crystallography at Kolkata (India), IT4.*

4. *Bihani S C, Ferrer J-L and Hosur M V. Molecular basis for mutual exclusivity of D30N and L90M in HIV-1 protease. Manuscript under preparation.*

5. *Bihani S C, Ferrer J-L and Hosur M V. Effect of drug resistance mutation D30N on the binding of drug nelfinavir. Manuscript under preparation.*

List of Figures

List of Tables

Chapter - 1

HIV-AIDS:
The Problem at The Atomic Level

Acquired Immunodeficiency Syndrome (AIDS) has become a serious health hazard with large scale social and economic impacts worldwide. Causative agent of AIDS was identified in 1983 as a retrovirus, the human immunodeficiency virus (HIV). Since then a variety of studies have been performed across the world leading to accumulation of wealth of information on different aspects of HIV biology. This detailed information about HIV infection and life cycle has led to the discovery of several inhibitors which are being successfully used for the treatment of HIV/AIDS. However, complete cure is not yet reached and emergence of drug resistance mutations is further limiting the effectiveness of currently available drugs. Development of a vaccine for HIV infection is still in progress and therefore, continued study of virus biology and the disease is required to develop novel drugs which can help in better management of HIV/AIDS menace.

AIDS: An Epidemic

The world first knew about AIDS on 5th June 1981 when, Centers for Disease Control and prevention (CDC), USA in its Morbidity and Mortality Weekly Report described a group of homosexuals suffered by some rare type of pneumonia and Kaposi's Sarcoma [Gottlieb et al 1981, MMWR 1981]. These patients also had very low level of $CD4^+$ T-lymphocytes. In 1982, this disease was identified as "acquired immunodeficiency syndrome" or AIDS. In 1983, causal organism of AIDS was identified as a retrovirus [Barre-Sinoussi et al 1983, Gallo et al 1983]. Different names like Lymphadenopathy-Associated Virus (LAV) and Human T-Lymphotrophic Virus - III (HTLV-III) were given to the virus at that time. In 1986, the International Committee on Taxonomy of Viruses recommended giving the name human immunodeficiency virus (HIV) to AIDS

virus [Coffin et al 1986]. Luc Montagnier and Francois Barre-Sinoussi were awarded the 2008 Nobel Prize in Medicine and Physiology for their contribution to discovery of HIV and its establishment as causal organism of AIDS.

Since the discovery of AIDS, more than 30 million people have died from HIV/AIDS across the world. In 2009 alone, AIDS claimed an estimated 1.6-2.1 million lives. According to UNAIDS report on the global AIDS epidemic, 2010, there were about 33.3 million people infected with HIV at the end of 2009 compared with about 26.2 million in 1999. This is about 0.6% of world's population and this number is increasing every year [UNAIDS 2010]. An estimated 2.6 million people became newly infected with HIV in 2009. The worst affected region is Sub-Saharan Africa, with an estimated 22.5 million infected people.

HIV/AIDS in India

In India, first HIV infection was reported in 1986 in female sex workers of Chennai, Tamil Nadu [Simoes et al 1987]. Though, number of HIV/AIDS infections in India is decreasing over the years, India is still among one of the countries, worst affected by HIV/AIDS [UNAIDS 2010]. India is estimated to have 2.4 million people infected with HIV in 2009 with an estimated adult HIV prevalence of 0.3% [UNAIDS 2010]. Southern and northeastern regions of the country are worst affected by HIV/AIDS.

Taxonomy of HIV

HIV is a retrovirus and belongs to *Lentivirus* genus of family *Retroviridae* [ICTV 2002]. Lentiviruses are characterized by infection with a long latency period [Levy 1993]. HIV isolates from all over the world are divided into two serotypes, HIV type 1 (HIV-1)

[Barre-sinoussi 1983] and HIV type 2 (HIV-2) [Clavel 1986]. Majority of HIV infections across the world are caused by HIV-1 as it is more virulent and is transmitted relatively easily [Gilbert et al 2003]. HIV-2 is less pathogenic and is largely confined to West Africa [Whittle et al 1994, Reeves and Doms 2002]. HIV-2 has only 60% homology with HIV-1 [Guyader et al 1987]. It is believed that both HIV-1 and HIV-2 have originated from simian immunodeficiency viruses (SIV). HIV-1 is suggested to have originated from chimpanzees (*Pan troglodytes troglodytes*) in Central Africa [Gao et al 1999, Sharp and Hahn 2010], while HIV-2 has its origin in sooty mangabey monkeys (*Cercocebus atys*) in West Africa [Sharp and Hahn 2010].

Based on the genetic variations, HIV-1 subtype is classified into following 4 different groups [De Leys et al 1990, Charneau et al 1994, Simon et al 1998, Plantier et al 2009]:

- M for major,
- O for outlier,
- N for non M or O, and
- P.

Group M is the most prevalent of all HIV-1 groups around the world, and is further sub-divided into 10 subtypes, A through K [McCutchan 2000, Korber et al 2001, UNAIDS 2010). Subtype C accounts for approximately half of the global infections while subtype B is more common in North America and Europe, accounting for less than ~12% of the total global infections [Hemelar et al 2011]. Figure 1.1 shows the relationship between different groups and subtypes of HIV-1. The bar represents a 10% difference in nucleotide sequence.

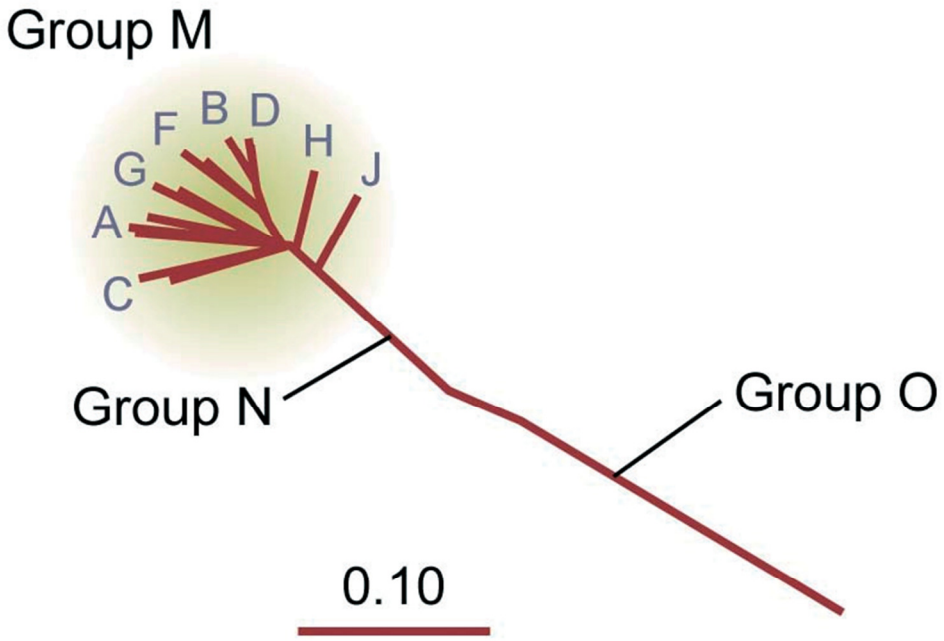

Figure 1.1: Relationship between different groups and subtypes of group M of HIV-1 (Reproduced from educational module of Virco Lab HIV resistance learning system, with permission from Virco Lab Inc, USA).

Genomic Architecture of HIV

HIV-1 has its genome in the form of plus-sense single-stranded RNA. Its life cycle however, involves conversion of this RNA to a DNA intermediate, a property unique to all retroviruses. The 9.2 kb HIV RNA genome contains nine overlapping open reading frames (ORFs) that encode 15 different proteins. Organization of these ORFs is shown in figure 1.2. Three major ORFs *gag, pol* and *env* encode major structural proteins and enzymes. Other six ORFs encode regulatory proteins, Tat, Rev, Vpu, Vpr, Vif, and Nef. HIV-2 and SIV also have similar genome organizations (Figure 1.2).

Basic Structure of HIV

HIV-1 is a complex virus with a diameter of about 90-120 nm. It has an outer phospholipid envelope and inner core containing RNA genome and enzymes (Figure 1.3). The envelope is derived from host cell membrane during budding from the infected host cell. It is studded with two viral glycoproteins, Gp120 and Gp41 [Chertova et al 2002, Zhu et al 2003, Zhu et al 2006, Liu et al 2008], which form a heterodimeric complex composed of trimers of each proteins [Chan et al 1997]. Gp120 is present on the surface whereas Gp41 forms the transmembrane anchor [Chan et al 1997, Liu et al 2008]. Both the glycoproteins play a crucial role in attachment and entry of virus into the host cell [Chan et al 1997]. The virus envelope also contains several host proteins derived from the host cell membrane during the budding process. These host proteins include a class of HLA proteins and adhesion proteins [Ott 2008, Fanales-Belasio et al 2010]. These proteins may also help during attachment of the virus onto the host cell surface [Fanales-Belasio et al 2010]. Inner surface of viral envelope, known as matrix is composed of another protein, p17 (also known as MA). HIV-1 core contains two single-

stranded, plus-sense RNA genomes tightly bound to a nucleocapsid protein p7 (also known as NC) [Turner & Summers 1999, Fanales-Belasio et al 2010]. The core also contains viral enzymes reverse transcriptase, integrase and protease [Turner & Summers 1999, Fanales-Belasio et al 2010]. The core is surrounded by capsid made up of protein p24 (also known as CA). Capsid protein, p24 is most abundant protein in HIV-1 virion.

Proteins of HIV-1

Different structural proteins and enzymes of HIV-1 are initially synthesized as part of large polyproteins Gag, Pol and Env, which are non-functional precursors of mature functional proteins. These precursors are then cleaved by proteases to yield mature functional proteins [Ganser-Pornillos et al 2008, Swanson and Malim 2008]. A 55kD Gag polyprotein, $p55^{Gag}$, produced from the *gag* gene is cleaved by viral protease to yield different structural proteins and spacer peptides, namely, p17 (MA), p24 (CA), p2, p7 (NC), p1 and p6 (Figure 1.4) [Freed 1998, Ganser-Pornillos et al 2008, Swanson and Malim 2008]. The *gag* and *pol* genes of HIV-1 overlap by 241 nucleotides (Figure 1.2), thus the *pol* gene is expressed as 160-kDa Gag-Pol fusion protein ($p160^{Gag\text{-}Pol}$) by ribosomal frame shifting. The *pol* gene encodes three viral enzymes, protease (PR), reverse transcriptase (RT) and integrase (IT). The *env* gene encodes a precursor protein $p160^{Env}$, which is cleaved into Gp120 and Gp41 by the host cell enzyme [Wlodawer and Vondrasek 1998]. Gp120 and Gp41 interact with CD4 receptors and chemokine co-receptors present on the host cell surface during attachment and entry of the virus into the host cell [Chan et al 1997, Douck et al 2002, Gougeon 2005]. Remaining six genes encode regulatory proteins (Tat and Rev) and accessory proteins (Vpu, Vpr, Vif, and Nef) of HIV [Emerman and Malim 1998, Turner & Summers 1999]. Regulatory protein Tat is

a viral transactivator. It interacts with transactivation response element (TAR) and several host factors, such as positive transcription elongation factor b. Tat promotes elongation of viral transcripts and prevents immature termination of provirus transcription [Emerman and Malim 1998, Coiras et al 2009]. The *rev* gene encodes another regulatory protein Rev that is a RNA-binding factor and controls export of viral mRNA from nucleus to cytoplasm [Pomerantz et al 1992, Pollard and Malim 1998, Hope 1999, Coiras et al 2009, Fanales-Belasio et al 2010]. This allows production of structural and functional proteins, Gag, Pol and Env [Pomerantz et al 1992]. Viral protein U (Vpu) is involved in the release of virus particles from infected cells [Bour and Strebel 2003]. The *vpr* gene encodes for Viral Protein R (Vpr) which stalls cell division in G2 phase of cell cycle [Goh et al 1998, Fanales-Belasio et al 2010]. Viral Infectivity Factor (Vif) binds to APOBEC3G from the host and induces its degradation [Marin et al 2003]. APOBEC3G restricts infection of the host cells by new virus particles. The Nef protein encoded by the gene *nef* has several functions that help in viral replication and infectivity. It is involved in signal transduction and down regulation of CD4 receptor on T-cell surface [Garcia and Miller 1991, Schwartz et al 1996, Stumptner-Cuvelette et al 2001, Alexander et al 2004, Zhu et al 2006, Fanales-Belasio et al 2010].

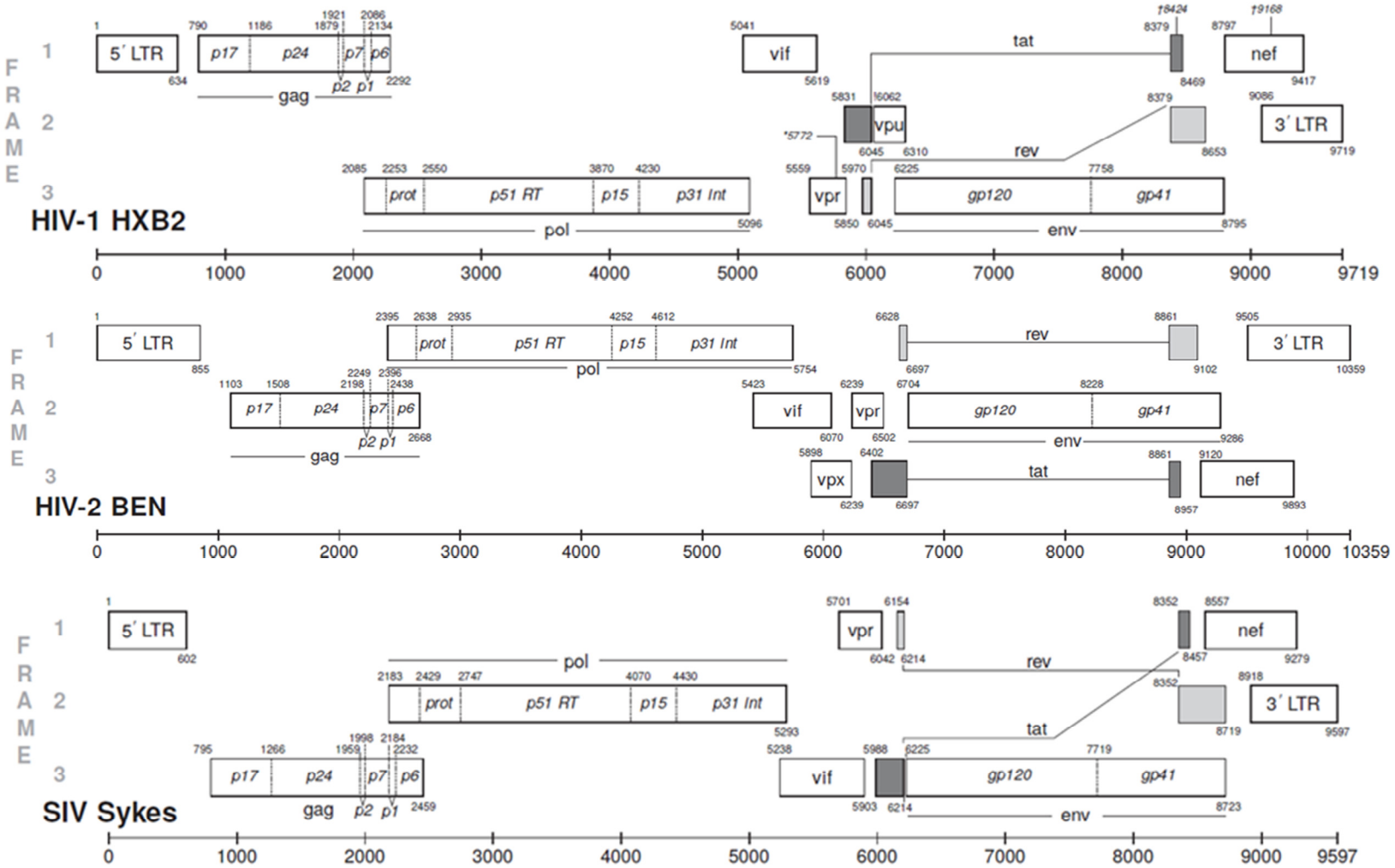

Figure 1.2: Genomic architecture and landmarks of HIV type 1, HIV type 2 and SIV: open reading frames are shown as rectangles. The gene start and end are marked by a number in upper left and lower right corners of each rectangle respectively. All three frames are shown (Reproduced with permission from HIV sequence compendium 2011, http://www.hiv.lanl.gov/).

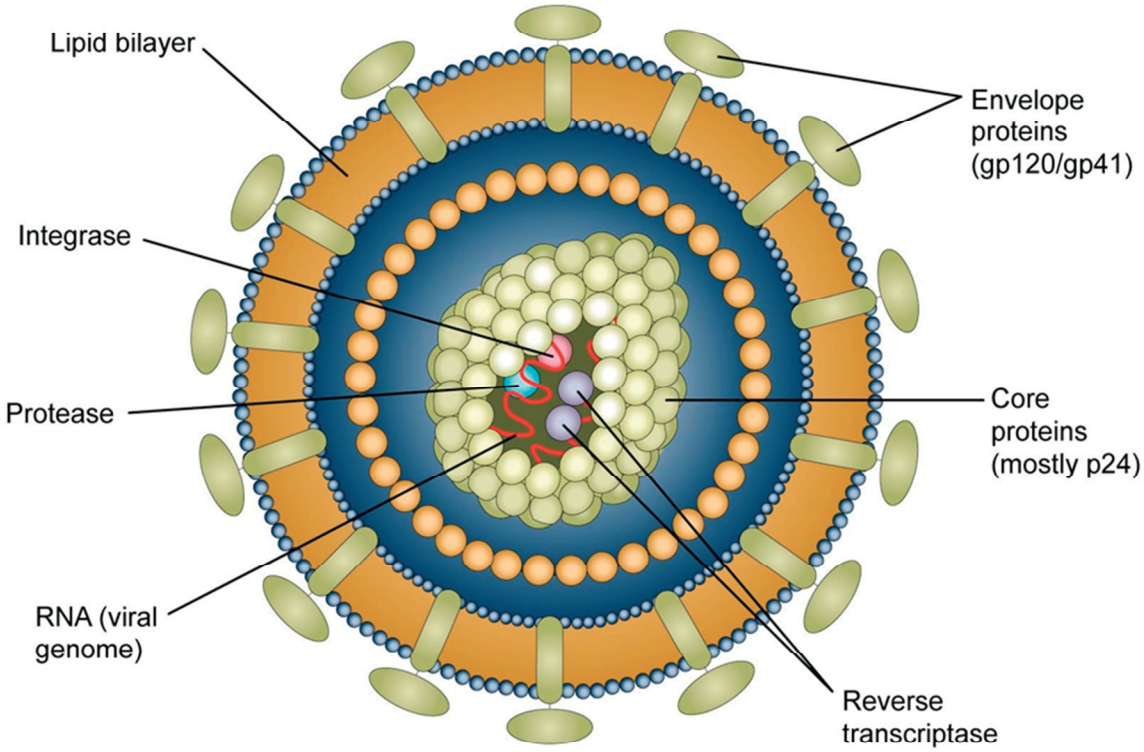

Figure 1.3: A schematic drawing of the mature human immunodeficiency virion (Reproduced from educational module of Virco Lab HIV resistance learning system, with permission from Virco Lab Inc, USA).

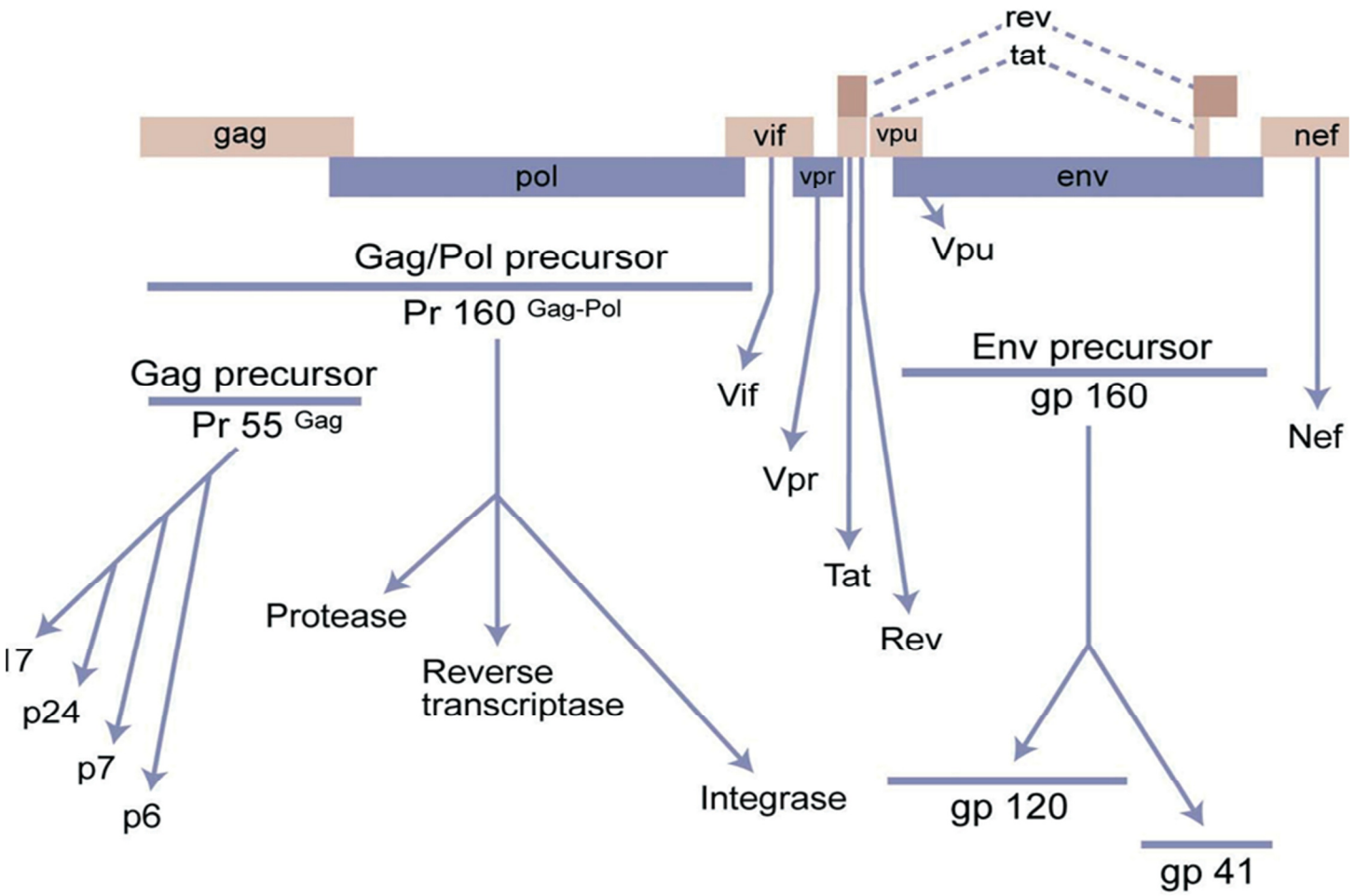

Figure 1.4: Different open reading frames and proteins of HIV (Reproduced from educational module of Virco Lab HIV resistance learning system with permission from Virco Lab Inc, USA).

HIV-1 Tropism

HIV-1 has the ability to infect different types of host cell. These host cells express different chemokine co-receptors on their surface, which are preferentially used by different strains of HIV-1 for the binding onto cell surface. This preferential use of a particular co-receptor is known as tropism. Based on chemokine co-receptor preference, three main tropic strains of HIV-1 are identified (Figure 1.5):

- M-tropic
- T-tropic
- Dual tropic

M-tropic HIV-1 strains preferentially use β-chemokine receptor CCR5 as co-receptor during binding and invasion of the host cell [Coakley et al 2005]. This CCR5 co-receptor is mainly expressed on macrophages and activated CD4 T-cells. This strain is also identified as R5 strain. T-tropic strains use CXCR4 chemokine receptor as its preferred co-receptor [Coakley et al 2005]. This co-receptor is primarily expressed on resting CD4 T-cells. These strains are also known as X4 strains. HIV-1 isolates that have the ability to infect host cells using any of these two different kinds of chemokine co-receptors are known as dual tropic or X4R5 strains.

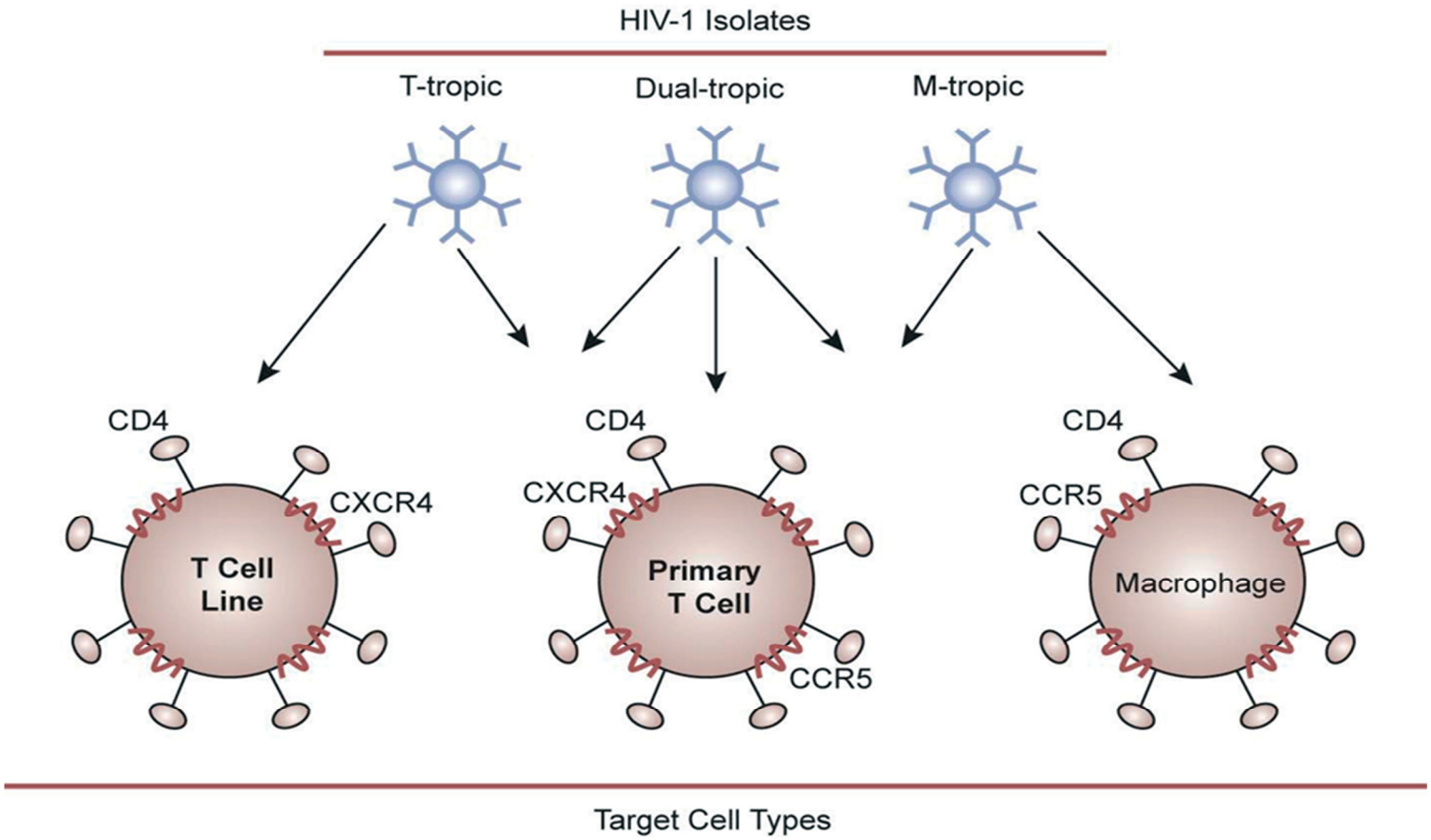

Figure 1.5: Different types of HIV tropism (Reproduced from educational module of Virco Lab HIV resistance learning system, with permission from Virco Lab Inc, USA).

Life cycle of HIV-1 and Drug Targets

Human immunodeficiency virus type-1 (HIV-1), attacks on host's immune defense system and primarily infects $CD4^+$ T-cells [Douck et al 2002, Gougeon 2005], dendritic cells and macrophages [Li et al 1999]. It uses host cell machinery to facilitate its own growth and in turn destroy the fabric of immune system. HIV/AIDS is characterized by a large reduction in number of $CD4^+$ T-cells. Three main mechanisms are identified by which HIV-1 reduces number of $CD4^+$ T-cells. First, direct killing of infected host cells, second, induction of apoptosis in surrounding host cells and third, destruction of infected host cells by $CD8^+$ killer T-cells. All these factors reduce the number of $CD4^+$ T-cells below a critical level whereby cell mediated immunity gets compromised and the host becomes prone to secondary bacterial and viral infections. According to the CDC, USA, a person is diagnosed with AIDS, when tested HIV-1 positive and has the one or both of the following conditions: a) number of CD4 positive T-cell below 200 per cubic millimeter of blood and b) one or more AIDS-related opportunistic infections. HIV-1 is primarily transmitted through three major ways: (1) sexual contact, (2) infected bodily fluids, including intravenous drugs, and (3) mother-to-child transmission, both during birth through placenta and after birth through mother's milk. Efforts all over the world have increased our understanding of HIV-1 and its life cycle. Detailed understanding of how HIV-1 infects and spreads inside the body helped scientists to develop drugs against HIV-1. These drugs act by inhibiting various stages of the viral life cycle. However, we are still far away from having a complete cure or vaccine against HIV/AIDS. The life cycle of HIV-1 is shown in figure 1.6 and can be described in following steps:

Attachment and entry into host cell

HIV-1 enters into the host cells *i.e.,* macrophages, dendritic cells, and $CD4^+$ T-cells, by a two-step procedure: first attachment on the host cell surface followed by fusion with the host cell membrane [Chan and Kim 1998, Wyatt and Sodroski 1998]. In the first step of attachment onto the host cell surface, HIV-1 interacts with receptors present on the target cell through its surface glycoproteins Gp120 and Gp41. First, Gp120 binds to CD4 receptor [Chan and Kim 1998, Wyatt and Sodroski 1998]. This interaction of Gp120 with CD4 receptor induces a conformational change in Gp120 and exposes a conserved region involved in the binding to chemokine co-receptors *i.e.,* either CCR5 or CXCR4 [Dimitrov 1997, Chan and Kim 1998, Wyatt and Sodroski 1998, Alkhatib and Berger 2007]. This dual binding of Gp120 with both CD4 receptor and chemokine co-receptor forms a more stable attachment of the virus, which, in turn, induces a conformational change in transmembrane protein Gp41 allowing it to insert its N-terminal hydrophobic fusion peptide into the host cell membrane [Fanales-Belasio et al 2010]. This facilitates fusion of the viral envelope with cell membrane and entry of the virus into the host cell.

- Binding of Gp120 to its co-receptor can be targeted to prevent virus attachment onto the host cell. One such drug Maraviroc that interferes with the binding of HIV-1 to CCR5 co-receptor is approved by FDA.
- The fusion of viral membrane with the host cell membrane is targeted by drug Enfuvirtide.

Reverse transcription and integration

After fusion of viral envelope with the host cell membrane, HIV-1 core releases its RNA genome and various enzymes, including reverse transcriptase, integrase, and protease into

the cytosol of the host cell. Reverse transcriptase converts single strand RNA genome into double stranded cDNA, which is then transported to the nucleus. The HIV-1 reverse transcriptase is highly error prone enzyme and introduces errors during synthesis of cDNA copy of RNA genome [Bebenek et al 1993, Preston et al 1988, Roberts et al 1988]. This results in high degree of genomic variability in a population of HIV-1. This double stranded cDNA is finally integrated into a host chromosome in a process mediated by viral enzyme integrase.

- Reverse transcriptase was the first enzyme targeted for anti-HIV drugs as it is essential for virus replication and unique to HIV-1. Two different types of anti-reverse transcriptase drugs have been approved by FDA: Nucleoside Reverse Trancriptase Inhibitors (NRTIs) and Non-Nucleoside Reverse Trancriptase Inhibitors (NNRTIs).
- Recently, inhibitors of another viral enzyme integrase were developed and one such inhibitor raltegravir, which blocks integrase activity, is approved by the Food and Drug Administration, USA (FDA).

Transcription and translation

During the integrated state, viral DNA is known as provirus. It can remain dormant for a very long time before its replication is triggered. The trigger is not fully understood, but is suggested to be an opportunistic infection such as tuberculosis, or deterioration of the immune system. After activation, proviral DNA is transcribed by the host cell machinery into two kinds of RNA: single-stranded full-length viral RNA genome and viral mRNA for synthesis of enzyme and other proteins.

Viral mRNA is then translated by the host cell machinery to produce viral enzymes and other proteins. Viral mRNA is initially translated in the form of precursor polyproteins, which need to be processed by HIV-1 protease and host proteases, to yield mature and functional proteins.

Assembly and release

After synthesis of viral genome and essential proteins, new virus particles assemble in the host cell. The assembly of new HIV-1 virions begins at the host cell membrane [Adamson and Freed 2007]. Two HIV-1 envelope glycoproteins Gp41 and Gp120 are transported through endoplasmic reticulum to the host cell membrane. Gag (p55) and Gag-Pol (p160) polyproteins along with RNA genome accumulate under the inner surface of the host cell membrane. At this stage, budding of immature virion from the host cell begins. During this process, it acquires a lipid-bilayer from the host cell membrane, which forms the viral envelope. Until this stage, the new virus particles are immature as their proteins are in the form of polyprotein precursors. During the budding stage or after it, a maturation process occurs which converts these immature virions into mature virus particles. During maturation process, which also involves morphological rearrangements, viral enzyme HIV-1 protease cleaves these polyproteins into mature functional proteins and enzymes [Navia and McKeever 1990, Adamson and Freed 2007]. The RNA genome and various structural proteins then assemble to produce a mature HIV-1 virion.

- HIV-1 protease plays a crucial role of processing viral Gag and Gag-Pol polyproteins into functional viral enzymes and structural proteins. This process, called viral maturation is essential for the production of infectious viral particles.

Inhibition of HIV-1 protease at this stage leads to the formation of immature and non-infectious virus particles and spreading of HIV-1 is therefore stopped. Several HIV-1 protease inhibitors have been developed and are in use for the treatment of HIV/AIDS.

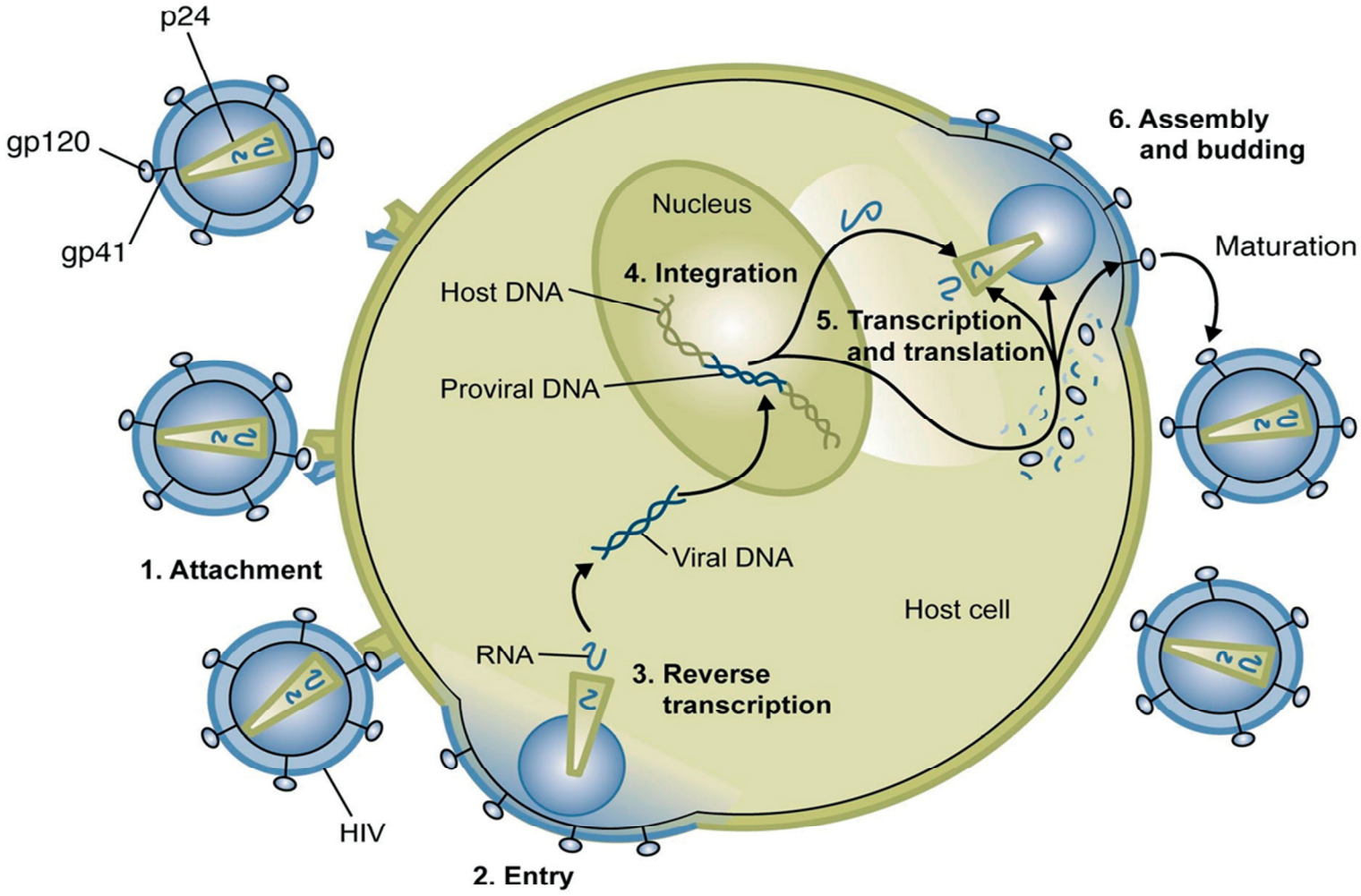

Figure 1.6: Schematic representation of various steps in HIV-1 life cycle (Reproduced from educational module of Virco Lab HIV resistance learning system with permission from Virco Lab Inc, USA).

Status of Treatment for HIV/AIDS

Current drugs do not cure or completely eradicate HIV-1 from the body of HIV/AIDS patient, but they are successful in delaying onset of AIDS after HIV-1 infection thereby prolonging the life of patient. These antiretroviral drugs are classified into six categories based on their target protein.

- Nucleoside analogue reverse transcriptase inhibitors
- Non-nucleoside reverse transcriptase inhibitors
- Protease inhibitors
- Integrase inhibitors
- Fusion inhibitors
- Co-receptor binding inhibitors

Current treatment for HIV-1 infection is known as Highly Active Anti-Retroviral Therapy (HAART), which consists of combinations/cocktails of at least three different drugs belonging to at least two different classes of antiretroviral drugs [Bonfanti P et al 1999, Shafer and Vuiton 1999]. HAART has proven to be very useful in treating HIV/AIDS patient and has produced tremendous results. HAART has reduced the severity of HIV-1 infection and has improved the quality of life of HIV/AIDS patients [Palella et al 1998]. Average life expectancy of HIV/AIDS patient is increased to 32 years by appropriate HAART therapy whereas in the absence of HAART, a HIV-infected patient dies in approximately 11 years [Morgan et al 2002, Schackman et al 2006]. According to an estimate, about 5.2 million patients were receiving antiretroviral therapy in 2009. HAART has contributed to decreasing AIDS related deaths over last few years

[UNAIDS 2010]. Estimates suggest that worldwide, about 14.4 million life-years have been gained due to antiretroviral therapy.

Protease inhibitors are an integral part of the HAART. The efficacy of HAART has improved significantly after introduction of HIV-1 protease inhibitors in 1995. HIV-1 protease inhibitors were instrumental in decreased morbidity and prolonged life expectancy of HIV/AIDS patients. Most of these protease inhibitors were developed using structure based drug-designing method, and represent one of the best examples of structure guided drug design strategy.

HIV-1 Protease

Soon after the discovery of HIV-1, Ratner et al [1985] by comparing genomic sequences of HIV-1 and other retroviruses predicted that HIV-1 genome encodes a protease. Same year, based on the presence of signature sequence (D-T-G) Toh et al [1985] suggested that this protease might be an aspartyl protease. The nature of HIV-1 protease as aspartyl protease is further supported by its inhibition by pepstatin [Hansen et al 1988, Seelmeier et al 1988, Darke et al 1989, Richards et al 1989], a known inhibitor of eukaryotic aspartyl proteases and inactivation by mutation of putative active site aspartases [Katoh et al 1985, Kohl et al 1988, Seelmeier et al 1988, Loeb 1989]. Inhibition of HIV-1 protease results in the production of immature and noninfectious virus particles. HIV-1 protease has wide specificity as it cleaves at nine different cleavage sites on Gag and Gag-Pol polyprotein to produce mature proteins. These cleavage sites are quite diverse in their sequences. One of the unique features of HIV-1 protease is that three of its nine cleavage sites contain a proline residue at C-terminus of the scissile peptide bond.

Structure of HIV-1 Protease

Initially, homology modelling using known structures of eukaryotic protease and RSV protease as template was used to study the three dimensional structure of HIV-1 protease [Pearl and Taylor 1987, Pechik et al 1989, and Weber et al 1989]. These studies suggested HIV-1 protease as homodimer, whereas eukaryotic proteases were monomers. Soon, several laboratories have independently determined the crystal structures of wild type HIV-1 protease [Lapatto et al 1989, Navia et al 1989, Wlodawer et al 1989, Spinelli et al 1991]. First crystal structure of recombinant HIV-1 protease was determined by Merck group to a resolution of 3 Å [Navia et al 1989]. These structures confirmed that HIV-1 protease is a homodimeric aspartyl protease consisting of two non-covalently associated chemically identical subunits of 99 amino acid residues (Figure 1.7) [Lapatto et al 1989, Navia et al 1989, Wlodawer et al 1989, Spinelli et al 1991]. The active site is situated in the cleft formed at the interface of the two subunits. Each monomer contributes one catalytic aspartyl residue in the active site. Overall structure of HIV-1 protease is primarily made up of β-sheet with a single α-helix from residue 86 to 94. The conserved active site triad (D25-T26-G27) is located in a wide loop at the floor of active site, and is stabilized by a network of hydrogen bonds known as "fireman's grip" [Davies 1990]. The carboxyl groups of catalytic aspartates are nearly coplanar and OD1 atoms of these two groups are in close contact with a separation of 2.6-2.9 Å [Das et al 2006]. In fireman's grip network, T26 forms two hydrogen bonds with the atoms from opposite monomer chain. OG1 atom of T26 accepts a hydrogen bond from main-chain NH of T1026 and also donates a hydrogen bond to carbonyl O atom of L1024 [Wlodawer and Vondrasek 1998]. Each monomer also contains a flexible β-ribbon known as 'flap',

which forms the roof of the active site and is also involved in substrate binding. These flaps are very dynamic and show approximately 7 Å movement at their tips to facilitate entry and proper positioning of substrates into the active site [Miller et al 1989, Gustchina and Weber 1990]. Dimer interface is mainly contributed by a four-stranded antiparallel β-sheet formed by amino (residues 1-4) and carboxy termini (residues 96-99) of each subunit. The arrangement of four-stranded antiparallel β-sheet is such that the two carboxy terminal β-strands are sandwiched between the two amino terminal β-strands. Also carboxy terminus of first monomer is just 5 Å away from amino terminus of second monomer. The dimer interface is also stabilized by interactions in the flap region and in the active site region.

Tethered Dimer of HIV-1 Protease

Crystal structures of HIV-1 protease revealed that amino and the carboxy termini of the two monomers are interlocked to form the dimer interface. This arrangement brings carboxy terminus of one monomer very close to amino terminus of the other. This allowed researchers to join the two monomers by a flexible linker to produce HIV-1 protease as a tethered dimer. Tethered dimer is expressed as a single polypeptide chain in which two copies of the monomeric genes are linked by a short linker. Several forms of tethered HIV-1 protease with various linker chain lengths and sequences have been explored [Dilanni et al 1990, Bizub et al 1991, Bhat et al 1994, Griffiths et al 1994, Cheng et al 1999]. Biochemical studies have shown that there are no major differences in activities of tethered HIV-1 protease as compared to the native HIV-1 protease, though it has been shown that tethered HIV-1 protease is more stable than native HIV-1 protease [Cheng et al 1999]. The crystal structure of a tethered dimer complexed with an inhibitor

has also shown that there is no significant change in structure of HIV-1 protease due to presence of the tether [Bhat et al 1994]. The major thrust behind tethered dimer gene was to introduce asymmetric modifications and produce heterodimers for structural studies [Dilanni et al 1990, Bhat et al 1994, Tozser et al 1997, Cheng et al 1999].

Structural Investigation of Inhibitor Complexes of HIV-1 Protease

Since HIV-1 protease was shown to be a prime target for anti-HIVdrugs, a large number of crystal structures of HIV-1 protease in complex with different types of inhibitors have been determined. These structures provided valuable information about the interactions of the inhibitors in the active site of the enzyme and helped in designing drugs against HIV-1 protease. The first such structure was that of inhibitor MVT-101 [Miller et al 1989]. Comparison of these structures with unliganded HIV-1 protease structures had shown that the binding of inhibitors induces substantial conformational changes into HIV-1 protease. Largest movement is in the flap region whose tips move approximately 7 Å to close over inhibitor bound in the active site [Miller et al 1989]. These inhibitors bind in an extended conformation forming hydrogen bonds with both active site and flap region of the enzyme. From these structures, a number of subsites that accommodate corresponding side chain residues of inhibitors or substrates can be identified in the active site of HIV-1 protease. These subsites are described using the Schechter and Berger nomenclature (Figure 1.8) [Schechter and Berger 1967]. Under this nomenclature subsites on N-terminal side of scissile peptide bond are unprimed while those on C-terminal side are primed. Substrate/inhibitor residues on either side of the scissile bond are labelled as: P1 and P1', the next two as P2 and P2', and so on. Corresponding subsites in the protease are marked as S1 and S1', S2 and S2', and so on. Three subsites on each

side (S1-S3 and S1'-S3') are better defined in HIV-1 protease, whereas other distant subsites are not very clear (Figure 1.8). S1 and S1' subsites are composed of mostly hydrophobic residues except catalytic aspartates. Inhibitors were designed to have bulky hydrophobic moieties at P1 and P1' sites which enable them to make extensive hydrophobic interactions with corresponding hydrophobic subsites. S2 and S2' subsites are also hydrophobic but these subsites can accommodate both hydrophobic and hydrophilic residues. Because of this feature, various kind of groups are tried at P2 and P2' subsites in the designing of HIV-1 protease inhibitors. Subsites distal to S2 and S2' are not very specific and can accommodate various types of residues.

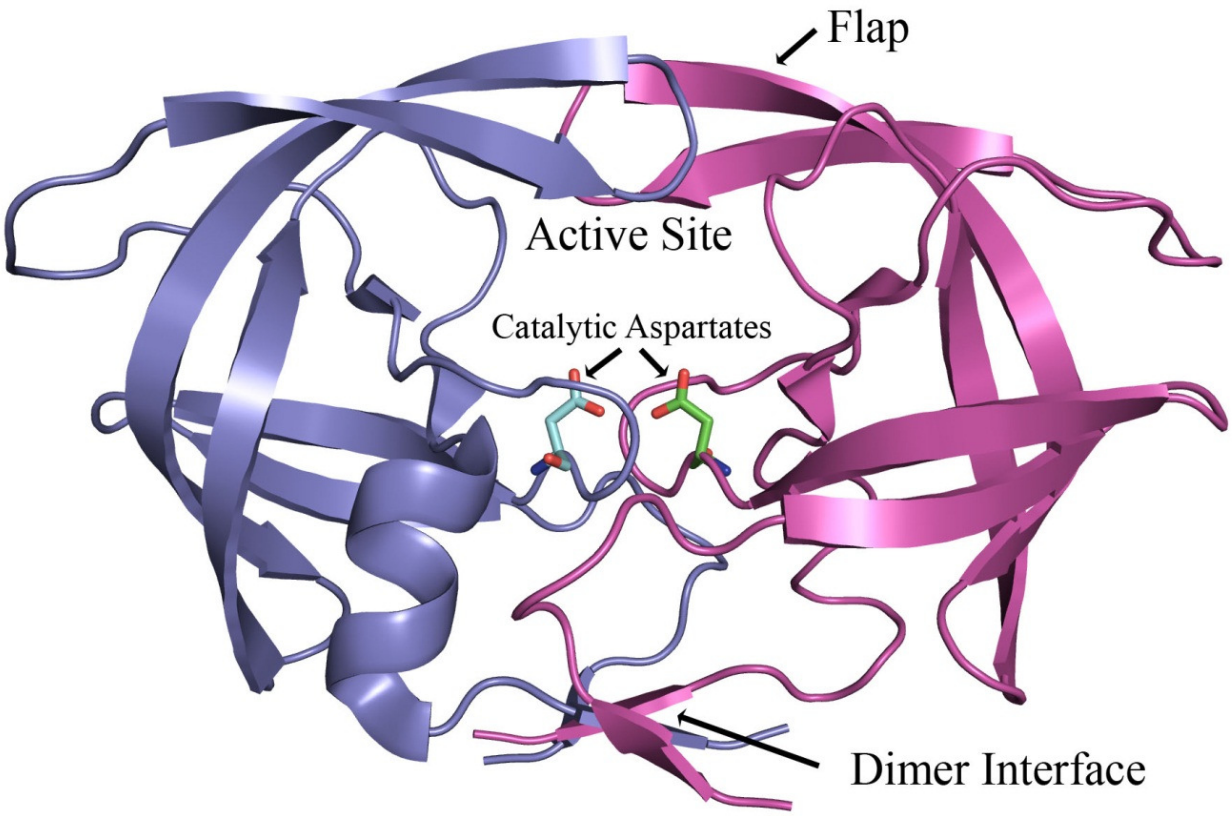

Figure 1.7: A cartoon drawing of the HIV-1 protease dimer in 'closed-flap' conformation with the two subunits coloured slate and magenta. Catalytic aspartates from both monomers are also shown. Active site, dimer interface and flap regions are marked. One hydrogen bond holds the two flaps together.

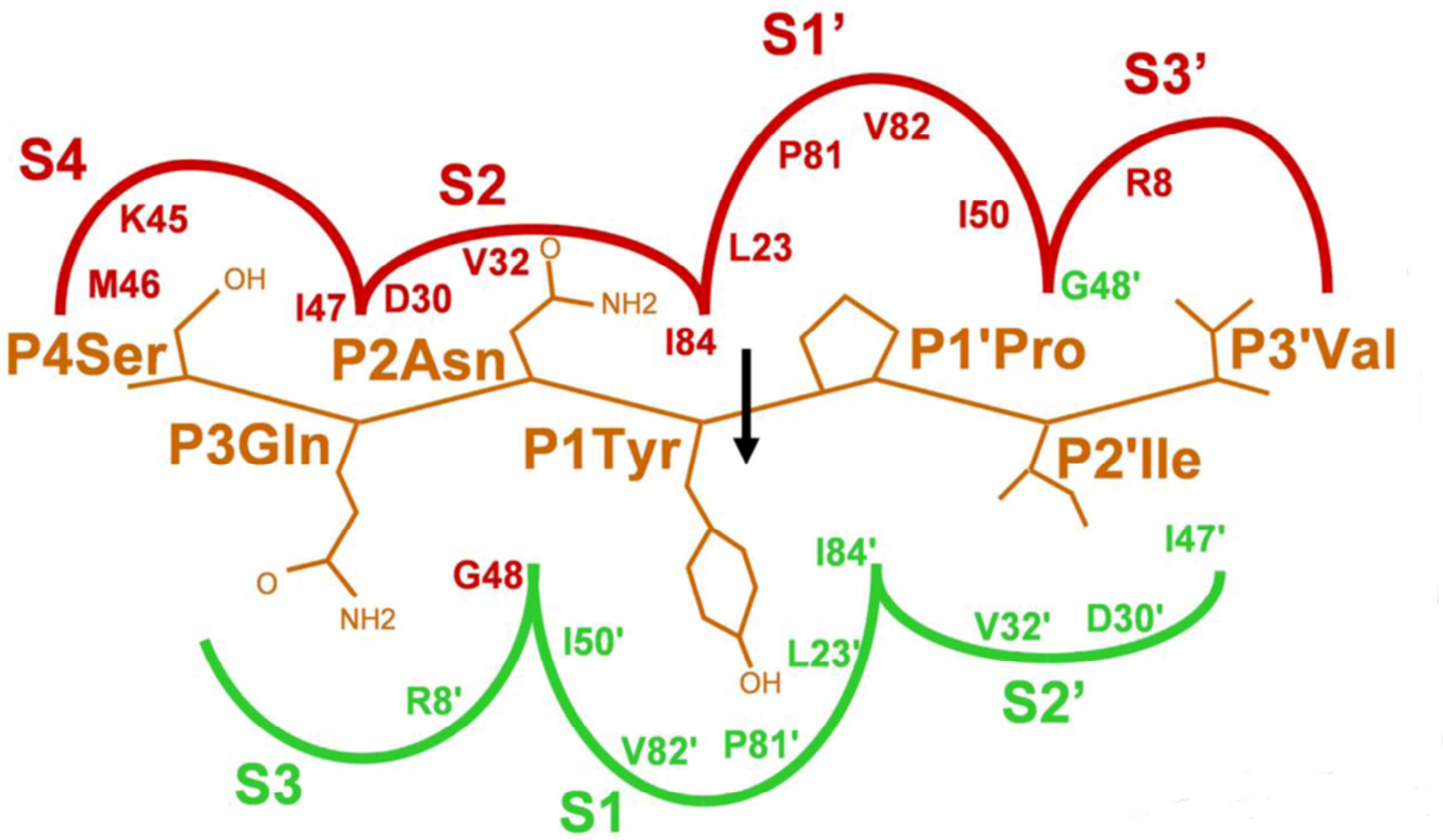

Figure 1.8: Schematic representation of different subsites of the HIV-1 protease: substrate shown represents MA-CA cleavage site and was modeled into the binding site of the crystal structure of the protease. The relative size of each subsite is also indicated approximately by the area enclosed by the curved line (Reproduced from Tozser J 2010, Viruses 2: 147-165).

Development of Protease Inhibitors

Protease inhibitors are one of the most potent agents developed to inhibit HIV-1 replication. Currently, there are nine FDA approved protease inhibitors available for the treatment of HIV/AIDS. All these inhibitors are competitive inhibitors. These drugs were developed to mimic the hypothetical tetrahedral transition state. All these inhibitors, except tipranavir are peptidomimetics. A common feature of all these drugs is the presence of a central hydroxyl group, which interacts with catalytic D25/1025 residues of HIV-1 protease and is required for tight binding in the active site. Binding of these peptidomimetic inhibitors also involves a unique water molecule, which is approximately tetrahedrally coordinated between the flap and the inhibitors (Figure 1.9). This water molecule, popularly known as flap water, links the amide N atoms of I50/I1050 in the flaps to the P2/P1' carbonyl O atoms of the inhibitors.

Development of HIV-1 protease inhibitors is a dream story of successful structure guided drug design approach. Initial HIV-1 protease inhibitors were developed using substrate cleavage sites at Gag and Gag-Pol polyproteins as template. During the development of inhibitors, scissile peptide bond is replaced with a noncleavable bond. Second round of HIV-1 protease inhibitor design was based on the transition state model. Several transition state analogs e.g., statine, norstatine, hydroxyethylene, dihydroxyethylene, hydroxyethylamine etc were developed (Figure 1.10). Crystal structures of HIV-1 protease in complex with these inhibitors have provided valuable information about interactions of inhibitors in the active site of the enzyme and allowed researchers to optimize them, thereby improving the efficacy of these inhibitors.

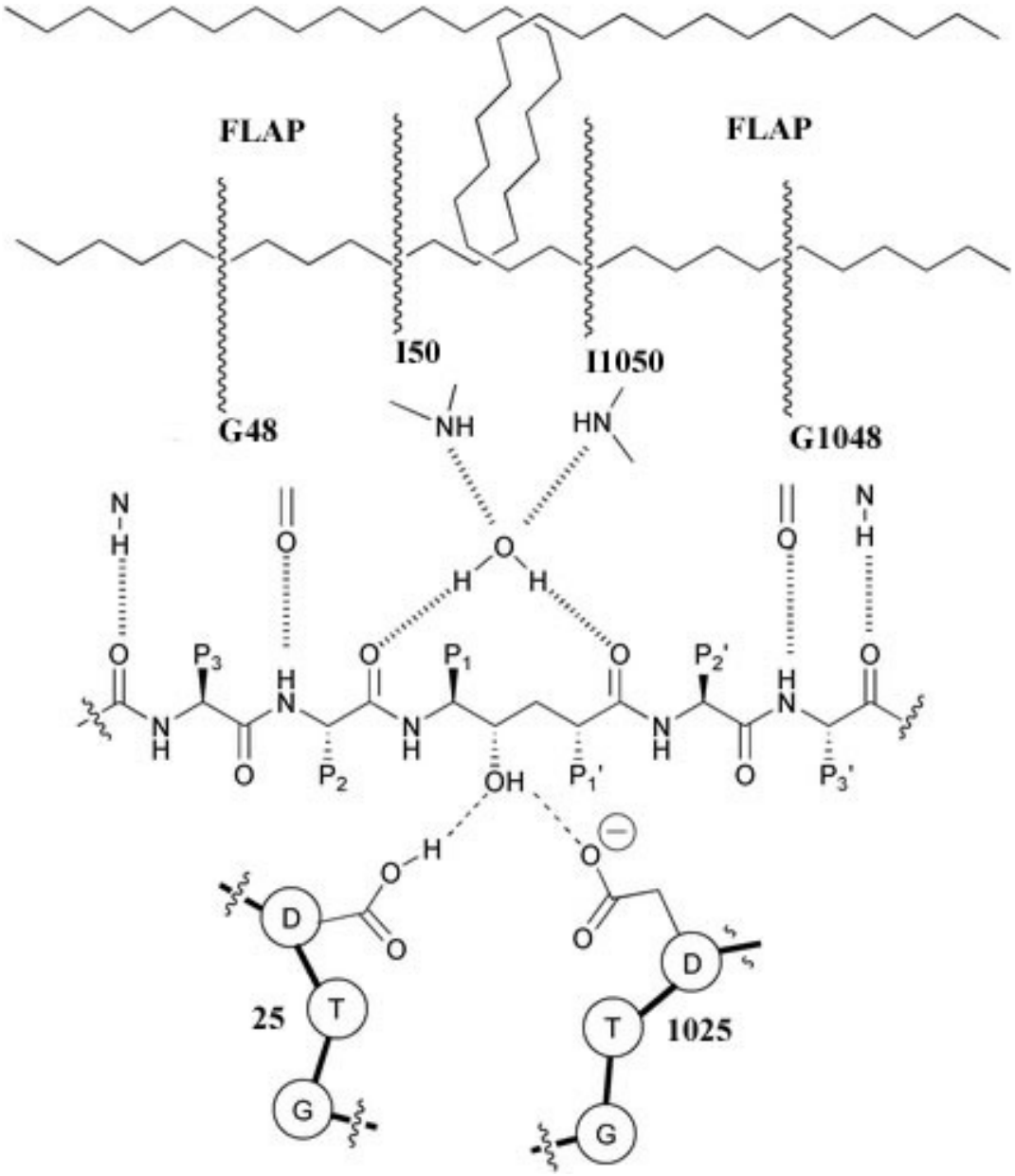

Figure 1.9: Schematic representation of flaps of HIV-1 protease and their hydrogen bonds with the water molecule. The same water molecule makes another two hydrogen bonds with the peptidyl inhibitor. Interactions of the central hydroxyl group of the inhibitor with the catalytic aspartates of HIV-1 protease are also shown (Brik and Wong 2003 Org Biomol Chem 1: 5-14. Adapted by permission of The Royal Society of Chemistry).

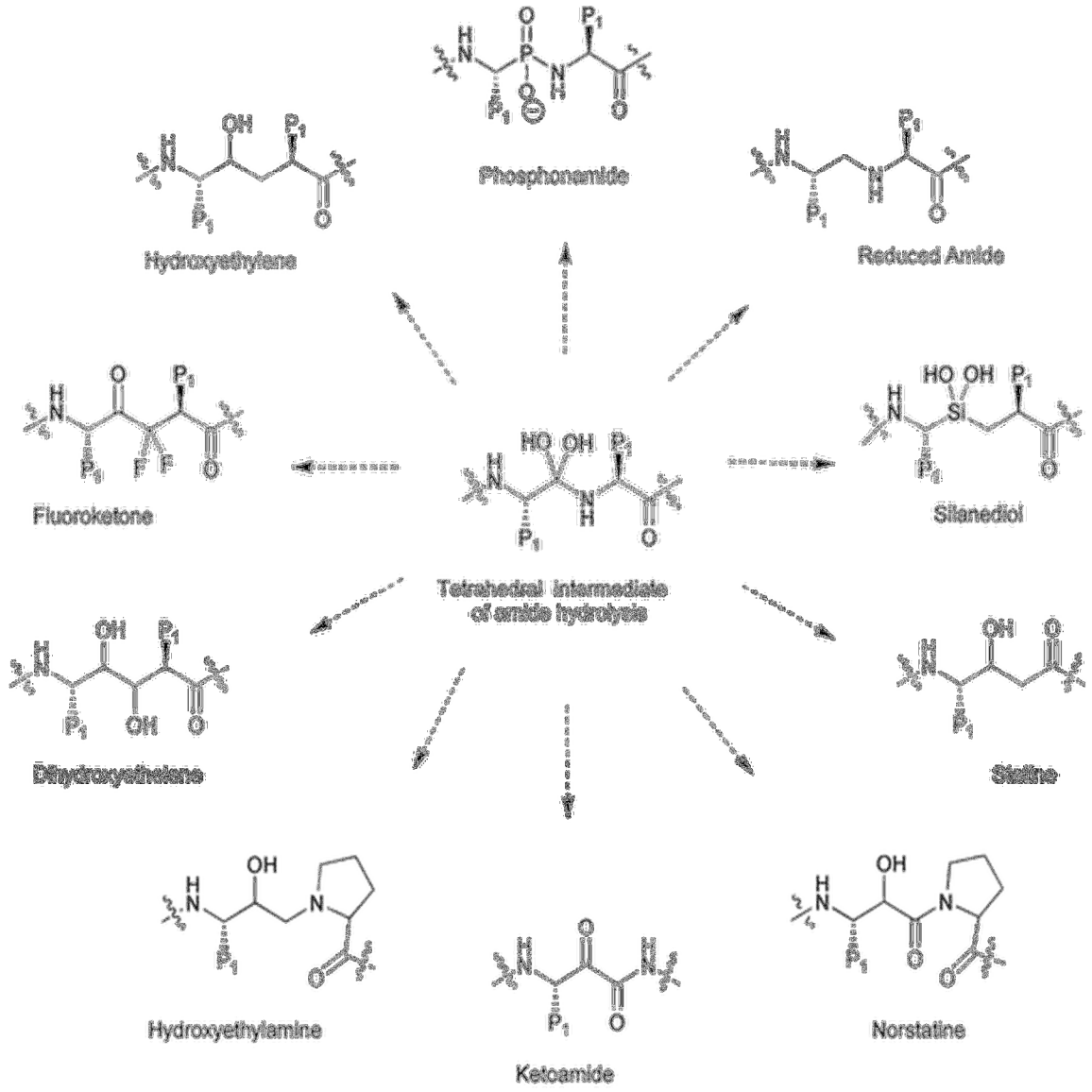

Figure 1.10: Noncleavable transition-state isostere used during the development of the HIV-1 protease inhibitors (Brik and Wong 2003 Org Biomol Chem 1: 5-14. Reproduced by permission of The Royal Society of Chemistry).

Saquinavir was the first HIV-1 protease inhibitor approved by FDA in 1995 for the treatment of HIV/AIDS. Saquinavir had a good potency with a Ki = 0.12 nM against HIV-1 protease. The design of saquinavir (SQV, Ro-31-8959, Invirase, marketed by Hoffman-La Roche) was based on the ability of HIV-1 protease to cleave specifically at Phe-Pro and Tyr-Pro sites in polyprotein substrate. During development of saquinavir, a non-cleavable hydroxyethylamine isostere was used to replace scissile peptide bond. Also, P1' proline was replaced with a large (S,S,S)-decahydroisoquinoline-3-carbonyl group (Figure 1.11). A tert-butyl group was added at P2' position. P1 and P2 sites were similar to the peptide substrate and a quinoline group was introduced at P3. Crystallographic studies of saquinavir had shown that it binds in an extended conformation with a characteristic set of hydrogen bonds with both the main chain and side chain of HIV-1 protease [Krohn et al 1991].

Soon, another protease inhibitor, ritonavir (RTV, ABT-538, Norvir) developed by Abott laboratories, was approved for clinical use [Kempf et al 1995, Boden and Markowitz 1998]. RTV is a potent inhibitor of HIV-1 protease with Ki of 0.022 nM [Kempf et al 1995]. Development of ritonavir had extensively used structure based drug design method. Ritonavir was initially developed on the line of C2 symmetric inhibitors [Erickson et al 1990], though final compound developed was asymmetric (Figure 1.12). Because of side effects, RTV is not used as a protease inhibitors. However, RTV has shown to be an inhibitor of Cytochrome P-450 [Kempf et al 1997] and therefore, RTV is currently used as a booster drug along with other protease inhibitors in anti retroviral therapy.

Indinavir (IDV, MK-639, L-735,524, Crixivan) was developed by Merck & Co, in a long and complex research project with more than 150 intermediate compounds described in the literature [Wlodaver and Vodrasek 1998]. A novel hydroxyaminopentaneamide dipeptide isostere was used to replace scissile peptide bond in indinavir (Figure 1.13). IDV has Ki against HIV-1 protease in the range of 0.3 -0.6 nM and has excellent oral bioavailability.

In 1997, nelfinavir (NFV, AG-1343, Viracept), was developed by Agouron Pharmaceuticals using protein-inhibitor co-crystal structure analysis and Monte Carlo simulations [Gehlhaar et al 1995]. Nelfinavir has Ki of 2 nM despite its lower molecular weight compared to other protease inhibitors [Kaldor et al 1997]. It was the first protease inhibitor approved for use in pediatric AIDS patients. Its P1' and P2' sub-sites are identical to that of saquinavir (Figure 1.14). At P1 site phenyl group was replaced with a unique S-phenyl group. It also incorporated a novel non-peptidic 2-methyl-3-hydroxybenzamide group replacing the P2 asparagine in SQV [Kaldor et al 1997].

Another drug amprenavir (APV, 141W94 or VX-478) developed by Vertex Pharmaceuticals and GlaxoSmithKline, is based on SQV backbone, with the P2 asparagine group replaced by tetrahydrofurancarbamate and the P1´-P2´ moiety by a sulfonamide derivative (Figure 1.15). APV is also approved as a prodrug, fosamprenavir. APV has a Ki value of 0.11 nM against HIV-1 protease [Klabe et al 1998] and has good bioavailability.

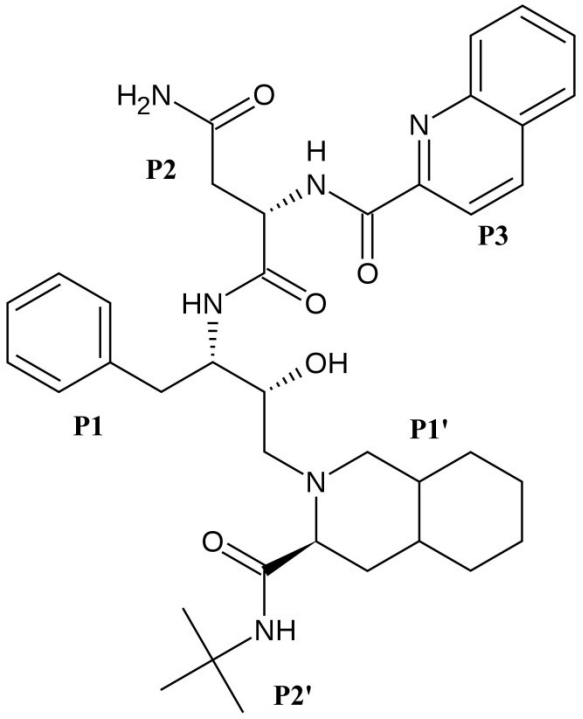

Figure 1.11: Saquinavir

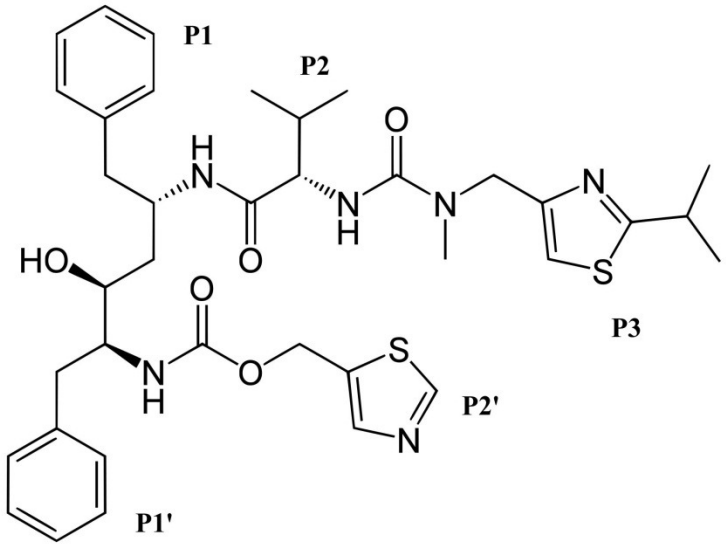

Figure 1.12: Ritonavir

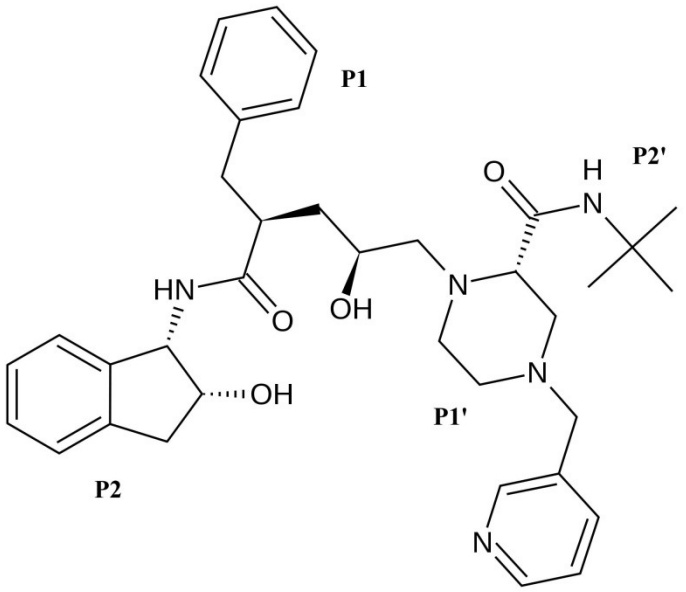

Figure 1.13: Indinavir

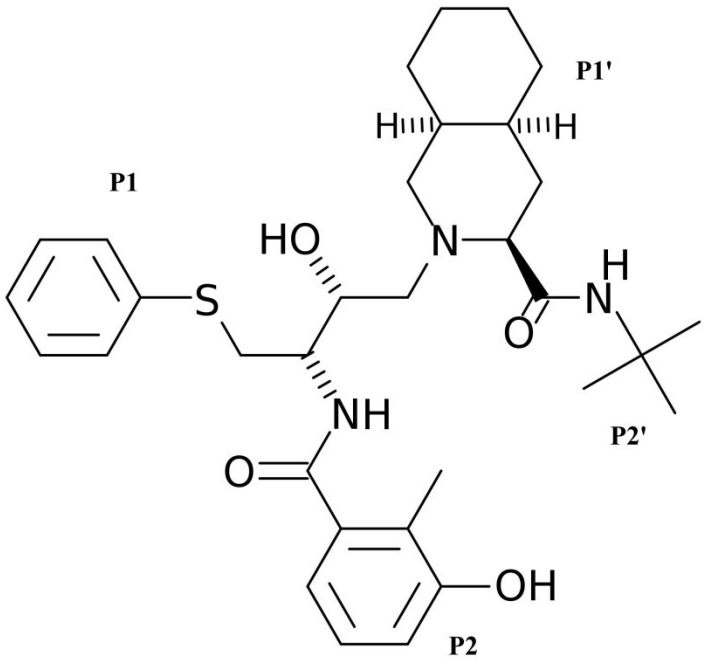

Figure 1.14: Nelfinavir

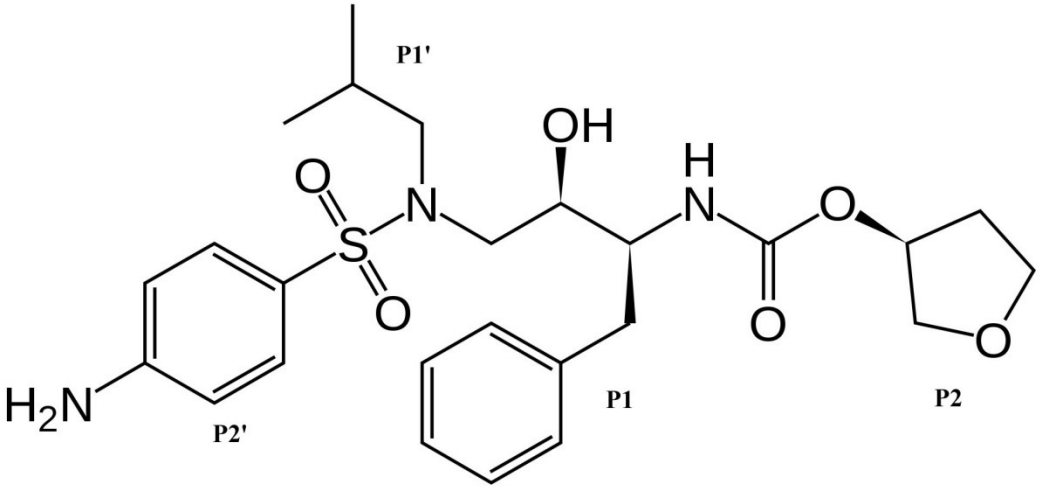

Figure 1.15: Amprenavir

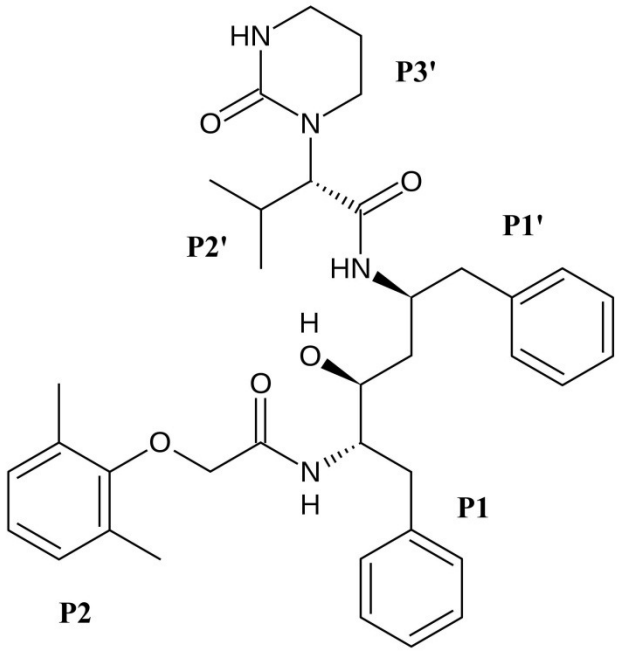

Figure 1.16: Lopinavir

Lopinavir was developed by Abbott, based on the first generation protease inhibitor RTV. Lopinavir has Ki of 0.0013 nM against HIV-1 protease with [Sham et al 1998]. The bulky (2-isopropylthiazolyl) methyl P3′ moiety was replaced with smaller cyclic urea group (Figure 1.16). A new dimethylphenoxyacetyl moiety was also added to P2 site.

Atazanavir developed by Bristol-Myers Squibb was approved by FDA in 2003. It has a novel (hydroxyethyl) hydrazine dipeptide isostere (Figure 1.17) [Robinson et al 2000]. It has very high potency against wild-type HIV-1 protease with a Ki value of 0.01 nM [Gong et al 2000].

Tipranavir (TPV) is developed by Boehringer Ingelheim from lead compounds hydroxycoumarin and dihydropyrones which were identified through high throughput screening (Figure 1.18) [Turner et al 1998]. Tipranavir has a Ki of 0.005 nM against HIV-1 protease [Turner et al 1998]. Tipranavir is the only nonpeptidic protease inhibitor approved for HIV/AIDS treatment. TPV is not a transition state mimetic. It binds to the protease in a distinct fashion and replaces conserved flap water. TPV potently inhibits multidrug-resistance protease variants.

Darunavir (DRV) [Surleraux et al 2005, Koh et al 2003, De Meyer et al 2005], developed by Tibotec Inc, is latest protease inhibitor approved by the FDA. It's development was based on the backbone of drug amprenavir. A novel bis-tetrahydrofuranyl (bis-THF) moiety at P2 replaced single tetrahydrofuranyl (THF) ring of APV in DRV. This bis-THF moiety makes novel interactions with the main chain atoms in S2 binding pocket of HIV-1 protease (Figure 1.19). These interactions make DRV a highly potent inhibitor of HIV-1 protease with a Ki = 0.015 nM.

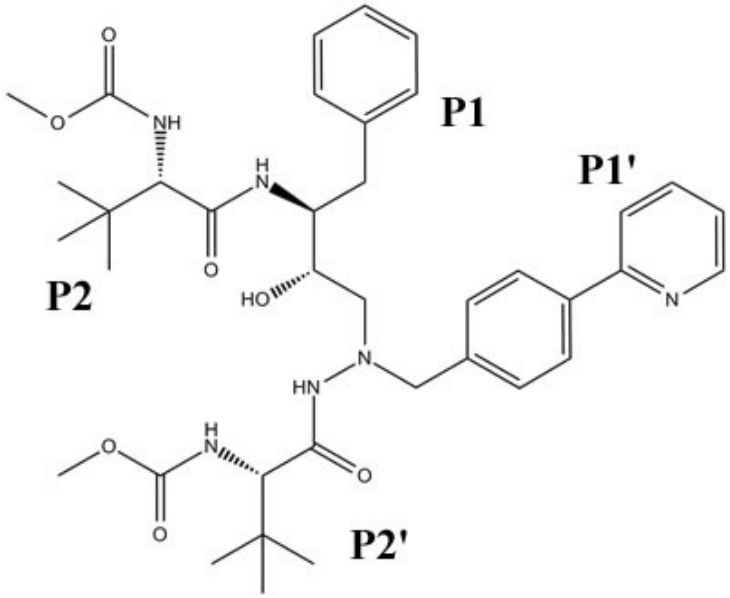

Figure 1.17: Atazanavir

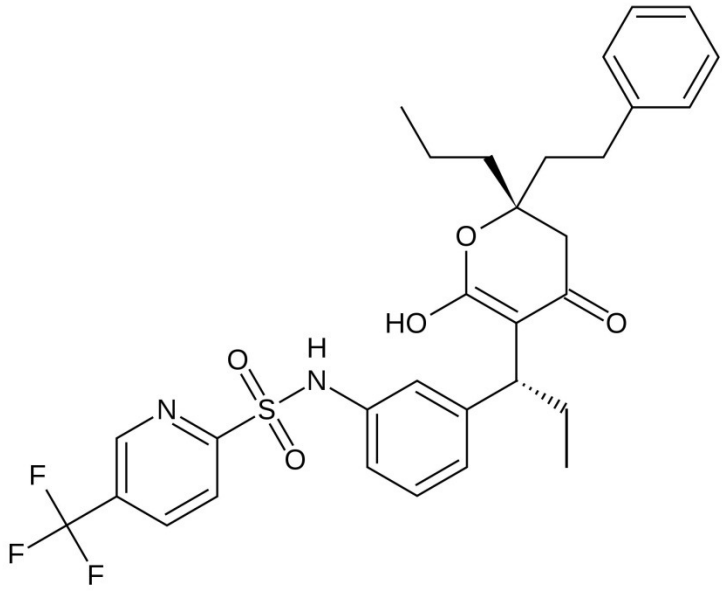

Figure 1.18: Tipranavir

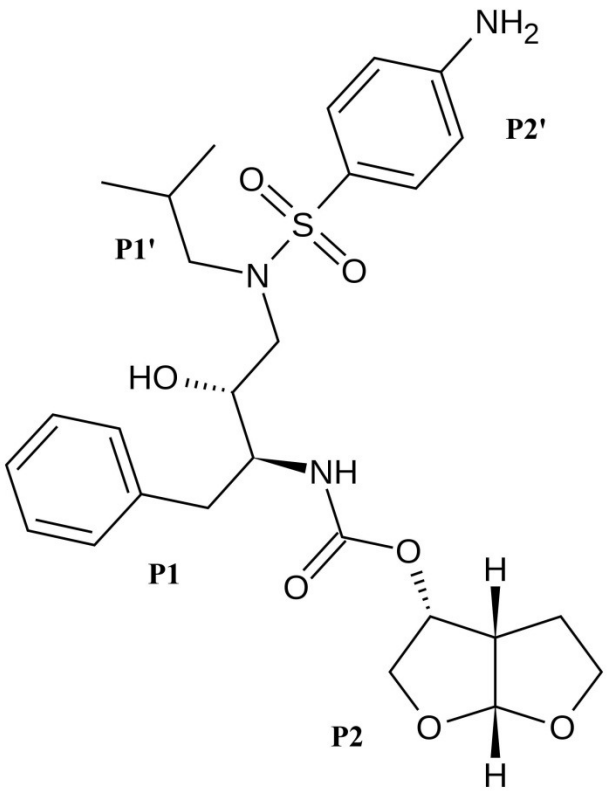

Figure 1.19: Darunavir

HIV Drug Resistance

The HAART and HIV-1 protease inhibitors in particular have shown remarkable effect in slowing down the progression of HIV-1 infection into AIDS, thus improving quality of life and survival years of HIV patients [Palella et al 1998]. However, emergence of drug resistant mutations limits the efficacy of antiretroviral inhibitors [Ledergerber et al 1999, DeGruttola et al 2000]. These drug resistant mutations are a major impediment in long-term success of antiretroviral treatment. HIV-1 is characterized by high mutation rate which is caused by error prone reverse transcriptase [Preston et al 1988, Roberts et al 1988, Bebenek et al 1993], recombination [Robertson et al 1995, Sharp et al 1995], and potential superinfection [Levy 2009]. High mutation rate coupled with very fast multiplication rate means that patient's blood will have a population of HIV with large diversity. Every day about 10^9 new virus particles are produced in a fully infected HIV/AIDS patient [Ho 1997] and rate of replication errors is 10^{-3} to 10^{-5} [Coffin 1995, Sarafianos et al 2009]. This results into a virus population in which most of the new virus particles are different from the parent virus at-least by one base. Most of these new viruses end up with one or more deleterious mutations that prevent them from being able to carry out new infections. However, some of these mutated viruses are able to infect new host cell and produce new copies of HIV. Many of these viruses have mutation/s in HIV-1 protease gene. These mutations, which allow the protease to carry out its essential functions but affect the binding of inhibitors, make the virus harboring mutated protease resistant to HIV-1 protease inhibitors. High rates of viral turnover rate will help in producing high populations of these resistant viruses in the presence of selection pressure of the drugs. The first clinical drug resistance case was published in USA in 1989 [Larder

and Kemp 1989]. Later, multiple drug resistance mutations in the reverse transcriptase against the drug azidothymidine (AZT) were observed [Larder and Kemp 1989]. Drug resistance mutations against all known HIV-1 protease inhibitors have been observed in HIV/AIDS patients who failed to respond to drug therapy. The pattern of HIV-1 protease mutations associated with the drug resistance is extremely complex.

Aims and Objectives

There is tremendous progress in the knowledge of HIV-1 biology and several drugs are now available for the treatment of HIV/AIDS. However, for several reasons we need to continue the efforts to develop novel, more potent compounds. Emergence of drug resistance mutations in the targeted proteins is one such reason. Other reason to continue work on HIV-1 protease is that currently available protease inhibitors are expensive and have problems of tolerability, side effects, and adherence. HIV/AIDS is a disease which primarily affects poor nations and poor people in rich nations. High prices of available inhibitors are affecting the reach of these drugs to the patients in developing and poor countries. Out of 33 million people infected with the HIV, only 5.2 million people were receiving antiretroviral therapy in 2009. So there is an urgent need for effective, affordable HIV-1 protease inhibitors, which will increase the reach of the anti-retroviral therapy. Taken together, in-spite of the tremendous success of the HAART, novel drugs including novel HIV-1 protease inhibitors are required to achieve better control of HIV infection.

The specific aims of present work are as follows:

Aim 1: To determine & analyse crystal structures of HIV-1 protease in complex with oligopeptide substrate to gain insight into the molecular mechanism of HIV-1 protease.

Aim 2: To determine & analyse crystal structures of protease harboring mutations selected against the drug nelfinavir to gain insight into the mechanism of resistance against nelfinavir.

Chapter - 2

Material and Methods

There is a direct relationship between three-dimensional structures of protein molecules and their functions. Understanding of this structure-function relationship has the potential applications in various fields ranging from basic science to pharmaceutical industries. Three different methods are used to determine 3-dimensional structures of biological macromolecules: X-ray crystallography, Nuclear magnetic resonance spectroscopy, and Cryo-electron microscopy. X-ray crystallography was the first method to be employed to solve structures, and even now is the first choice. X-rays were discovered by Conrad Rontgen in 1895. Subsequently Max von Laue proposed that the X-rays were waves with wavelength comparable to the spacing between atoms in crystals. He therefore, suggested that the crystals should diffract X-rays. Following this, the first X-ray diffraction experiment on the crystal of copper sulfate was performed by his students, Walter Friedrich and Paul Knipping in 1912. Laue also provided theoretical basis of diffraction from a crystal. For his discovery, Laue was awarded the 1914 Nobel Prize in Physics. William Bragg and Lawrence Bragg (father and son) gave a simple but elegant explanation of X-ray diffraction by crystals, and showed that structures of molecules can be determined using X-ray diffraction. They solved the structure of NaCl using X-ray crystallography. For their classical work in X-ray crystallography, they received the 1915 Nobel Prize in Physics. Initially, X-ray diffraction was used for the determination of structures of small inorganic and organic compounds. Since proteins could also be crystallized, it was suggested that crystal structures of proteins could also be determined using X-ray crystallography. However, it took 45 years from the time of the Braggs' work to determine first crystal structure of a protein. J.D. Bernal and Dorothy Hodgkin recorded the first X-ray diffraction pattern from crystals of pepsin in 1934. First structure

was determined only in 1958 when structure of myoglobin was solved by John C Kendrew [Kendrew et al 1958]. Subsequently, structure of hemoglobin was also determined by Max F Perutz [Perutz et al 1960]. The 1962 Nobel Prize in Chemistry was awarded jointly to Max F Perutz and John C Kendrew for their work on protein structures. Since then several other scientists working in the field of protein crystallography have been awarded the Nobel Prize. Recently, Venki Ramakrishnan, Thomas A. Steitz and Ada Yonath were awarded 2009 Nobel prize of Chemistry for the determination of crystal structure of the ribosome.

Since the first structure of myoglobin, there is an exponential increase in the number of protein structures, particularly in structures determined by X- ray diffraction. Crystal structure determination is based on the diffraction of X-rays by electrons of the atoms, arranged periodically in a crystal. Diffraction pattern from a protein crystal is observed as numerous discrete spots on the detector frame. Using intensities of each of these spots obtained from the diffraction pattern and phases obtained from additional methods, the electron density can be reconstructed using Fourier synthesis. A molecular model representing the structure of the protein molecule is then built into electron density map. Therefore, structure determination in protein crystallography broadly involves following steps: 1) Protein purification, 2) Crystallization, 3) X-ray diffraction data collection and processing, 4) Phase determination, 5) Model Building and refinement, and 6) Structure validation and analysis. In this chapter, I will discuss these steps with special emphasis on methods that were used in determining the structures reported in the present thesis. However, I will first describe the gene construct used in the present study and molecular biology experiments carried out to prepare mutant proteins.

Tethered Dimer Gene Construct

A HIV-1 protease tethered dimer gene construct was used in present study. In this construct, two protease genes are linked in frame by a linker region having codons for G-G-S-S-G (Figure 2.1). This construct has additional 57 amino acids at amino terminus linked to protease with a Phe-Pro cleavage site. The flanking peptide is a part of the amino terminal polyprotein of the *pol* gene.

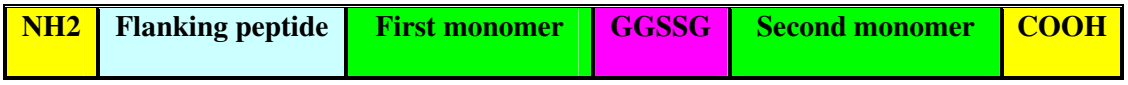

Figure 2.1: A schematic diagram of the tethered dimer of HIV-1 protease constructed by linking two monomers with a 5 residue linker peptide and a 57 amino acids flanking peptide at amino terminus.

Due to the presence of a flanking peptide, the enzyme is initially expressed as a 29 kDa precursor. Autoprocessing at the Phe-Pro cleavage site yields 22 kDa mature HIV-1 protease dimer. This construct allows for easy and rapid determination of activity of the enzyme. Amino acid residues in first subunit are numbered from 1 to 99 and those in second subunit are numbered from 1001 to 1099. Linker residues are numbered 101-105.

Site Directed Mutagenesis

All the mutations were introduced in tethered HIV-1 protease gene using Stratagene's quick-change multi-site-directed mutagenesis method [Stratagene, La Jolla, CA]. The method is based on the amplification of whole plasmid using mutagenic primers and DNA polymerase enzyme lacking primer displacement activity. The mutant strands are synthesized using a thermal cycling procedure. This step requires plasmid DNA template, mutagenic oligonucleotide primers, and the enzyme blend including a high fidelity DNA polymerase lacking strand displacement activity like PfuTurbo DNA polymerase and a

thermo-stable DNA ligase. In next step, reaction mixture from the first step is incubated with restriction endonuclease DpnI that digests the parental plasmid DNA template. This helps in easy selection of mutated gene as only the closed circular single stranded mutated strands are left in the reaction mixture. In the final step, appropriate amount of reaction mixture is used to transform the chosen *E.coli* cells. After transformation, mutant closed circular single stranded DNA is then converted into double stranded form by the bacterial cells.

Primer designing

All the mutagenic primers were designed using either PrimerX software [Lapid 2003] or Stratagene's primer design software [Stratagene, La Jolla, CA]. Briefly, following parameters were considered for primer designing:

a) All the mutagenic primers used in one reaction must anneal to the same strand of the template plasmid DNA.
b) Primers should be between 25 and 45 bases in length and the desired mutation should be in the middle of the primer.
c) Primers should have a minimum GC content of 40% and terminate in one or more C or G bases at the 3'-end.
d) Melting temperature (T_m) of the primers should be ≥75°C.

The following formula is recommended by Stratagene to be used for estimating the Tm of primers:

$T_m = 81.5 + 0.41$ (%GC) – (675/N) - % mismatch,

Where,

- *N* is the number of bases in the primer.
- % GC and % mismatch are input as whole numbers.

Reaction conditions for thermal cycling step

Following cycling conditions were used for the mutant strands synthesis:

Segment	Cycles	Temperature (°C)	Time (Minutes)
1	1	95	1
2	25-30	95	1
		55	1
		65-72	4-13
3	1	72	7-10

DpnI digestion of the amplification products

1 µl of DpnI (10 U/µl) was added to the product from thermal cycling step and incubated at 37°C for 2-3 hours to digest the parental plasmid DNA.

Transformation

Mutated HIV-1 protease gene construct was then introduced into *E. coli* XL-1 and BL-21 (DE3) cells. Heat shock method of transformation was used with competent cells prepared by chemical method. *E.coli* cells were inoculated in 10 ml Luria-Bertani (LB) medium and incubated overnight at 37^{o}C on an orbital shaker operated at 90 rpm. Next morning, 1ml from this starter culture was added to larger culture medium (typically 100ml of LB broth) and incubated at 37^{o}C on an orbital shaker operated at 120 rpm. The cells were allowed to grow until desired growth was achieved, as determined

spectrophotometrically by measuring absorbance at 600 nm (OD of 0.3-0.4). The cells were harvested by centrifugation at 5000 x g for 10 min at 4°C. The cell pellet was washed with 0.1 M freshly prepared ice cooled $CaCl_2$ followed by resuspension in 2ml 0.1 M $CaCl_2$. All these steps were carried out on ice. From the suspension, cells (200µl) were transferred into ice cooled microfuge tubes. These are competent cells. Products of site directed mutagenesis reaction (1µl) was incubated with competent cells and incubated on ice for 30 minutes along with appropriate control. The cells were subjected to heat shock by incubating at 42°C for 90 seconds followed by immediate incubation on ice for 2 min. The cells were then mixed with 800µl LB broth and incubated at 37°C for 45 min. The cell suspension (200µl) was plated on LB plate (with ampicillin) along with control and incubated overnight. Only those bacterial cells will grow which have been transformed with plasmid containing desired mutated HIV-1 protease gene. The transformed microbial colonies that appeared on petri plates were grown for sufficient growth (OD of 0.2-0.3). Stocks of recombinant colonies were prepared by adding 25% glycerol in cell culture and stored at –70°C. Plasmids from the transformed cells were isolated using alkaline lysis method and HIV-1 protease gene was sequenced to confirm the presence of mutation.

Protein Expression and Purification

E.coli BL21 (DE3) cells carrying plasmid pET11a with appropriate tethered HIV-1 protease dimer insert were grown in LB/TB broth at 37°C on an orbital shaker operated at 120 rpm till cells reach mid-log growth phase (OD_{600} of 0.5-0.7). At this stage, expression of HIV-1 protease was induced by adding 1 mM IPTG. Cells were then allowed to grow for 3-4 hours, after which they were harvested by centrifugation at 5000

x g for 10 minutes at 4°C. The cell pellet was resuspended in 15 ml ice-cold lysis buffer (20mM Tris, 10mM EDTA and 1% Triton X100). The resuspended cells were lysed by ultrasonication under the ice-cold condition by giving 1 sec on and 2 sec off pulses for a total of 4 minutes of pulse-on time. The lysed homogenous suspension was centrifuged at 10000 x g for 10 minutes at 4°C. HIV-1 protease is accumulated as inclusion bodies as determined by running SDS-PAGE of both supernatant and pellet fractions. Supernatant was then discarded and grayish white pellet of inclusion bodies was obtained. The pellet was resuspended in 15 ml of ice-cold wash buffer, which was same as lysis buffer. The obtained suspension was briefly sonicated under ice-cold condition for 60 seconds using the same procedure. The suspension was centrifuged at 10000 x g for 10 minutes at 4°C. This washing procedure was repeated 2 more times. The protein from the final pellet was extracted with 3-4 ml of ice-cold 67% acetic acid followed by centrifugation at 10000 x g for 15 min at 4°C. The supernatant was diluted by 33 times with ice-cold distilled water and transferred into a dialyzing bag with a molecular weight cut-off (MWCO) of 10 kDa. The diluted protein was dialyzed against 5 liters of cold distilled water at 4 °C with continuous stirring for 3 hours and then incubated in fresh cold distilled water for overnight with continuous stirring. The dialyzed protein sample was then transferred into 3-5 liters of pre-cooled refolding buffer (20 mM PIPES pH 6.5, 100mM NaCl, 1mM DTT and 10% glycerol) and incubated for 5 hours to overnight with stirring at 4°C. Some precipitation was observed which was removed by centrifugation at 10000x g at 4°C for 15 minutes and the purified protein was concentrated by centrifugal concentrators up to the desired concentration. The final concentration (1-5 mg/ml) was obtained by exchanging with 50 mM sodium acetate buffer (pH 4.5) containing 1mM DTT. The

protein purity was determined by running SDS-PAGE with appropriate protein samples and standards.

Protein Crystallization

Good diffraction quality protein crystals are a prerequisite of protein crystallography. However, growing crystals is also one of the most difficult steps and a major limitation in protein crystal structure determination. Crystallization of proteins requires the solubility of protein molecules to be reduced slowly to generate a supersaturation stage so that the protein molecules come out from the solution phase into the solid phase in a controlled manner to form a three dimensional crystal. This is usually done by adding different types of precipitants. Several oragnic additives and salts were used as precipitant. Among them polyethylene glycol (PEG) of different molecular weights and ammonium sulfate are two most commonly used precipitants. The crystallization of proteins is influenced by a number of factors. Concentration, purity and homogeneity of protein sample, type and concentration of precipitant, ionic strength and pH of buffer, temperature, ligands or additives, are some of the factors which can affect outcome of a protein crystallization experiment. Finding the right combinations of these factors which lead to crystallization of protein in question is an empirical process. In a protein crystallization experiment, this is achieved by identifying important components of this set by manipulating each component individually to determine its effect on crystallization. Precise mechanism and theoretical basis of crystallization is not yet fully established. Therefore, it is not possible to predict the crystallization conditions of a particular protein. Despite tremendous progress in protein crystallography, crystallization of protein has largely remained a trial and error procedure, and is normally considered as an art rather than science.

Different methods have been developed to explore chemical conditions for protein crystallization. Some of the more frequently used methods are vapor diffusion, free-interface diffusion, microbatch and dialysis. Among these crystallization techniques, most commonly used method for crystallization of proteins is vapor diffusion method. In vapor diffusion method, a drop containing purified protein, buffer, and precipitant is placed onto a cover slip and then inverted over a chamber containing a larger reservoir volume having same buffer and precipitants. Usually, the drop contains precipitant at half the concentration to that of the concentration of precipitant in the reservoir. Due to the concentration gradient, water vapor diffuses from the drop into the reservoir solution. This in turn increases the concentration of protein and precipitant in the drop. If conditions are right, eventually this causes protein sample inside the drop to attain super saturation level where nuclei can form spontaneously. As the protein comes out of the solution, concentration of protein in the drop decreases from supersaturation level to a more moderate level where these nuclei can grow to form large crystals.

Vapor diffusion method can be set up into two different forms: hanging drop vapor diffusion and sitting drop vapor diffusion. The main difference between the two methods is the orientation of the drop with respect to the reservoir. In hanging drop method, protein drop is suspended over the reservoir chamber. In sitting drop method, the protein drop is placed on some support and kept in upright orientation. Vapor diffusion is a simple and powerful method of setting up protein crystallization experiments. It consumes less protein, so more number of chemical conditions can be explored with a limited amount of protein. It also allows easy monitoring of the progress of

crystallization. Vapor diffusion method is also convenient for crystal manipulation and harvesting. For the present study, hanging drop vapor diffusion method was used.

In the free interface diffusion method, protein solution is placed in direct contact with precipitant solution creating an interface between the two solutions. Protein and precipitant diffuse into one another creating a concentration gradient and crystallization may occur at appropriate protein and precipitant concentration. Free interface diffusion is a method of choice for fine-tuning the crystallization conditions.

In the microbatch method, the protein is mixed with high concentration of precipitant to reach super saturation stage and then left undisturbed. In the microbatch under oil, a small drop of protein mixed with the precipitant is placed under a layer of oil. Generally, paraffin oil is used because it allows little or no evaporation so no change in the concentration of the protein or precipitant takes place. A mixture of paraffin and silicon oil can be used that allow very slow evaporation from the crystallization drop. For the microbatch method to be successful, one must be very close to the condition, optimum for the crystallization.

The dialysis method involves putting protein sample in a small dialysis bag or micro-dialysis buttons. The dialysis bag/button is placed in chamber containing a larger reservoir solution with precipitant. Dialysis membrane allows water and some precipitants to move freely while protein sample is retained inside the bag. In dialysis method, concentration of protein does not change during the course of the experiment.

HIV-1 protease was crystallized by the hanging-drop vapor diffusion method. 1-4 µl of protein solution mixed with equal volume of reservoir solution is placed onto a

siliconized coverslip. The coverslip is then inverted and sealed by silicon grease over the reservoir chamber. Chemical conditions for crystallization of native and mutant HIV-1 proteases were explored by systematically varying: (1) pH (2) type and concentration of buffer (3) type and concentration of precipitant, and (4) additives. Hexagonal rod shaped crystals were obtained with sodium citrate/ sodium dihydrogen phosphate as reservoir buffer (200/100mM, pH 6.2) and ammonium sulfate (1-6% of saturated solution) as precipitant. Mutant HIV-1 proteases were also co-crystallized with the drug nelfinavir. Nelfinavir mesylate was dissolved in 100 % DMSO. Mutant HIV-1 protease was mixed with five molar excess of the drug and incubated at room temperature for 2-3 hours before setting up the crystallization experiments. Crystals of mutant HIV-1 proteases and drug complexes were obtained in two different conditions. One of the conditions was similar to that of unliganded HIV-1 proteases. In a different condition, NaCl (0.5-1.75M) was used as a precipitant with sodium acetate buffer (50 mM, pH 4.5) as reservoir. Effect of different additives was also explored. DMSO (1-10 %) was found to be effective as an additive.

For obtaining co-crystals of wild type HIV-1 protease and substrate oligopeptide soaking method was used. Oligopeptide was dissolved in 100% DMSO. HIV-1 protease crystals were transferred to a drop containing oligopeptide at a final concentration of 1mM in appropriate buffer. For the present study, sodium citrate/ sodium dihydrogen phosphate buffer (200/100mM, pH 6.2) was used. The crystals were incubated in the drop for varying time and then harvested.

X-ray Diffraction Data Collection

In a X-ray diffraction experiment, a crystal is illuminated with X-rays by placing the crystal in the path of a X-ray beam and diffracted X-rays are collected with a suitable detector. Two types of X-ray sources are used in protein crystallography experiments, the conventional laboratory X-ray generator (so-called home source) and the synchrotron radiation source. The crystal is mounted either in a capillary or in a loop and then attached to the goniometer head. The goniometer allows proper positioning of the crystal in the path of the X-ray beam. X-rays scattered by the crystal are recorded onto a suitable detector in the form of diffraction images. Various types of area detectors are in use for protein crystallography. Most commonly used detectors are image plate, CCD and pixel array detectors.

The diffraction of X-rays by crystals is the result of interaction of X-rays with electrons from the atoms of the macromolecules arranged periodically in a crystal. The crystal can only diffract X-rays in certain directions that are recorded on the detector frame as discrete spots known as reflections. The directions of these reflections are given by the Bragg's law:

$$2d_{hkl} \sin\theta_{hkl} = n\lambda$$

where λ is the wavelength, θ_{hkl} is the angle of incidence or reflection, n is an integer and d is the distance between the parallel lattice planes with Miller indices h,k,l. hkl are also the indices for the Bragg reflections. Each atom in the structure contributes to intensity of all the Bragg reflections, and all the Bragg reflections contain information about any given atom in the structure. Bragg considered the crystal to consist of sets of a series of parallel planes and diffracted ray as reflections from these atomic planes. According to Bragg's

law a reflection is produced only when difference in the path length for rays reflected from successive planes is equal to an integral number of wavelength of the impinging X-rays. In all other cases waves reflected from successive planes in the crystal are out of phase causing destructive interference among them and no diffracted beam emerges from the crystal. Thousands of such diffraction spots need to be collected to solve a protein structure. The goal of a X-ray diffraction data collection experiment is to collect as many Bragg reflections as possible. This is mostly achieved by using the rotation method where the crystal is rotated by small angles. A series of diffraction images (also called frames) are collected in which each image contains several reflections which appear during the rotation of crystal through a specified angle.

X-ray diffraction data on protein crystals can be collected either at room temperature or at cryogenic temperature (100K) with the crystal bathed in a liquid nitrogen stream. Room temperature data collection was initially used for protein crystals but with the advent of synchrotron sources for X-ray diffraction data collection, radiation damage became a major problem in protein crystallography. Collecting data at cryogenic temperature *i.e.,* cryo-crystallography then became essential for protein crystallography. Collecting data at cryogenic temperature greatly reduces the X-ray induced radiation damage. The formation of crystalline ice within and outside the protein crystal has to be prevented to avoid damage to the crystal lattice. This is achieved by flash cooling the crystal in liquid nitrogen in the presence of a cryo-protectant. Various cryo-protectants have been explored in protein crystallography. Glycerol is the most commonly used cryo-protectant.

For the present work, X-ray diffraction data were collected at 100K at European Synchroton Radiation Facility (ESRF), Grenoble, France, using rotation method. Crystals were equilibrated briefly in cryoprotectant solution (25% glycerol and 75% reservoir solution) before flash freezing in liquid nitrogen. A total of 60 to 150 frames were collected for different crystals, each for an oscillation angle of 1º and exposure time of 1-120 seconds. Exposure time for different crystals were decided by looking at the intensity of spots in initial diffraction images.

Data Processing

During data processing step, reflections on diffraction images are assigned their respective Miller indices in a process known as indexing. After indexing, their intensities along with error estimates were obtained. This process is known as integration. The integrated intensities are corrected for Lorentz, polarization and other effects. During the data collection, the intensities of identical and symmetry related reflections are collected on separate frames of data. Integrated intensities of these reflections are then compared and a scaling factor is applied to each frame which corrects the data for experimental drifts and makes the data internally consistent. After scaling, multiple measurements of a particular Bragg reflection are averaged, yielding a reduced data set containing the index (hkl) and its averaged intensity (I_{hkl}).

The data processing can be summarized in following seven steps:

1. Visualization and inspection of diffraction images.

2. Indexing of the diffraction pattern.

3. Refinement of the crystal and detector parameters.

4. Integration of the diffraction intensities.

5. Scaling of the data.

6. Determination of point group symmetry and merging of the symmetry related reflections.

7. Statistical analysis and error estimation.

Several software programs are available for data processing. Some of the more popular programs are Mosflm, XDS and HKL-2000. For present study, XDS suite of programs (Kabsch 2010) was used. XDS provides modules that carry out all the above functions. It has 3 major modules, XDS, XSCALE and XDSCONV. XDS is the major program which performs indexing and integration. From a set of diffraction images, XDS infers crystal symmetry and produces a list of Miller indices 'hkl', with their corresponding corrected integrated intensities. XSCALE is scaling module which puts the data obtained from the XDS on a common scale and merges intensities of symmetry related reflections to produce a reduced dataset. It also corrects the data for radiation damage, absorption effects and detector sensitivity differences. XDSCONV converts the files obtained from processing with XDS or XSCALE into various variety of formats required by different software programs for subsequent structure determination steps.

Data Statistics

Final step in the data processing is the statistical analysis of the processed data. All data processing programs provide several statistics which help in deciding the quality of the data. These statistics are primarily based on the internal consistency of the data [Evans 2006]. One of the more frequently used class of statistical quantity, to judge the quality of

the data is R-factors. The conventional R-factor, R_{merge} (R_{symm}) is given by following formula:

$$R = \frac{\sum_{hkl} \sum_j | I_{hkl,j} - < I_{hkl} > |}{\sum_{hkl} \sum_j I_{hkl,j}}$$

It is suggested that R_{merge} is not a very good parameter for judging the quality of the data as it does not account for the improvement in the merged intensities gained by averaging over multiple observations [Evans 2006]. For this reason R_{merge} generally increases with increasing redundancy. Improved R factors have been suggested [Diedrichs and Karplus 1997, Weiss 2001].

Deciding resolution cut-off for diffraction data is a difficult decision which also depends on future use of the data. For structure determination and refinement a general rule is to use resolution cut-off where <I/σ(I)> (averaging within resolution bins) falls below 2.0 [Evans 2006].

Structure Determination

Since, X-rays are diffracted by electrons from atoms in the molecule, the result of a X-ray experiment is a 3-dimensional map showing distribution of electrons in the structure. This electron density distribution represents molecular structure of the protein. The crucial step is to convert the diffraction data to electron density distribution. Diffracted X-rays cannot be focused and recombined to form an image as in the case of optical microscopy. Therefore, the re-combining has to be done by a mathematical equivalent known as Fourier synthesis. In this mathematical model, each reflection is observed as the summation of all the waves diffracted in that direction. This is described as a structure

factor equation, which is a Fourier series in which each term corresponds to the wave diffracted by each atom (or volume element) in the structure in the direction of that reflection.

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

The overall structure factor $\mathbf{F}_{hkl}$ is related to the electron density by Fourier transform. So, the electron density can be calculated by a Fourier series in which each term is a structure factor.

$$\boldsymbol{\rho(xyz) = \frac{1}{V}\sum_{hkl} \left|\, F_{hkl} \,\right| exp(i\alpha_{hkl}) exp\,[-2\pi i\,(hx + ky + lz)]}$$

Where, $|F_{hkl}|$ and α_{hkl} are structure factor amplitude and phase respectively, of the Bragg reflection hkl. This equation is known as electron density equation and is the basis of protein crystallography. Electron density equation allows us to calculate electron density distribution inside the crystal from a set of structure factor amplitudes and phases. The amplitude depends on the type of element and phase depends upon the position of atom. To calculate an electron density distribution, both of these parameters are required for each reflection. Structure factor amplitudes are directly obtained from intensity of the reflection recorded on the diffraction image, in the following manner:

$$I_{hkl} = |F_{hkl}|^2$$

However, phases of diffracted rays cannot be measured during a X-ray diffraction experiment. This inability to measure phase directly is known as the crystallographic 'phase problem'. Solution of the phase problem is one of the major challenges in the determination of three-dimensional structures of proteins by X-ray crystallography. Over the years, several innovative methods have been developed to solve the phase problem in protein crystallography. Prominent among them are listed here:

1. Multiple isomorphous replacement method (MIR)
2. Single anomalous dispersion method (SAD)
3. Multi-wavelength anomalous dispersion method (MAD)
4. Difference Fourier method
5. Molecular replacement method (MR)
6. Direct methods

Isomorphous replacement (IR) method involves the introduction of one or more heavy atoms (with more number of electrons) in protein crystals in a way that does not affect the integrity of crystal and derivative crystal remains isomorphous to the native crystal. Presence of heavy atoms results into measurable differences in the intensities of some of the reflections in the diffraction pattern. These intensity differences between native and heavy atom derivative crystal are then used to obtain phase information. A single derivative provides two equally possible phase solutions so data from more than one heavy atom derivatives is required to get a single unambiguous phase solution, hence the method is known as Multiple Isomorphous Replacement (MIR).

Anomalous dispersion methods exploit the anomalous scattering that takes place when wavelength of the incident X-rays is near the absorption edge of atoms scattering the X-

rays. If crystal contains any such atom known as anomalous scaterrer, anomalous scattering takes place if diffraction data is collected at or near the wavelength of absorption edge. Anomalous scattering results into breakdown of Friedel's law *i.e.,* the intensities of Friedel pairs are no longer equal. This difference in intensities of Friedel mates can be used in phase determination. Single anomalous dispersion (SAD) requires data to be collected only at a single wavelength around the X-ray absorption edge of the anomalous scatterer; however, phase solution by SAD leaves ambiguity in two equally likely results. This phase ambiguity is removed by density modification. Multi-wavelength anomalous dispersion (MAD) method requires diffraction data to be collected at at-least two or preferably three different wavelengths from the same crystal. As this method requires a tunable X-ray source, it can only be performed at a synchrotron. The phases obtained from the MAD method are of good quality and often produce very high quality electron density maps.

If the structure to be solved is almost identical to a previously solved structure, then difference-Fourier method can be used. In this method, phases from the model are directly used as the phases of new protein. These phases are combined with the structure factor amplitudes of experimental data to obtain weighted electron density maps. These maps show the structural features, which are different in the two structures. For example, structure of mutant protein can be determined using phases from native protein structure.

If the known structure is somewhat different from the protein in question, then molecular replacement (MR) method can be used. In MR, initial phase estimates are produced by placing the model of a known homologous molecule into the crystal of the unknown

structure. All possible orientations and positions of the model are tried and the one, which gives best match with the observed diffraction data, is chosen. Phases calculated from this solution are then combined with observed structure factor amplitudes to produce an electron density map using the electron density equation.

In direct methods, statistical relationships among structure factors are used to derive phases from the observed structure factor amplitudes. Direct methods are based on the atomicity and positivity of the electron density. The reliability of the estimated phases decreases rapidly with increase in number of atoms in the structure. Therefore, direct methods are not suitable for large protein structures. These methods are mostly used for finding heavy atom substructure during phase solution using experimental methods.

The structures reported in the present thesis have been solved by the difference Fourier method. Structure of wild type tethered HIV-1 protease (PDB ID 1LV1) was used as the template.

Electron Density Maps and Model Building

Electron density map thus generated is then interpreted as molecular model of the protein. Different types of electron density maps can be used for this purpose. Most commonly used maps are mFo-DFc; 2mFo-DFc map and omit maps. Here, Fo is the observed structure factor amplitude and Fc is the calculated structure factor amplitude. *m* is a figure of merit which is an estimate of the cosine of the error in the phase, and *D* is a scale factor. The mFo-DFc map contains both positive and negative peaks. Positive peaks signify that the electron density is not yet accounted by the model used to calculate the Fc and hence is very useful to complete the model. Negative peaks in this map show atoms

placed wrongly in the model. Thus, this map also helps in removing the errors in the model. The 2mFo- DFc map is the sum of a Fo map plus a mFo-DFc map, and contains information from both the Fo map and the difference mFo-DFc map and is used in the model building. Another useful type of electron density map is the omit map [Bhat 1988]. Omit maps are calculated by leaving out the portion of the model to be examined from the phase calculation. The remaining part of the model provide the phase information for the omitted portion of the model. This feature of the Fourier is possible because all parts of the structure contribute to each reflection. Omit maps are a very useful type of map which allows to remove model bias in the phase calcualtion.

Model building is the process of interpreting electron density map in terms of a set of atomic coordinates representing molecular model. Electron density map interpretation is done using computer graphics programs like O [Jones et al 1991] and Coot [Emsley and Cowton 2004, Emsley et al 2010]. For the present work both of these programs were used. These programs have several tools that help in model building and manipulation. The process of model building starts with the construction of a Cα backbone into the electron density maps. Main chain and side chains of the protein residues can then be built using sequence information. This process yields a model that fits the electron density but possibly still incomplete and includes a number of errors. This initial molecular model has to be refined against observed intensities and stereo-chemical information. In difference-Fourier method, model building is relatively straightforward as we already have a sufficiently correct model. But this also introduces bias into the model towards the template structure. Removal of this bias requires manual model manipulation and refinement. After each round of refinement, the model and an electron density map

were visualized using O and Coot. Model was then manipulated to remove local errors and this improved model is then again refined.

Crystallographic Refinement

Initial model built into the electron density map is rarely complete and accurate. Building a more accurate model requires manual model building and refinement. The purpose of the refinement is to optimize simultaneously the agreement of an atomic model with both the observed diffraction data and known stereo chemical information [Brunger et al 1998]. This is usually achieved by adjusting the model parameters in a way that minimizes the difference between structure factor amplitudes obtained from the observed intensities and structure factor amplitudes calculated from the atomic model. Historically, protein structure refinement is done by minimization of a least squares residual function given below:

$$\boldsymbol{LS} = \sum_{\boldsymbol{h}} \frac{\mathbf{1}}{\boldsymbol{\sigma^2}} \left(\left| \boldsymbol{Fo}\,(\boldsymbol{h}) \right| - \boldsymbol{k} \left| \boldsymbol{Fc}\,(\boldsymbol{h}) \right| \right)^{\mathbf{2}}$$

Where LS is the least squares crystallographic residual, h=hkl, $\frac{\mathbf{1}}{\boldsymbol{\sigma^2}}$ is weight applied to each observation, Fo and Fc are the observed and calculated structure factor amplitudes respectively and k is a scale factor. A successful least squares refinement requires observation to parameters ratio to be high enough to allow a degree of over-determination of the system. Unfortunately, this value is very low for proteins. This makes a full least squares refinement unstable. Addition of stereo-chemical knowledge in the form of restraints increases the observation to parameters ratio and makes the refinement stable. Other form of addition of stereo-chemical information is to add them as fixed value

known as constraints. Constraints reduce the number of parameters. Least squares refinement is highly suitable for small molecule crystallography, where high resolution X-ray diffraction data is available and number of parameters to be determined is small. However, least squares refinement is less successful in protein crystallography. In protein crystallography, number of parameters is very high and only moderate resolution data is generally available which makes data to parameter ratio poor. A more successful generalized refinement method known as maximum likelihood is now implemented in several protein crystallographic refinement packages, including CNS [Brunger et al 1998] and PHENIX [Adams et al 2010]. The goal of maximum likelihood method is to find the model, which is most consistent with the observations. Likelihood is described as the probability of a set of observations given the model. In the maximum likelihood method, Likelihood (L) *i.e.,* the joint probability, P(Fo;Fc), of all the observed structure factors given the model is calculated. In practice, negative log likelihood is calculated and minimized to obtain the best model. The maximum likelihood provides better results than least squares refinement because it can account for incompleteness and errors in the model. Maximum likelihood provides results that are less biased toward the model.

R-factors

R factors are statistical quantities used to monitor the progress of refinement and ensure that the refinement is moving towards convergence. When cycles of refinement provide no further improvement in R-values, then the refinement is considered as converged. These statistics provide a handy tool to assess quality of fit of the model to experimental data. Crystallographic R-factor (R_{cryst}) is the traditionally used statistic to judge the quality of the model. The lower R_{cryst} values describe better agreement between the

observed and calculated structure factor amplitudes. However, it was shown that it was possible to reduce the R_{cryst} artificially, by increasing number of parameters used to describe the model. Thus, diffraction data can be overfit without a real improvement in the atomic model. To alleviate this problem, free R-factor (R_{free}) was introduced which is a more reliable indicator of model quality [Brunger 1992]. 5-10% of diffraction data is set aside as a test set (T) and is never used in the refinement. Refinement is performed on remaining 95% of the data, which makes the working set. The R-value is then computed for both the test and working set. R-values calculated using test data is known as R_{free}. It is used to prevent over-refinement of the model and is a very useful guide during model building and refinement. In a refinement step, both R_{cryst} and R_{free} should decrease simultaneously. For present work, 5% reflections were set aside in each data set to provide a test set for cross validation calculation.

Addition of solvent

A protein crystal contains approximately 20-70% solvent content. Most of these solvent molecules are disordered hence are not visible in electron density map. However, the water molecules bound to protein molecule are ordered and can be located accurately in electron density map. In the present thesis, water molecules are added by using programs CNS, PHENIX and COOT. These water molecules are than checked manually. Simultaneous presence of electron density at 2.5σ level in mFo- DFc map and at 1σ level in 2mFo-DFc map are used to validate these water molecules.

Structure Validation and Deposition

Finally, refined model need to be checked for stereo-chemical and geometrical errors. This is done in a process known as validation. Several structure validation programs are

available, which check stereochemical and geometrical parameters of the model by comparing them with the ideal values. Covalent geometry, planarity, dihedral angles and chirality are checked. An useful criteria to judge the quality of the model is Ramchandran's map. It is a map of backbone torsion angles. Most of the residues in a good model should lie within allowed regions of the Ramachandran map. Residues which lie outside the allowed regions need thorough examination. The last step in the process of crystal structure determination is deposition of the structure in the Protein Data Bank (PDB), which is the depository of three dimensional structures of macromolecules.

Superposition of structures

Comparison of macromolecular structures to find similarity or differences between them is done by least squares superposition of one structure on the other. In the present study, structures were superposed using programs O [Jones et al 1991], Coot [Emsley and Cowton 2004, Emsley et al 2010] and superposition module of CCP4.

Circular Dichroism Spectroscopy

Far UV Circular Dichroism (CD) spectra on wild type and mutant proteins were recorded using a Jasco J-815 CD spectropolarimeter. Protein solutions were used at a concentration of 10 mM in 50 mM sodium acetate buffer, pH 4.5. Thermal stability of the protein can be assessed using CD spectroscopy by observing changes in the CD spectrum with increasing temperature. For the present thesis, thermal denaturation experiments were carried out by monitoring CD signal at 215 nm over a temperature range of 20-80ºC with a scan rate of 2 ºC/min. Band width was 1 nm with 2 second response time. Thermal denaturation midpoint (T_m) was calculated as the temperature, where the first derivative of the ellipticity versus temperature graph reached a minimum.

Chapter - 3

X-Ray Structure of HIV-1 Protease *In-situ* Product Complex

The role of HIV-1 protease is to cleave viral precursor polyproteins at different cleavage sites on Gag and Gag-Pol polyprotein to produce mature structural and functional proteins. Nine of these cleavage sites are well established, and are given in table 3.1. One can observe certain preference for type of amino acids at different subsites within the diverse sequences of cleavage sites. Based on this preference, cleavage sites of HIV-1 protease can be grouped into two classes: first group of peptides contains hydrophobic/hydrophobic residues at P1 and P1', while the second group contains aromatic residue/proline at P1 and P1' positions. HIV-1 protease is unique in its ability to cleave a substrate with Tyr-Pro or Phe-Pro cleavage site flanked by several amino acids residue. Though, proline appears frequently at P1 subsites of other proteases but has not been found at subsite P1' of natural cleavage sites.

The molecular basis of cleavage by HIV-1 protease is still not fully understood. A variety of experimental and theoretical studies have been used to explore the detailed cleavage mechanism [Hyland et al 1991, Jaskolski et al 1991, Rodriguez et al 1993, Lee et al 1996, Liu et al 1996, Rose et al 1996, Silva et al 1996, Park et al 2000, Northrop 2001, Dunn 2002, Prabu-Jeyabalan et al 2002, Brik and Wong 2003, Prabu-Jeyabalan et al 2003, Kovalevsky et al 2007]. Based on these studies, a general acid-general base catalysis mechanism similar to that of non-viral aspartyl proteases has been proposed for HIV-1 protease. A schematic representation of the basic mechanism is outlined in figure 3.1. According to the proposed mechanism, a water molecule activated by charged aspartate side chain acts as nucleophile and attacks the scissile peptide bond leading to the formation of tetrahedral intermediate. The next step is the protonation of scissile N, which results in the breakage of peptide bond to generate peptide products. It is widely

accepted that the water molecule coordinated to both catalytic aspartates in the free enzyme is the nucleophilic water. There is, however no consensus about the functional group which carries out this scissile N protonation. While most of the proposals invoke protonation by outer oxygen of a catalytic aspartate, the proposals by Stroud and his colleagues [Rose et al 1996], and more recently, by Kovalevsky et al [2007] suggest protonation by the gem-hydroxyl of the tetrahedral intermediate. Interestingly, in all these proposals, the cleavage mechanism is presumed to be the same irrespective of whether the amino acid at the cleavage site is a proline residue or a non-proline residue. Since oligopeptides have been shown to serve as effective substrates for HIV-1 protease, we have undertaken to solve crystal structures of HIV-1 protease oligopeptide substrate complexes to explore the mechanism. Although hundreds of high-resolution structures of active HIV-1 protease complexed with inhibitors are reported, structural data on complexes with true substrates has been difficult to obtain. The first attempts were made by Stroud and co-workers to co-crystallize HIV-1 protease with oligopeptide substrates [Rose et al 1996]. These authors, however, found that the substrate was cleaved during crystallization and that only one of the products was present in the crystal. In more recent crystallographic work, the full oligopeptide substrates could be found in the active site when the enzyme used for co-crystallization was an inactive D25N mutant [Prabu-Jeyabalan et al 2002, Prabu-Jeyabalan et al 2003]. These structures could, however, serve as approximate models for Michaelis complex. Our lab has approached the problem differently [Cheng et al 1990, Pillai et al 2001, Pillai et al 2004], and have been able to prepare crystalline complexes between active enzyme and unmodified oligopeptide substrates [Kumar et al 2005, Das et al 2006]. This report is only the third example of

HIV-1 protease substrate complex and is the first one involving a P1' proline. An interesting feature of the present complex is that there is no hydrogen bond between the scissile N atom of P1' residue and an outer oxygen atom of catalytic aspartates in sharp contrast to formation of such a hydrogen bond when the substrate peptides did not contain a proline residue in P1' position [Kumar et al 2005, Das et al 2006]. Because a hydrogen bond to the scissile nitrogen atom could act as a conduit for protonation of the scissile nitrogen, its absence in the present structure opens up the possibility that HIV-1 protease uses different mechanisms to cleave substrates with and without proline residue in P1' position. The second interesting result is the absence of a low-barrier hydrogen bond (LBHB) [Cleland et al 1998] between inner oxygens of catalytic aspartates that was observed in the earlier study [Das et al 2006].

Table 3.1. Cleavage sites of the HIV-1 protease in the Gag and Gag-Pol precursor proteins. * shows the actual site of the cleavage. Proline containing cleavage sites are shown in bold.

CLEAVAGE SITES	**JUNCTION IN POLYPROTEINS**
Cleavage sites in Gag	
SQN**Y*P**IVQ	MA-CA
ARVL*AEAM	CA-p2
ATIM*MQRG	p2-NC
QANF*LGKI	NC-p1
PGNF*LQSR	p1-p6
Cleavage sites in Pol	
SFN**F*P**QIT	TF-PR
TLN**F*P**ISP	PR-RT
AETF*YVDG	RT-RH
RKIL*FLDG	RH-IN

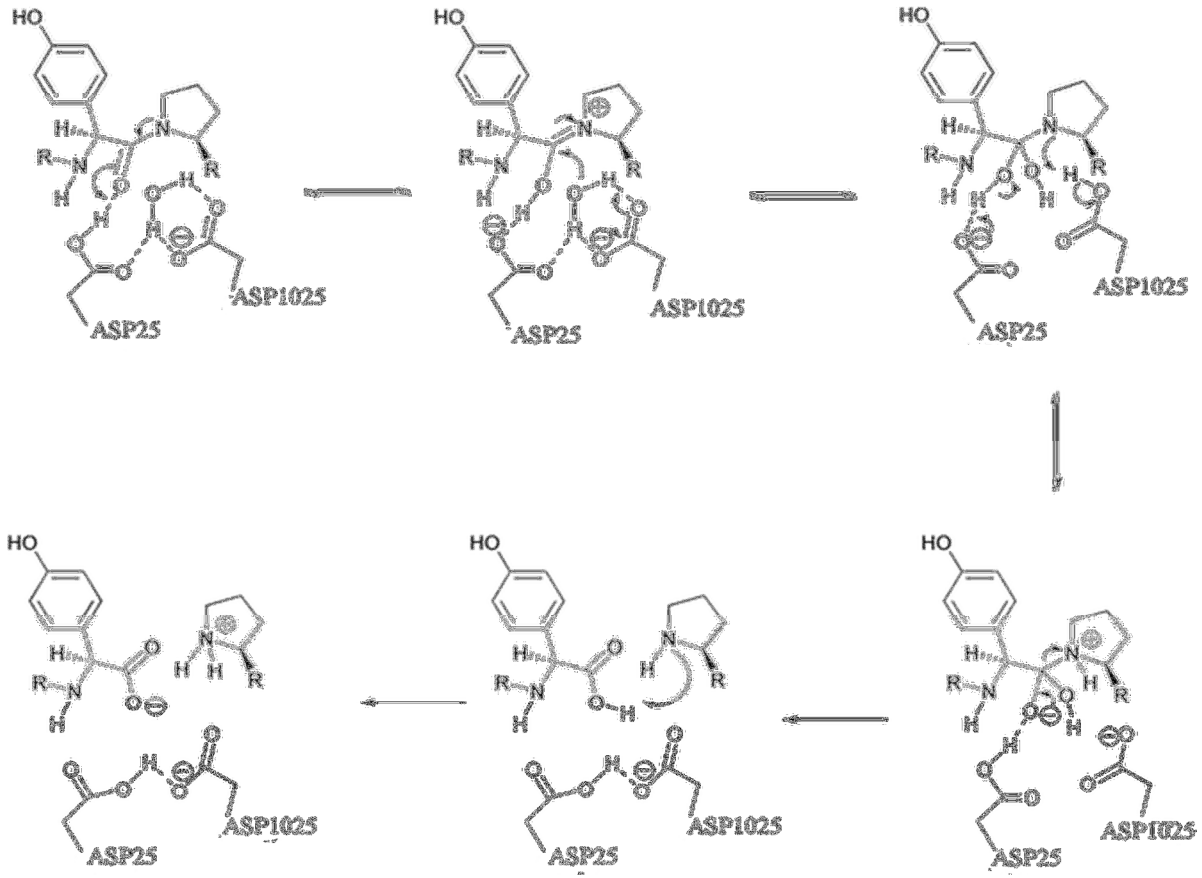

Figure 3.1: Proposed cleavage mechanism for HIV-1 protease [Brik and Wong, Org Biomol Chem 2003 - adapted by permission of The Royal Society of Chemistry]

Experimental Procedures

Expression, purification, and crystallization of HIV-1 protease

HIV-1 protease tethered dimer described in material and methods chapter was used for the present study. Protein was over-expressed and purified using the protocol described in chapter 2. Final concentration of the protein used for crystallization was 1-4 mg/ml in 50 mM sodium acetate buffer containing 1 mM DTT. Crystallization was done by hanging-drop vapor diffusion method at room temperature. 1-6% saturated ammonium sulfate in 200/100 mM sodium citrate/sodium dihydrogen phosphate buffer (pH 6.2) was used as reservoir. A hepta-peptide of sequence SQNY*PIV was used in the present study. It is derived from MA/CA junction in Gag polyprotein (Table 3.1). Crystals of the enzyme-hepta-peptide complex were prepared by the soaking method as detailed in material and method chapter. The pH of the soaking solution was 6.2.

X-ray diffraction data collection

The HIV-1 protease crystal soaked with hepta-peptide was equilibrated for 5 minutes in the cryoprotectant (25% glycerol and 75% reservoir solution) before flash freezing under liquid nitrogen stream for exposure to X-rays on the X06SA beamline at PSI-SLS. 150 images were collected, each of 1.0^{o} oscillation and 2 minutes of exposure. The crystals diffracted to 2.0 Å resolution. The diffraction data were indexed, integrated, and scaled using the software package XDS [Kabsch 2010]. Diffraction data statistics are shown in Table 3.2.

Table 3.2: Reflection data and refinement statistics for HIV-1 protease heptapeptide substrate complex.

Wavelength of radiation used (Å)	0.97644
Resolution range used in refinement (Å)	45 - 2.00
Space group	$P6_1$
Unit cell dimensions (Å)	a=b=62.38, c=82.00
Asymmetric unit	One tethered dimer
Total number of reflections measured	91,324
Total number of unique reflections	12,070
R_{merge} (%)	8.0 (49.4)
Completeness (%)	99.2 (97.0)
$\langle I/\sigma(I) \rangle$	18.03 (3.99)
R_{cryst} (%)	20.54
R_{free} (%)	23.75
Number of water molecules	183
RMSD from ideal values	
Bond lengths (Å)	0.006
Bond angles (°)	1.265

*The numbers between parentheses indicate the values in the outermost resolution shell.

Rmsd = root mean square deviation.

Structure determination and refinement

In view of good isomorphism, the protein coordinates extracted from the unliganded structure (PDB ID 1LV1) were used directly in refinement. The structure was refined in CNS using standard simulated annealing protocols and maximum likelihood target function for structure factor amplitudes [Brunger et al 1998a, b]. Overlapping ligand molecules related by pseudo-two fold symmetry of the active site were simultaneously included in refinement by switching-off van der Waal interactions between them, through appropriate IGROUP statements. The molecular motif refined was a 1:1 complex between the tethered protease dimer and the two product oligopeptides still bound in the active site cavity. Water molecules were added manually by examining mFo-DFc and 2mFo-DFc maps. In the calculation of simulated annealed omit maps, the annealing temperature was 600K, and all atomic parameters were varied. The entire model building and structural superpositions were carried out using the software O [Jones et al 1991]. For structural comparisons with other substrate/inhibitor complexes, only the protein Cα atoms were used in the superposition. All figures have been prepared using the software PyMOL [Schrodinger, LLC].

Results

The amino acid residues in the two sub-units of the tethered enzyme have been labeled 1 to 99 and 1001 to 1099 respectively. Residues in the linker region are numbered 101 to 105. The electron density for the protein part is of excellent quality, except for the linker region. There was no density for the linker (residues 101 to 105), even at low contour levels, suggesting disorder of this region. Positive maxima of the mFo-DFc map (Figure 3.2) in the active site indicated the presence of the substrate. In view of the pseudo two-

fold symmetry of the active site, the substrate was placed in the two symmetry related orientations with occupancy of 0.5 each. In each orientation, the substrate was modeled in three different ways: 1) as an oligopeptide cleaved at the scissile bond, but with the two products still bound in the active site, 2) as a regular oligopeptide uncleaved at the scissile bond, and 3) as uncleaved hydrated oligopeptide (tetrahedral reaction intermediate) with scissile carbonyl converted into a gem-diol. Thus, these three models differ from one another only at the scissile peptide linkage. However, during refinement the two-orientation model led to higher R-factors (higher by approximately 0.5% in R_{free} for all models considered for the substrate) and poorer fitting into electron density compared to a single orientation model. The crystallographic R-factors obtained for single orientation interpretations are given in Table 3.3. It may be seen that lowest R-factors are obtained for the cleaved substrate model. Further, the occupancy of the cleaved substrate was not unity, but converged to a value of 0.73. In the remaining 27% unit cells the difference density is interpreted as eight water molecules, which are shown as red spheres in figure 3.2. Therefore, this set of eight water molecules and the two product peptides are mutually exclusive. Of the eight water molecules, two each are located at positions corresponding to P2 and P2' sites, while the remaining four are along the P1-P1' linkage. Real space correlation coefficient calculated using RS-fit option in the modeling program O, was 0.569 when averaged over all residues in the substrate. Simulated annealed omit map calculated by omitting the scissile N atom is shown in figure 3.2 along with the mFo-DFc map covering the P1-P1' region of the substrate.

Table 3.3: Crystallographic R-factors for various models of the difference density.

Atomic model for difference density in the active site	**R_{cryst} %**	**R_{free} %**
Cleaved substrate (73% occupancy) along with water molecules (27% occupancy)	19.36	23.41
Uncleaved substrate (73% occupancy) along with active site water molecules (27% occupancy)	19.39	23.92
Tetrahedral Intermediate (73% occupancy) along with active site water molecules (27% occupancy)	19.60	24.15
Only water molecules (100% occupancy)	19.91	24.10
Cleaved substrate (100% occupancy)	19.62	23.81
Uncleaved substrate (100% occupancy)	19.63	24.00

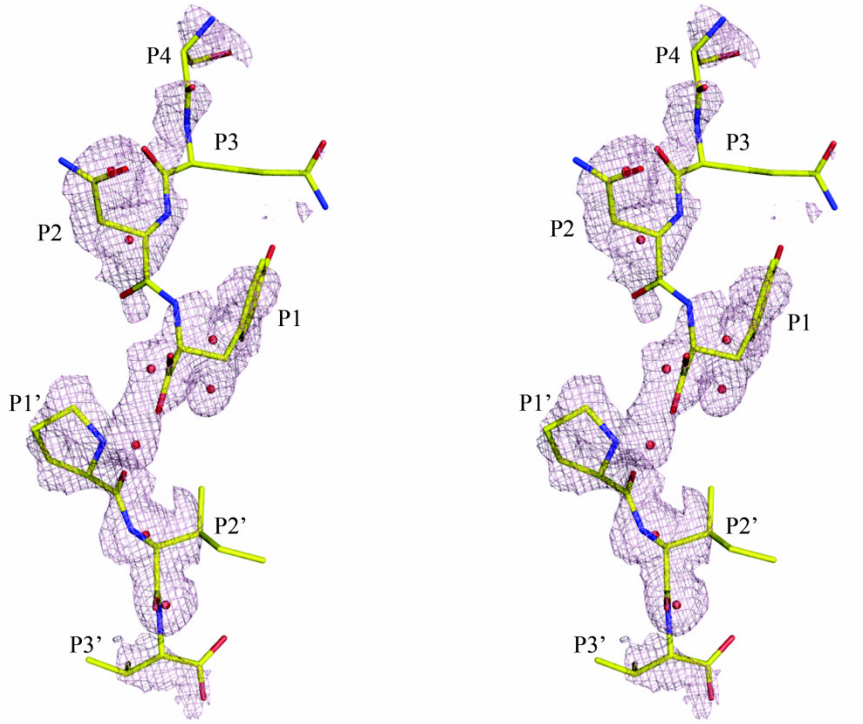

Figure 3.2: Stereodiagram of electron density maps in the active site region: a) σ_A weighted 2Fo-Fc map for all substrate residues and partial water molecules, which are shown as red spheres.

The partial water molecule WAT1 in the active site coordinates both catalytic aspartates and figure 3.3 shows density for this water molecule in a simulated annealed omit map. This water molecule is suggested to perform the nucleophilic attack during peptide bond cleavage. The position of water molecule WAT1 is only 1.5 Å away from the position of hydroxyl oxygen of hydroxyethylene inhibitor JG365, in the very high-resolution structure of the complex [Johnson et al 2007] (Figure 3.4), providing further support that WAT1 could be the catalytic water molecule. The P product carboxyl oxygen O1 is nearest to WAT1, at 1.2 Å, and may therefore have been derived from nucleophilic attack, leaving the O2 atom of the product carboxyl to be the original peptide oxygen atom.

Hydrogen bonding at the catalytic center

Electron density for the ligand/product molecules is defined better for residues ranging from P2 to P2' (Figure 3.2). There is no density for residues on either side of these limits, implying that these residues are very mobile and that there are no strong interactions between the substrate and the enzyme at these positions. Even in the P2-P2' region, the electron density splits into three distinct islands when the contour level of the 2mFo – DFc map is increased to higher and higher values. These islands are at residues P2 and P2' and at the P1-P1' linkage. Interestingly, these are the only regions of the substrate molecule, which interact directly with the protein through hydrogen bonds as shown in figure 3.5. The main chain carbonyl oxygen of P2' isoleucine is at 3.1 Å from the amide nitrogens of protein residue 29. Interestingly, on the N-terminal side of the scissile bond, the asparagine side-chain of P2 rather than main chain carbonyl forms hydrogen bonds with the backbone nitrogen atoms of residues 1029 and 1030. It is interesting that these

hydrogen bonds have been described as conserved hydrogen bonds that play significant roles in substrate recognition before peptide bond cleavage [Prabu-Jeyabalan 2002, Prabu-Jeyabalan 2003]. The oxygens of the P1 carboxyl group interact with the catalytic aspartate residues as shown in figure 3.6.

The carboxyl oxygen (O1) is hydrogen bonded to D1025 OD2 and with the scissile N on the Q product (O1…N(PRO) = 2.6 Å). The two inner oxygen atoms (OD1s) of the catalytic aspartates form hydrogen bond between themselves. However, this hydrogen bond of length 2.8 Å is weaker than that observed in protease-product complex reported earlier [Das et al 2006]. The two OD1 atoms also accept hydrogen bonds from peptide NH groups of residues G27 and G1027, as is found in all reported structures of HIV-1 protease. Hydrogen bonds from partial water molecule WAT1 to catalytic aspartates are also shown in figure 3.6. The product peptides are linked to the protein also through water mediated interactions (Figure 3.5). The carbonyl oxygens of P2 and P1' residues interact with the flaps through a water molecule. Although hydrogen atoms cannot be located at this resolution, the hydrogen bonds at the catalytic center provide clues to the ionization states of the aspartates. There are a total of three hydrogens (one from the catalytic aspartates in the free enzyme and two from the lytic water molecule), whose positions need be determined in the complex. Since the peptide bond is broken, there has to be one hydrogen atom bonded to the N. The other two hydrogens have to form part of the shortest hydrogen bonds between the P product carboxyl group and the carboxyl oxygens of aspartates. The hydrogen atoms could be attached either to the aspartate oxygens or to the carboxyl oxygens, which, it is not possible to distinguish from the present structure of 2 Å resolution. Analysis of X-ray structures of HIV-1 protease

inhibitor complexes has led to a correlation between protonation state of the aspartates and their hydrogen-bonding pattern [Piana and Carloni 2000]. Whenever the aspartates are monoprotonated, they form strong hydrogen bonds to donor groups from the substrate/inhibitor or to water molecules. In the diprotonated state, there are no strong hydrogen bonds from aspartates. Therefore, according to this hypothesis, the catalytic aspartates in the present structure should be monoprotonated.

Discussion

The cleavage reaction

Structures of enzyme substrate complexes can help map, at the atomic level, the sequence of events that actually happen during the process of bond cleavage followed by product separation. Figure 3.7 shows structural comparison of present structure with two other structures where the substrate oligopeptides have been complexed with an inactive D25N mutant enzyme [Prabu-Jeyabalan 2002, Prabu-Jeyabalan 2003].

The substrate peptides in the inactive complexes are uncleaved, and the structures may be assumed to represent the first step in the cleavage process, that is substrate binding. It is interesting that the P1 carboxyl group is drawn closer to the catalytic aspartates in the active enzyme complex compared to the inactive complex. When the two inactive complexes are compared, the Cα atom of P1' proline is shifted laterally, and is very close to the Cα atom of P1' proline in the present structure. Structural comparison of two product complexes with active enzyme is shown in figure 3.8.

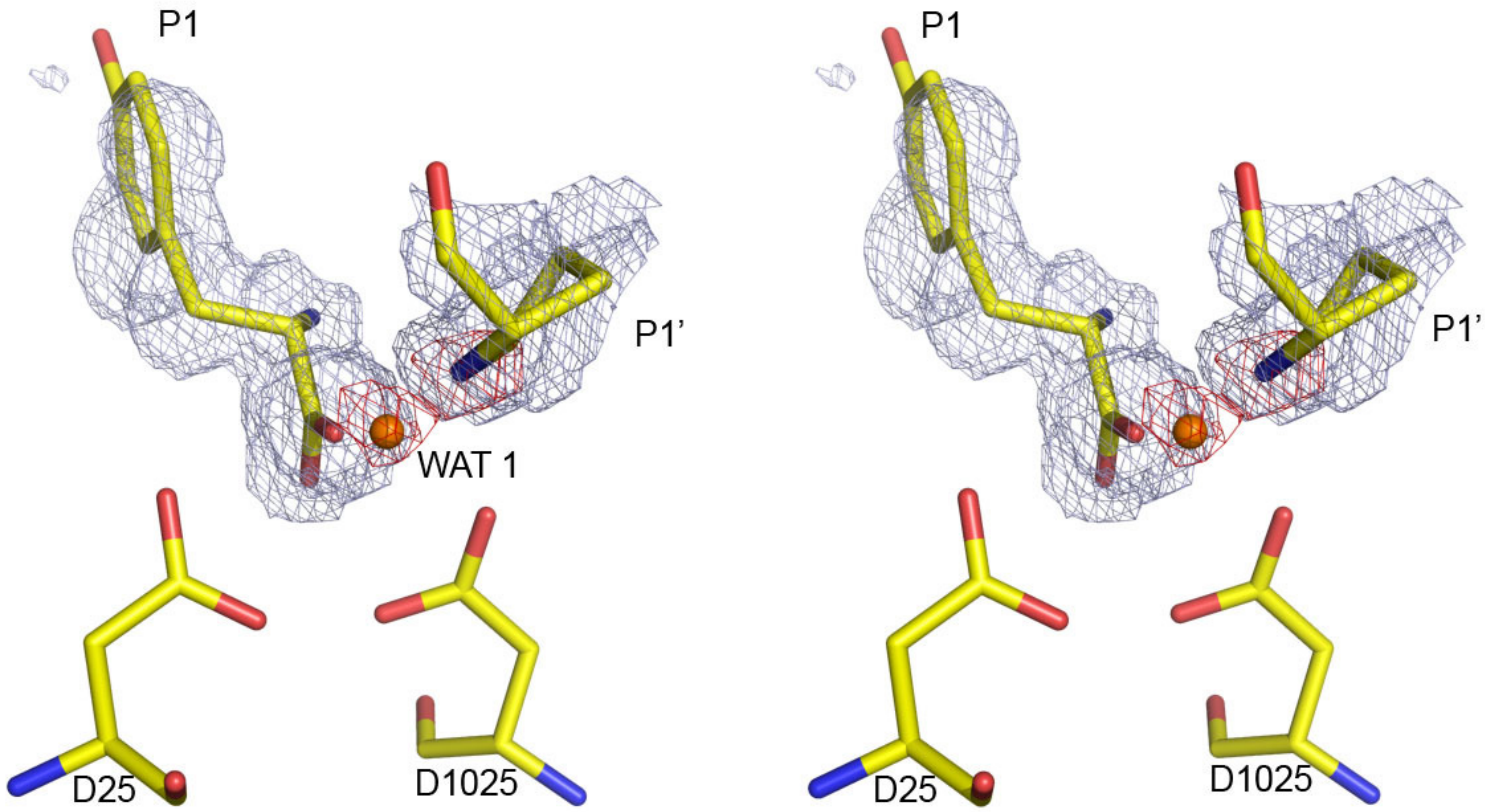

Figure 3.3: Stereodiagram of electron density maps: (1) σ_A weighted Fo-Fc map in the P1-P1' region of substrate (blue, contoured at 2σ), (2) scissile nitrogen in simulated annealed omit map (red, contoured at 3σ) and (3) simulated annealed omit map for partial water molecule, WAT1 (orange, contoured at 3σ). Refined positions of P1, P1' and WAT1 are shown along with catalytic aspartates.

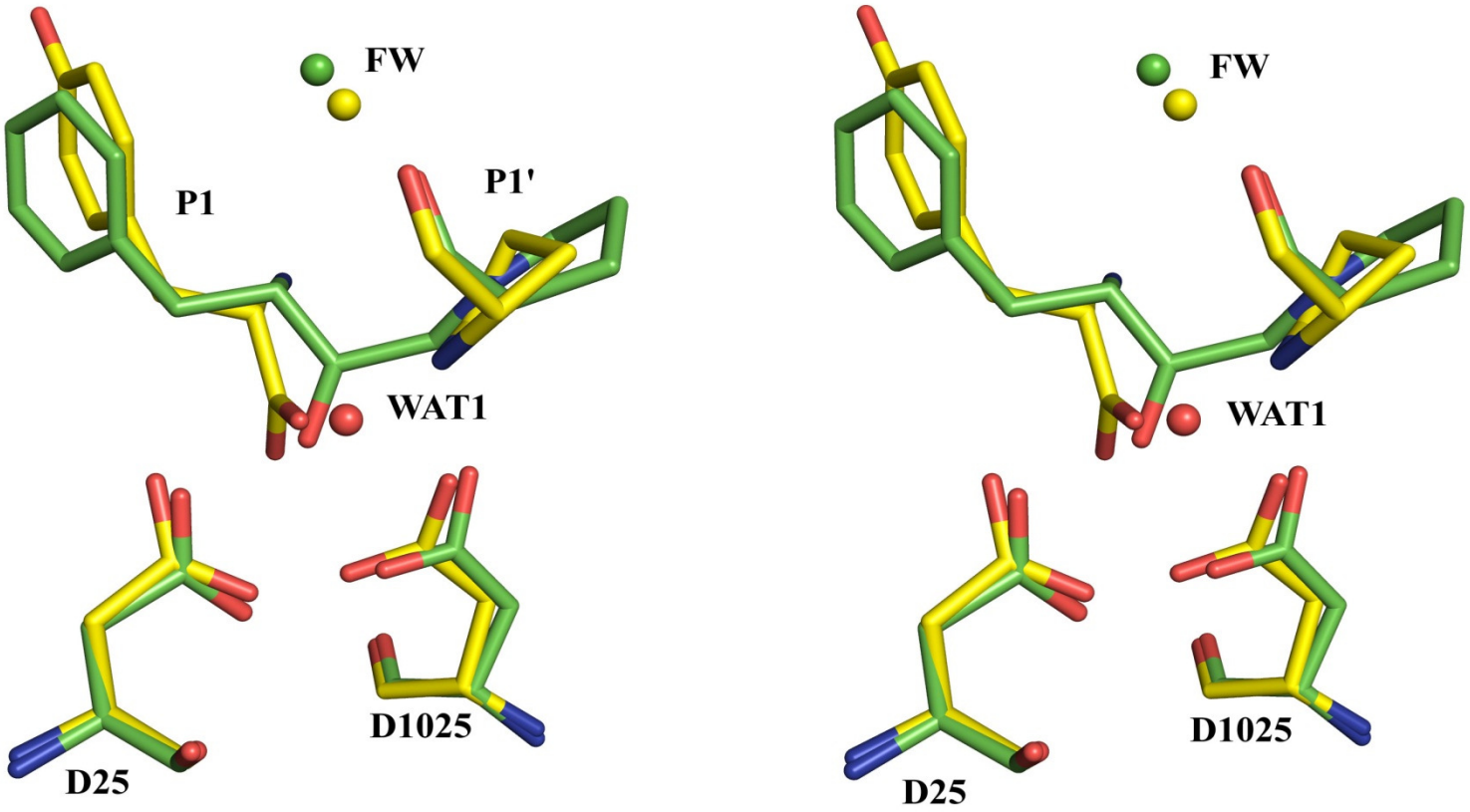

Figure 3.4: Stereodiagram superimposing present structure (yellow carbon) with the structure of hydroxy-inhibitor (JG365) complex (green) [PDB ID 2J9J]. The putative catalytic water WAT1 (orange sphere) is at 1.5 Å from the hydroxyl oxygen of the inhibitor. The flap water molecules (FW) for the two structures are also drawn.

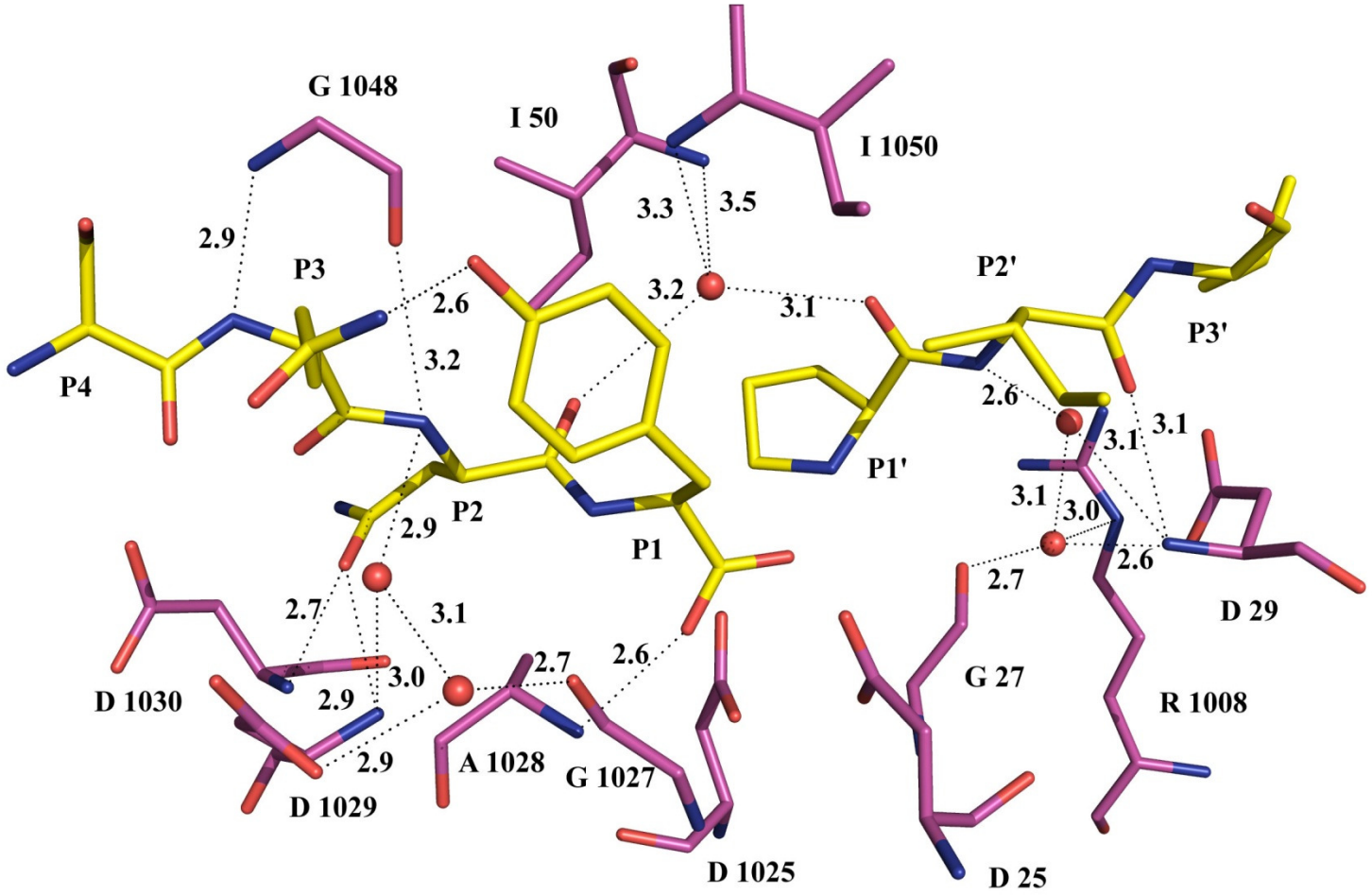

Figure 3.5: Hydrogen bonding interactions of product peptides in the active site cavity. Both direct and water mediated interactions are shown.

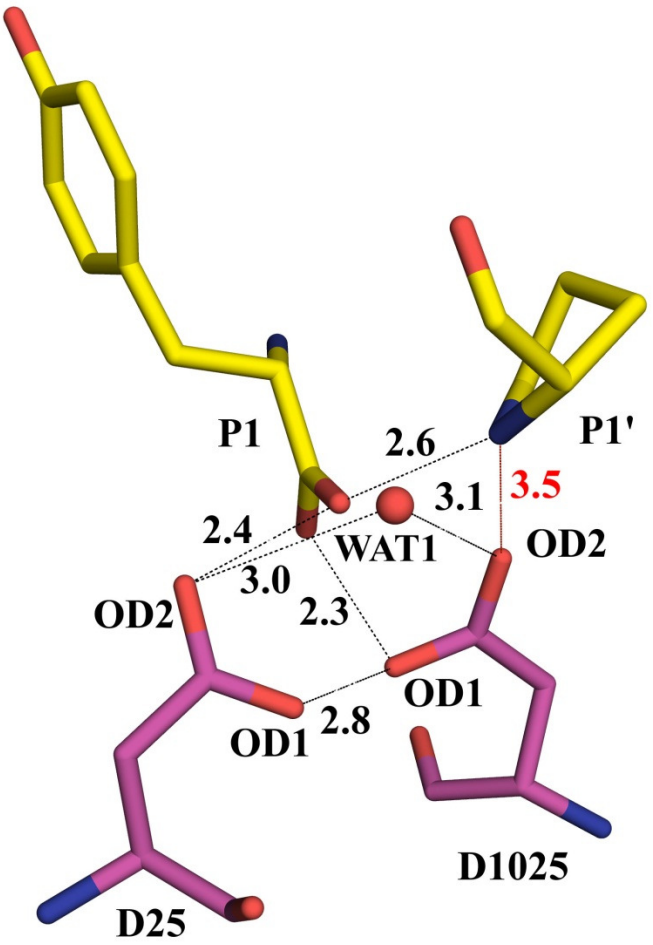

Figure 3.6: Hydrogen bonding interactions involving P1, P1', WAT1 and catalytic aspartate residues. Interatomic distances are given in Å units. Partial occupancy WAT1 and product peptides donot exist simultaneously.

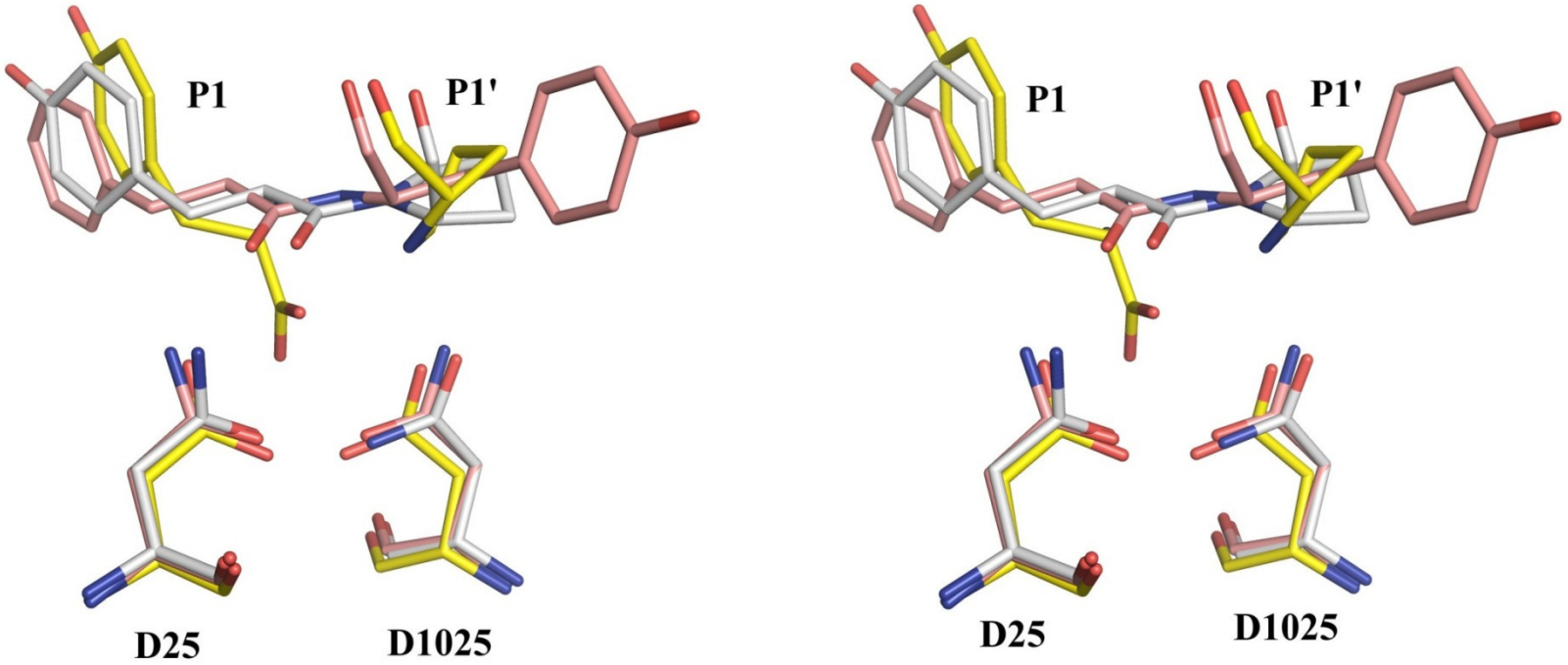

Figure 3.7: Stereo diagram of structural comparison of three complexes: 1) present structure (yellow carbon) 2) inactive D25N mutant enzyme with oligopeptide GAETF*YVDGA [Prabu-Jeyabalan 2002] (grey carbon) and 3) inactive D25N mutant enzyme with oligopeptide VSQNY*PIVQN [Prabu-Jeyabalan 2002] (salmon carbon).

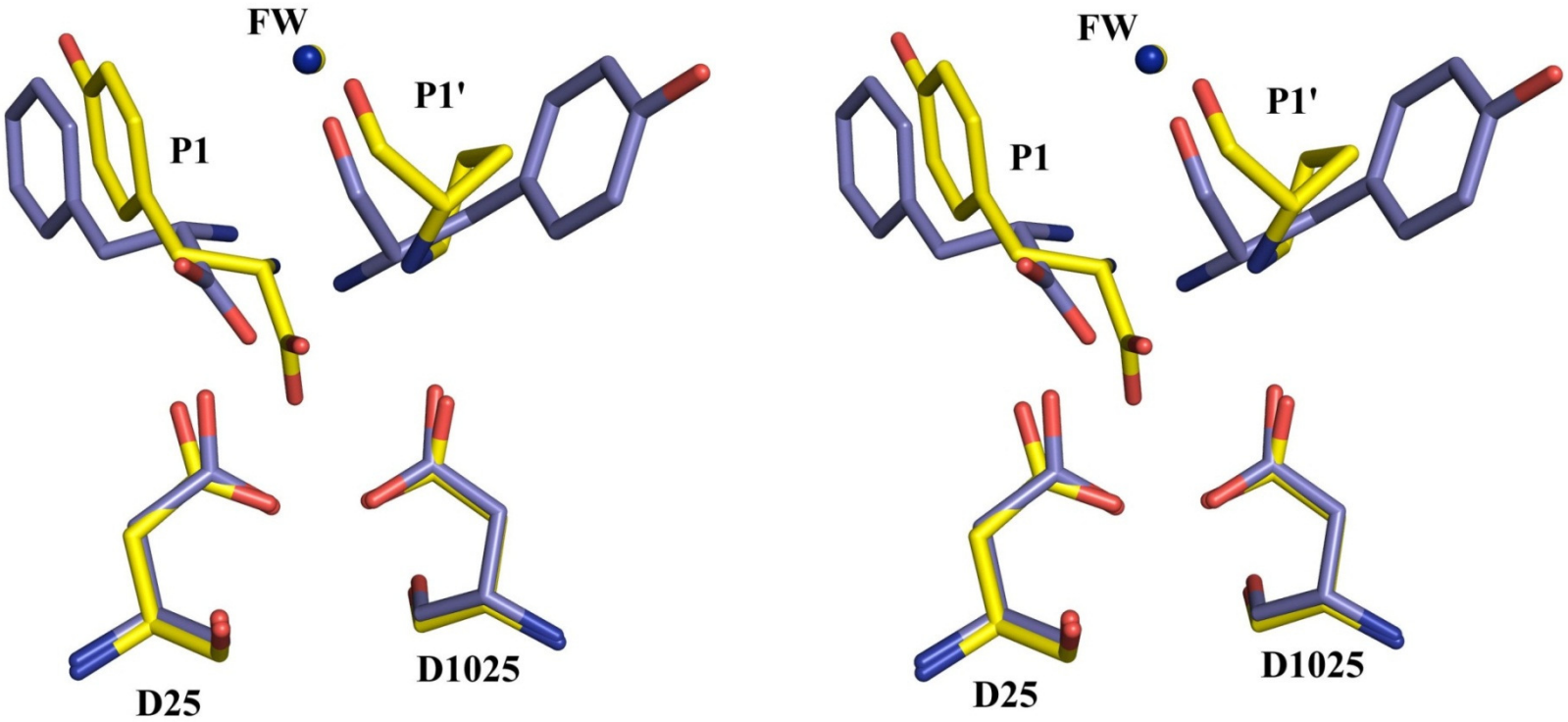

Figure 3.8: Stereo diagram of superposition of cleaved substrate peptides in the present structure (yellow carbon) and in the earlier study (slate carbon) [Das et al 2006]. The products in the present complex are shifted across the aspartates by about 1.7 Å.

It may be seen that both the products (P and Q) have shifted laterally in the same direction across the catalytic aspartates in the present complex. The shift has been in the direction to increase the distance of N atom of P1' proline from the catalytic aspartates. For example, the distance D25 Cγ – N (PRO) has increased to 4.5 Å from 3.69 Å. At the same time, the distance D25 Cγ – C (TYR) has decreased to 3.4 Å from 4.9 Å showing that both product peptides have moved together in the same direction within the active site. The lateral shift of the scissile nitrogen atom is about 1.7 Å. Interestingly a similar lateral shift, but to a lesser extent, of the proline containing structure is seen when complexes of D25N mutant protease with substrates containing a proline and a non-proline residue in P1' position are compared (Figure 3.7). These observations suggest that positioning of the substrate in the active site may be different whenever there is a proline ring in the P1' position. Because of this lateral shift, the distance between scissile N and the closest aspartate outer oxygen is 3.5 Å (Figure 3.6), which is longer than hydrogen bond distance. The absence of this hydrogen bond combined with presence of a hydrogen bond between separating products P and Q, which was mentioned earlier, may be suggesting that the proton transfer to scissile nitrogen atom required for bond cleavage, does not happen through catalytic aspartate, either in all substrates or in substrates containing proline at P1' position.

Inter-aspartate hydrogen bond and the scissile C-N distance.

In the peptide bond cleavage reaction, the C-N separation changes from that of a covalent bond at the start of the reaction to that of van der Waal separation at the end, with intermediate values representing intermediate stages of the reaction. Table 3.4 gives interatomic distances between inner oxygens of catalytic aspartates and the corresponding

scissile C-N distances in three substrate complexes [Kumar et al 2005, Das et al 2006, and present structure]. These distances are suggestive of a correlation between strength of the inter-aspartate hydrogen bond and the stage of the cleavage reaction. The C…N separation of 3.5 Å in the present structure is normal van der Waal separation and the steric repulsion between the scissile C and N atoms is expected to be less. The corresponding inter-aspartate hydrogen bond is relatively long and weak. Similarly, in the tetrahedral reaction intermediate before bond-breakage, where the scissile covalent bond is only stretched and not yet broken, the steric repulsion is again less and the inter-aspartate hydrogen bond is again relatively long and weak (row 1 in Table 3.4). On the other hand, when the C…N separation is significantly lower than van der Waal separation (row 2 in Table 3.4) the inter-aspartate hydrogen bond is very short and strong. Since these three structures are closely related and were obtained using almost identical experimental methodology, the correlation mentioned above may be significant. It is therefore suggested that in the reaction mechanism of HIV-1 protease, the inter-aspartate hydrogen bond modulates to make up for inter-product steric repulsion at different stages of the cleavage reaction. Modulation of hydrogen bond strengths during enzyme catalysis is widely prevalent, but these hydrogen bonds are normally between functional groups from the enzyme and the reactants. However, what is suggested here is modulation of a hydrogen bond between functional groups of the enzyme alone. This suggestion is different from the inter-aspartate hydrogen bond modulation envisaged in the two stage kinetic *iso*-mechanism proposed by Northrop [Northrop 2001].

Table 3.4: Correlation between inter-aspartate hydrogen bond length and separation between C and N atoms of the scissile peptide bond

Tethered HIV-1 protease complexed with	Hydrogen bond length (D25 OD1… D1025 OD1)	Separation between atoms C and N of the scissile peptide bond	Reference
Tetrahedral intermediate	2.98 Å	1.65 Å	[Kumar et al 2005]
Cleaved substrate	2.3 Å	2.7 Å	[Das et al 2006]
Cleaved substrate	2.8 Å	3.5 Å	Present structure

Conclusion

The crystal structure of HIV-1 protease tethered dimer complexed with a heptapeptide substrate that contains a proline residue in the P1' position has been determined to 2 Å resolution. The substrate is found cleaved in the crystals, with the two products still bound in the enzyme active site. Because of lateral displacement of the peptide in the active site, the scissile nitrogen atom of the product peptide does not form a hydrogen bond with the latter. Instead, it hydrogen bonds with the carboxyl oxygen atom from the other product peptide. This inter-product hydrogen bond was also observed in the earlier study involving non-proline containing substrate [Das et al 2006]. These observations are consistent with the suggestion that gem-hydroxyl protonates scissile nitrogen atom prior to bond cleavage [Rose et al 1996, Kovalevsky et al 2007]. The hydrogen bond between inner oxygens of catalytic aspartates in the present complex is very weak, while the scissile carbon and nitrogen atoms are at normal van der Waal separation. It is suggested that a correlation exists between the strength of the inter-aspartate hydrogen bond and the stage of peptide bond cleavage by HIV-1 protease.

Chapter - 4

Drug Resistance Probed through Crystallographic Methods

Drug resistance mutations against all the approved protease inhibitors have been reported. All these inhibitors are competitive inhibitors of HIV-1 protease *i.e.,* their effectiveness is determined by the balance between substrate affinity and inhibitor affinity. Most of these drug resistance mutations also affect the catalytic efficiency of the enzyme but the effect on substrate is less than that of inhibitor binding. Commonly found drug-resistance mutations against these drugs are shown in figure 4.1. It is clear from the table that many mutations are cross resistant to several drugs. For example, M46I, V82A, and L90M provide broad cross-resistance to most of the drugs. Drug resistance mutations are classified into primary and secondary mutations based on their contribution to resistance. Primary mutations can cause drug resistance in the absence of any other mutation. Some of the primary mutations are D30N, G48V, M46I/L, I50L, and V82A. Secondary mutations usually accompany some primary mutations and complement them by various mechanisms. These secondary mutations include N88D, A71V, M36I, etc. Some of these secondary mutations increase the fitness of mutant protease, which is diminished by primary mutations. Other secondary mutations increase the level of resistance against the drug.

Drug resistance mutations are divided into active site mutations and non-active site mutations based on their location in the HIV-1 protease structure. While majority of the primary mutations are mapped to the active site region, most secondary mutations are found to be located outside the active site. Structural studies of mutant protease-inhibitor complexes have shown that most of the active site mutations directly interfere with the binding of inhibitor to the enzyme and thereby alter the affinity of the inhibitor. However, mechanisms of drug resistance by the non-active site mutations are not properly

understood and are difficult to rationalize. It has been suggested that they may act by affecting the active site shape, conformational flexibility, and stability of the HIV-1 protease monomer and dimer [Xie et al 1999, Piana et al 2002]. Effect on the inhibitor binding via long-range structural interactions also has been suggested [Liu et al 2005].

Nelfinavir is one of the FDA approved drug used for the treatment of HIV/AIDS patients as part of HAART therapy. It was developed by Agouron Pharmaceuticals using structure based drug design strategy and Monte Carlo simulations [Gehlhaar et al 1995]. It has a novel 2-methyl-3-hydroxy benzamide group at P2 subsite and S-phenyl group at P1 subsite (Figure 4.2). A dodecahydroisoquinoline ring is present at P1' site while P2' site is occupied by tert-butylcarboxamide group (Figure 4.2). Nelfinavir is a highly potent inhibitor of HIV-1 protease with Ki of 2 nM and was the first protease inhibitor approved for use in pediatric AIDS patients. Nelfinavir has a unique mutation pattern among the protease inhibitors. Primary mutations observed against nelfinavir were D30N, L90M, and N88S [Rhee et al 2003]. In addition to these primary mutations, several secondary mutations like M36I, M46I/L, I54V, A71V, V82A, I84V, and N88D *etc.* were also observed in combination with primary mutations providing very high level of drug resistance [Rhee et al 2003]. Understanding the structural effects of these drug resistance mutations and their molecular basis of resistance using the three-dimensional atomic structures of these mutant proteases will be highly beneficial for modification of nelfinavir and future development of HIV-1 protease inhibitors. For present work, mutations D30N, L90M, N88D and N88S were chosen and their crystal structures both in unliganded form and in complex with nelfinavir were determined. This chapter is divided into three sub-chapters.

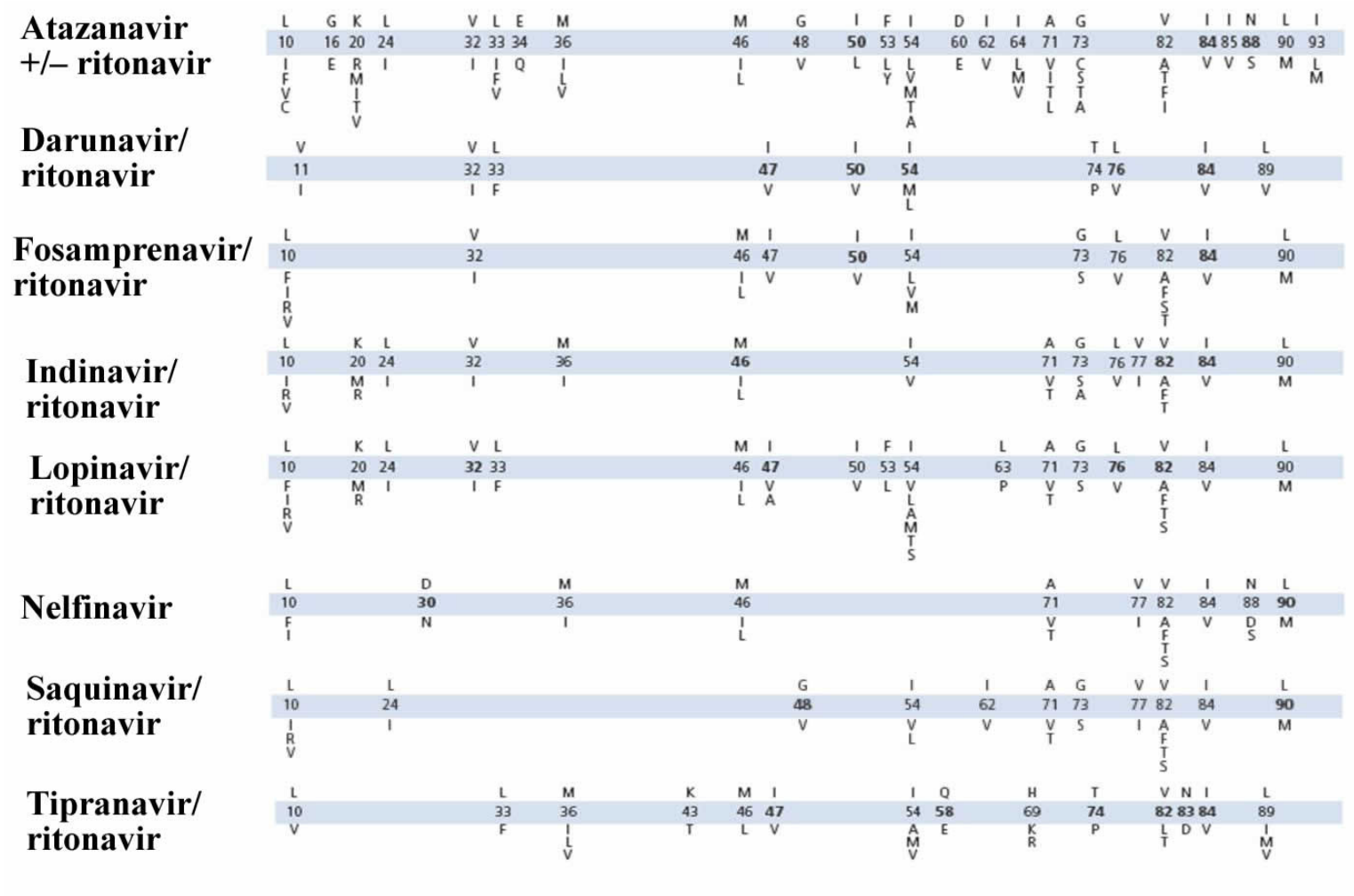

Figure 4.1: Drug resistance mutations associated with protease inhibitors. Primary mutations are shown in bold [reproduced with permission from Johnson et al (2011) Top Antivir Med 19: 156-164].

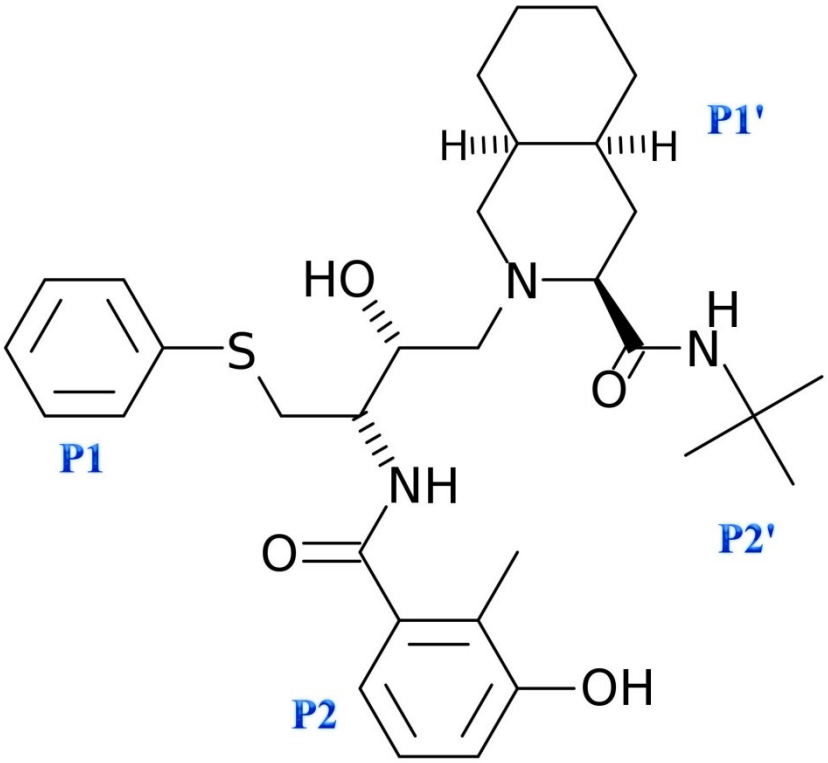

Figure 4.2: A two dimensional schematic diagram of nelfinavir.

Chapter - 4A

X-Ray Structures of D30N HIV-1 Protease in Unliganded Form and Its Nelfinavir Complex: Implications for Reduced Cleavage Activity and Drug Resistance

The mutation D30N is the most frequent mutation and is observed in 36-59% population of HIV infected patients undergoing nelfinavir therapy [Patick et al 1998, Yerly et al 2001, Perrin and Mammano 2003, Saah et al 2003]. Interestingly, mutation D30N is selected exclusively by nelfinavir and it does not provide cross resistance to any other drug [Perrin and Mammano 2003]. The mutation D30N causes ~39 fold resistance against nelfinavir as measured by PhenoSense assay [Rhee et al 2003], while the K_i value is increased approximately 20-fold compared to the wild type [Kozisek et al 2007]. D30N mutation also affects the catalytic activity of HIV-1 protease. The k_{cat} value of HIV-1 protease has been shown to decrease by 50-90 % depending upon the substrate oligopeptide used [Mahalingam et al 1999, Kozisek et al 2007]. Parera et al [2007] had observed, using phage λ based genetic screening method, that D30N HIV-1 protease retained only 2.6 % of the wild type HIV-1 protease proteolytic activity. Martinez-Picado et al [1999] have shown that the mutation D30N is highly deleterious for virus and reduces its replicative capacity. Kozisek et al [2007] earlier determined a low-resolution structure of D30N HIV-1 protease-nelfinavir complex and suggested that reduced binding affinity of nelfinavir is not due to structural changes in the active site pockets. High-resolution crystal structure of D30N mutant HIV-1 protease will help in understanding the molecular basis of drug resistance by this mutation and its exclusivity against nelfinavir. Present sub-chapter describes crystal structures of D30N HIV-1 protease both in unliganded form and in complex with nelfinavir. These structures have been determined to resolutions of 1.65 Å and 1.91 Å respectively. The structure of D30N HIV-1 protease in unliganded form shows structural effects that might be responsible for reduced catalytic activity of D30N HIV-1 protease. The structure of D30N HIV-1

protease-nelfinavir complex shows that a crucial O-H…O ionic hydrogen bond between the hydroxyl oxygen of 2-methyl-3-hydroxy benzamide group from nelfinavir and carboxyl oxygen from D30 side chain in the wild type protease is changed to an O-H…N/ O…H-N hydrogen bond in the D30N mutant. This change is likely to reduce the strength of the hydrogen bond due to loss of coulombic interaction part of the original interaction. The present structure therefore indicates that drug resistance to nelfinavir through D30N mutation is because of weakening of crucial hydrogen bond between the drug and the enzyme.

Experimental Section

Site-directed mutagenesis, purification, and crystallization of D30N HIV-1 protease

The tethered dimer of HIV-1 protease described in the chapter 2 was used in the present study. Mutation D30N was introduced into HIV-1 protease gene construct by Quick-change multi-site directed mutagenesis method [Stratagene, La Jolla, CA]. The primer was designed using Primer X software [Lapid 2003] and the sequence of the primer used was 5'-CTGGATACCGGTGCTGAT*A*ATACTGTACTGGAGGAG -3'. Incorporation of mutation in the protease gene was confirmed by DNA sequencing. Expression and purification of mutant protease is similar to that of the wild type HIV-1 protease as described in the previous chapters. The protein concentration during crystallization trials was 1-3 mg/ml in 50mM sodium acetate buffer (pH 4.5) containing 1mM DTT. Crystallization was done at room temperature by the hanging-drop vapor diffusion method. D30N HIV-1 protease crystallized under conditions similar to those for wild type HIV-1 protease described in the previous chapters. 1-6% of saturated ammonium sulfate was used as precipitant with 200/100 mM sodium citrate/sodium dihydrogen

phosphate buffer at pH 6.2. Different additives were also explored to improve the quality of the crystals. Low molecular weight poly ethylene glycols (PEGs) PEG 400 and PEG 600 were found to be effective at 0.5 to 1.0% (v/v) concentration. Crystals of D30N HIV-1 protease-nelfinavir complex were obtained by co-crystallization under two different conditions. One of the conditions is similar to that of the unliganded protease where ammonium sulfate was used as precipitant. For drug complex crystal, 1-10% DMSO was tried as additive and was found to be effective in some of the cases. However, the effects of additives were unpredictable, random, and sometime even negative. In an alternative condition, NaCl was used as precipitant with 50 mM sodium acetate buffer at pH 4.5 and DMSO 1-10% as additive. Concentration of NaCl was systematically varied from 0.25 M to 1.75 M and it was found that 0.75 M – 1.25 M sodium chloride was best for obtaining good quality crystals. Effect of DMSO concentration was explored systematically by using from 1% to 10% v/v DMSO with different precipitant concentration. It was found that optimum concentration of DMSO was dependent on precipitant concentration and protein preparation.

Diffraction data collection

X-ray diffraction data were collected by the oscillation method, under cryo conditions at ESRF, using a crystal equilibrated in cryoprotectant solution (25% glycerol and 75% reservoir solution). For the D30N unliganded crystal, 70 frames, each for an oscillation angle of 1° were collected on ID14-1 beamline. For the D30N-nelfinavir complex crystal, 90 frames, each for an oscillation angle of 1° were collected on FIP beamline. The oscillation frames were indexed, integrated and scaled with program suite XDS [Kabsch 2010]. Data collection statistics for both the crystals are summarized in Table 4.1.

Structure refinement

Crystals of D30N HIV-1 protease and D30N HIV-1 protease - nelfinavir complex were isomorphous with crystals of the unliganded enzyme [PDB ID 1LV1]. Therefore, these structures were determined by difference-Fourier methods, using protein part of the unliganded tethered HIV-1 protease dimer (PDB ID 1LV1) as the phasing model. The D30N HIV-1 protease unliganded structure was refined with crystallography software suite PHENIX [Adams et al 2010]. First, a rigid body refinement and simulated annealing step was performed and only restrained refinement without simulated annealing was used at all subsequent stages. Water molecules were identified by the PHENIX program, which were later validated during manual model building stage using programs O [Jones et al 1991] and Coot [Emsley and Cowtan 2004, Emsley et al 2010]. Likelihood weighted electron density maps were used for manual model building. Translational/Liberation/Screw (TLS) refinement was performed at final stages [Painter and Merritt 2006a]. TLS groups were identified using TLSMD server [Painter and Merritt 2006b]. D30N HIV-1 protease-nelfinavir complex structure was refined using computer program CNS [Brunger et al 1998a, b]. The molecular model was first subjected to rigid body refinement and standard simulated annealing torsional dynamics refinement protocols [Brunger et al 1998 a, b] and at later stages, restrained refinement without simulated annealing was employed. The topology and parameter files for nelfinavir molecule were generated on the HIC-UP server [Kleywegt 2007, Kleywegt and Jones 1998]. Likelihood weighted electron density maps were interpreted using the computer graphics programs O [Jones et al 1991] and Coot [Emsley and Cowtan 2004, Emsley et al 2010]. Water molecules were picked automatically using the molecular

graphics program WinCoot [Lohkamp et al 2005]. These water positions were then edited by visual inspection of both 2Fo-Fc and Fo-Fc electron density maps. Molecular superpositions in three dimensions were carried out using softwares O, Coot, and CCP4 suite. All figures were prepared using the software PyMol [Schrodinger, LLC].

Thermal denaturation analysis

Far UV Circular Dichroism (CD) spectra on wild type and mutant proteins were recorded using a Jasco J-815 CD spectropolarimeter. Protein solutions were used at a concentration of 10 mM 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 20-80ºC with a scan rate of 2 ºC/min. Melting temperature (T_m) was calculated as the temperature where the first derivative of the melting profile *i.e.,* ellipticity versus temperature graph reached a minimum.

Results

Melting temperature

Thermal unfolding of wild type and mutant HIV-1 protease were studied using CD spectroscopy and melting temperature (T_m) was calculated. For wild type HIV-1 protease the T_m is 68ºC whereas for the D30N HIV-1 protease, it decreases to 63ºC.

Table 4.1: Crystallographic statistics for D30N and D30N- nelfinavir complex structures.

	D30N	D30N-Nelf
Wavelength of radiation used (Å)	0.934000	0.979778
Resolution range used in refinement (Å)	32.73-1.65	45 - 1.91
Space group	$P6_1$	$P6_1$
Unit cell dimensions (Å)	a=b= 62.37, c= 82.3	a=b= 62.05, c= 81.46
Asymmetric unit	One tethered dimer	One tethered dimer
Total number of reflections measured	121767	58940
Total number of unique reflections	21751	13778
R_{merge} -F (%)	7.9 (53.1)	12.7 (49.7)
Completeness (%)	99.5 (97.8)	99.5 (99.2)
$\langle I/\sigma(I) \rangle$	24.23 (3.06)	12.46 (3.08)
R_{cryst} (%)	18.87	21.87
R_{free} (%)	22.43	24.66
Number of water molecules	197	178
RMSD from ideal values		
Bond lengths (Å)	0.007	0.007
Bond angles (°)	1.071	1.34

Crystal structures

The D30N HIV-1 protease and D30N HIV-1 protease-nelfinavir complex both crystallized in $P6_1$ space group and contain one tethered dimer in asymmetric unit. Crystals of unliganded D30N HIV-1 protease had low mosaicity of 0.131º and diffracted to 1.65 Å resolution while crystals of D30N-nelfinavir complex diffracted to 1.91 Å, and displayed higher mosaicity of 0.344º. D30N HIV-1 protease and D30N HIV-1 protease-nelfinavir complex structures were refined to R_{free} values of 22.43% and 24.66% respectively with very good stereochemistry. Refinement statistics for both the structures are given in Table 4.1.

Two fold orientation of nelfinavir

A clear difference density was visible in the active site of D30N HIV-1 protease-nelfinavir complex structure. The difference density displayed approximate two-fold symmetry suggesting that nelfinavir is bound in two orientations corresponding to the pseudo-two fold symmetry of the active site. Therefore, nelfinavir is modeled with two alternate conformations (M and N) (Figure 4.3). The two orientations M and N were included with occupancy of 0.5 each at the start of the refinement. The occupancies of M and N orientations of the drug were systematically varied in steps of 0.05, but maintaining their sum to be unity. The R_{free} and R_{cryst} were lowest when the occupancies of M/N orientations were 0.45/0.55.

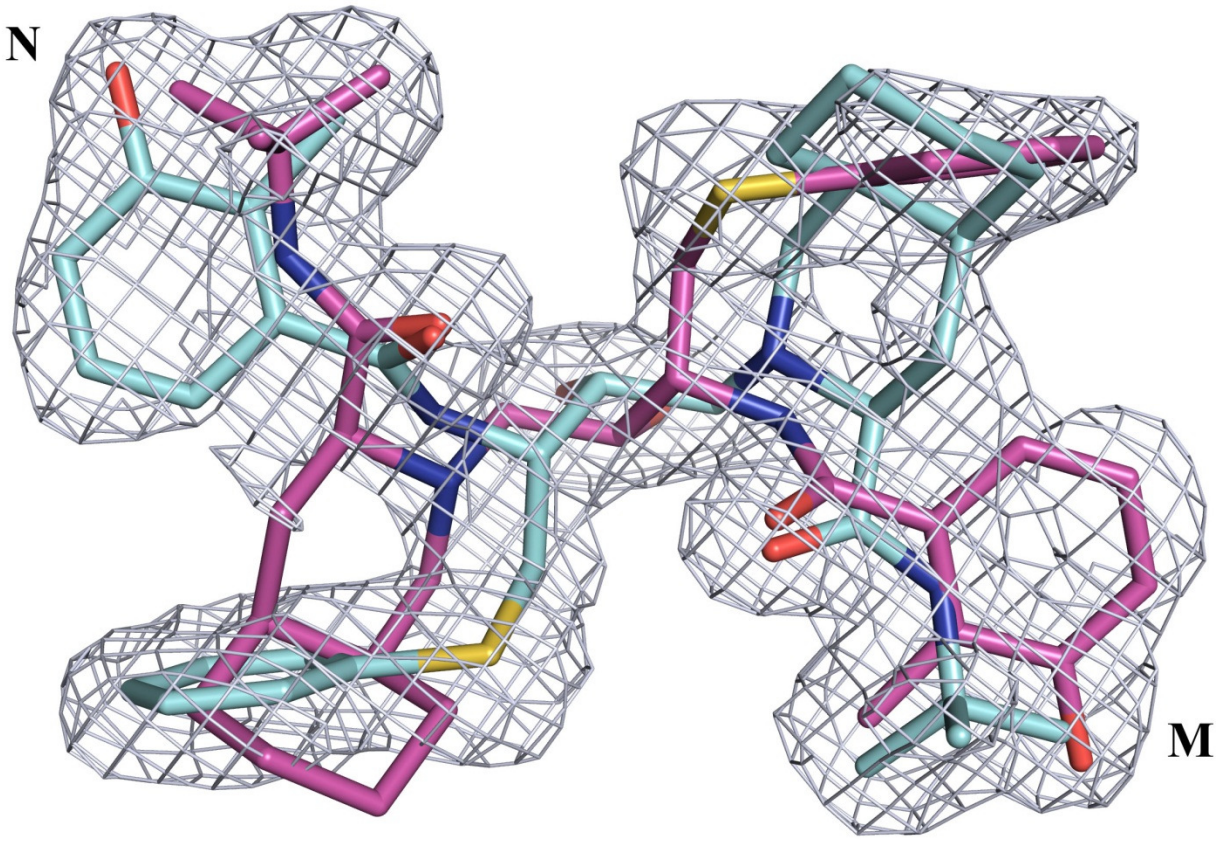

Figure 4.3: Two fold orientation of nelfinavir and 2mFo-dFc map at 1σ level. Orientation M is shown in magenta and N in cyan.

Determining the rotamer of asparagine 30

Correct assignment of the orientation of side chain amide group of asparagine is rather difficult because hydrogen atoms on NH_2 cannot be identified in X-ray structures except at extremely high resolution, and N and O atoms cannot be differentiated by direct observation of the electron density map. One of the approaches to overcome this problem is based on the hydrogen bonding interactions of the side chain. In the present structure, hydroxyl O of 2-methyl-3-hydroxyl benzamide group from nelfinavir can act as either an acceptor or donor for hydrogen bonding. In such scenarios, where the interacting atoms are both potential donors and acceptors, correct assignment requires further analysis of the surrounding environment as well [Word et al 1999]. Therefore, in present structure, asparagine 30 was built in both the orientation and the two scenarios were compared. These two orientations are described as Nas or Oas respectively depending on whether ND2 or OD1 of N30 is towards the drug. Nas rotamer is slightly favored over Oas rotamer configuration in standard rotamer library for asparagine. For the Nas orientation the R_{free} was only slightly lower (0.05%) than for Oas orientation. Analysis of the hydrogen bonding interactions after adding hydrogen atoms using reduce tool of MolProbity server [Davis et al 2007, Chen et al 2010] suggested that both situations are equally likely.

Molecular conformation: comparison of D30N protease with wild-type HIV-1 protease

The overall conformation of the protein in the D30N unliganded structure was very similar to that of the unliganded HIV-1 protease structure [PDB ID 1LV1], as the two structures superposed with an overall RMSD of 0.318 Å for 198 Cα pairs. This suggests that D30N mutation does not have any major structural effect on HIV-1 protease. All the

residues that deviated by more than 0.5 Å were found to be on the surface and in the loop regions, which are intrinsically flexible, and have been found to have higher RMSDs in different HIV-1 protease structures. However, a closer look of the structure shows that D30N mutation do have some effect on the less-flexible residues in the active site loop (Figure 4.4). Cα atoms of 30N/1030N moved by 0.26 Å/ 0.25 Å while maximum side chain deviation is 0.78 Å /1.05 Å. Residues from 23-31 in both subunits also have shown some displacements with RMSD ranging from 0.14 Å/0.07 Å to 0.26 Å/0.28 Å, which is larger than what is normally observed for these residues. Cα atoms of catalytic aspartates from both the subunits show deviations of 0.26 Å/0.28 Å. Catalytic residues are one of the most rigid parts of the HIV-1 protease structure. Overall effect of these variations is that the active site loop is pushed down towards the dimer interface (Figure 4.4). The volume of the active site cavity calculated using CASTp server [Binkowski et al 2003] in the present structure is 1058.5 $Å^3$ which is approximately 6% higher as compared to the 996.6 $Å^3$ observed for wild-type structure [PDB ID 1LV1].

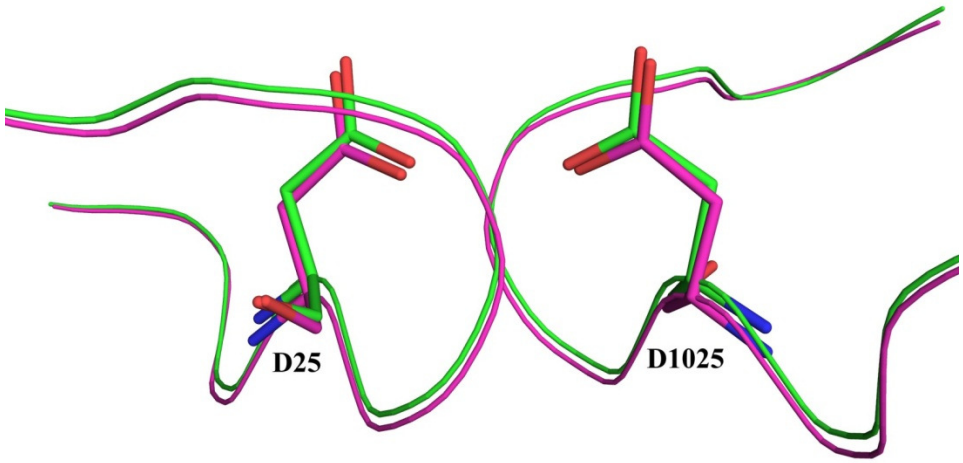

Figure 4.4: Deviations at catalytic loop in D30N HIV-1 protease (magenta) as compared to the wild type HIV-1 protease (green). Catalytic aspartates are also shown.

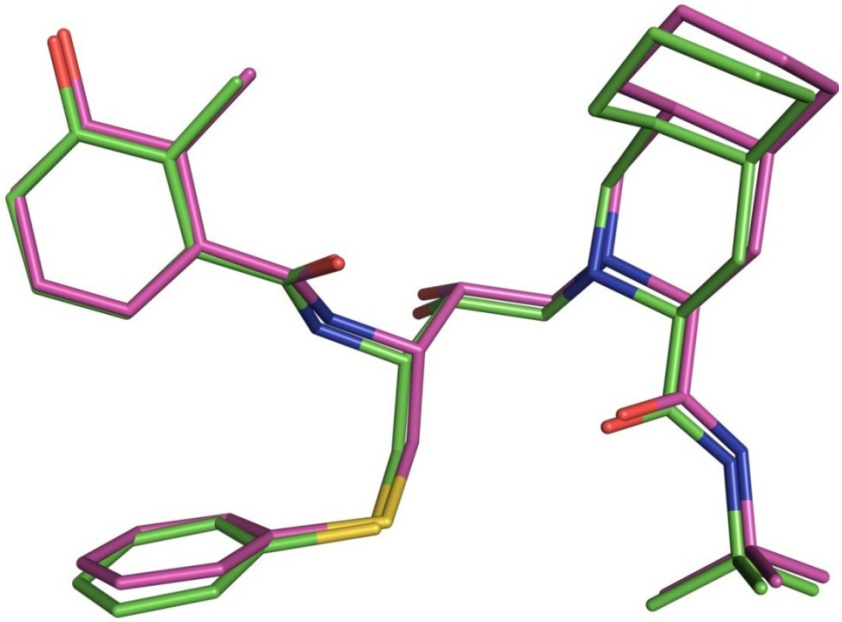

Figure 4.5. Superposition of nelfinavir in D30N-nelfinavir complex (magenta) with wild-type-nelfinavir complex structure (green).

Molecular conformation: comparison of D30N protease-nelfinavir complex with wild-type protease-nelfinavir complex

The overall conformation of the protein in the D30N complex structure was very similar to that of the wild type-nelfinavir complex structure (PDB ID 1OHR), with RMSD of 0.629 Å for 198 Cα pairs, despite having been crystallized in different space groups. The present structure was very symmetric as the two subunits superimposes on each other with RMSD of 0.152 Å. Interestingly, in the wild-type complex (PDB ID 1OHR), where nelfinavir molecule binds in a single orientation, the RMSD for superposition of the two subunits is somewhat higher at 0.55 Å. Notable deviations in the two structures are at residues that are on the surface and are part of crystal contacts. The deviations of Cα atoms at residue positions 29 and 30 were less than 0.3 Å. Interestingly, theoretical simulations by Ode et al., predicted larger shifts of the order of 0.7 Å– 0.8 Å for these residues [Ode et al 2007]. The orientation and conformation of nelfinavir in the active site of the mutant complex is also similar to that of the wild type complex with small displacements distributed in all four subsites (Figure 4.5). The volume of the active site cavities calculated by ignoring the nelfinavir in the present structure is 1109 $Å^3$, which is approximately 6.5 % lower than 1186.1 $Å^3$ observed for the wild-type complex.

D30N HIV-1 protease-nelfinavir hydrogen bonding interactions and comparison with the wild type complex structure

The hydrogen bonding interactions of nelfinavir with D30N HIV-1 protease are shown in Figure 4.6. Nelfinavir molecule has four oxygens in two hydroxyl and two carbonyl groups, which can participate in hydrogen bonding interactions. All these oxygens are involved in hydrogen bonding with the enzyme, either directly or through water

molecule, in the present structure. Central secondary hydroxyl group (O21) is at distances of 2.6 Å- 3.1 Å from all four oxygens of the catalytic aspartates. These interactions are similar to those found in all the structures of protease complexed with hydroxyl containing inhibitors. Flap water, mediates hydrogen bonds between main chain amide nitrogen atoms of I50 and I1050 of protease and carbonyl oxygens O17 and O25 of nelfinavir molecule. Nelfinavir in the wild type HIV-1 protease also makes similar interactions. However, there is difference in the hydrogen bonding at the mutation site. In the wild type protease, the D30 side chain is involved in a strong (d = 2.9 Å) hydrogen bonding interaction with the hydroxyl oxygen O38 of 2-methyl-3-hydroxy benzamide group from nelfinavir (Figure 4.7). In the present structure, mutation D30N causes slight deviations in the main chain and in the orientation of side chain of asparagine resulting into a weak hydrogen bond (d = 3.2 Å/3.1 Å) with N30 ND2/OD1 (Figure 4.7). Of the three nitrogen atoms in nelfinavir, N12 interacts with residue D1029 and G1027 through water mediated hydrogen bonds. Similar interactions are found in the wild type structure. These interactions are probably responsible for the proper positioning of the tert-butyl group. N22 atom of nelfinavir forms hydrogen bonds to main chain carbonyl O of G27 in both the wild type and D30N mutant HIV-1 protease.

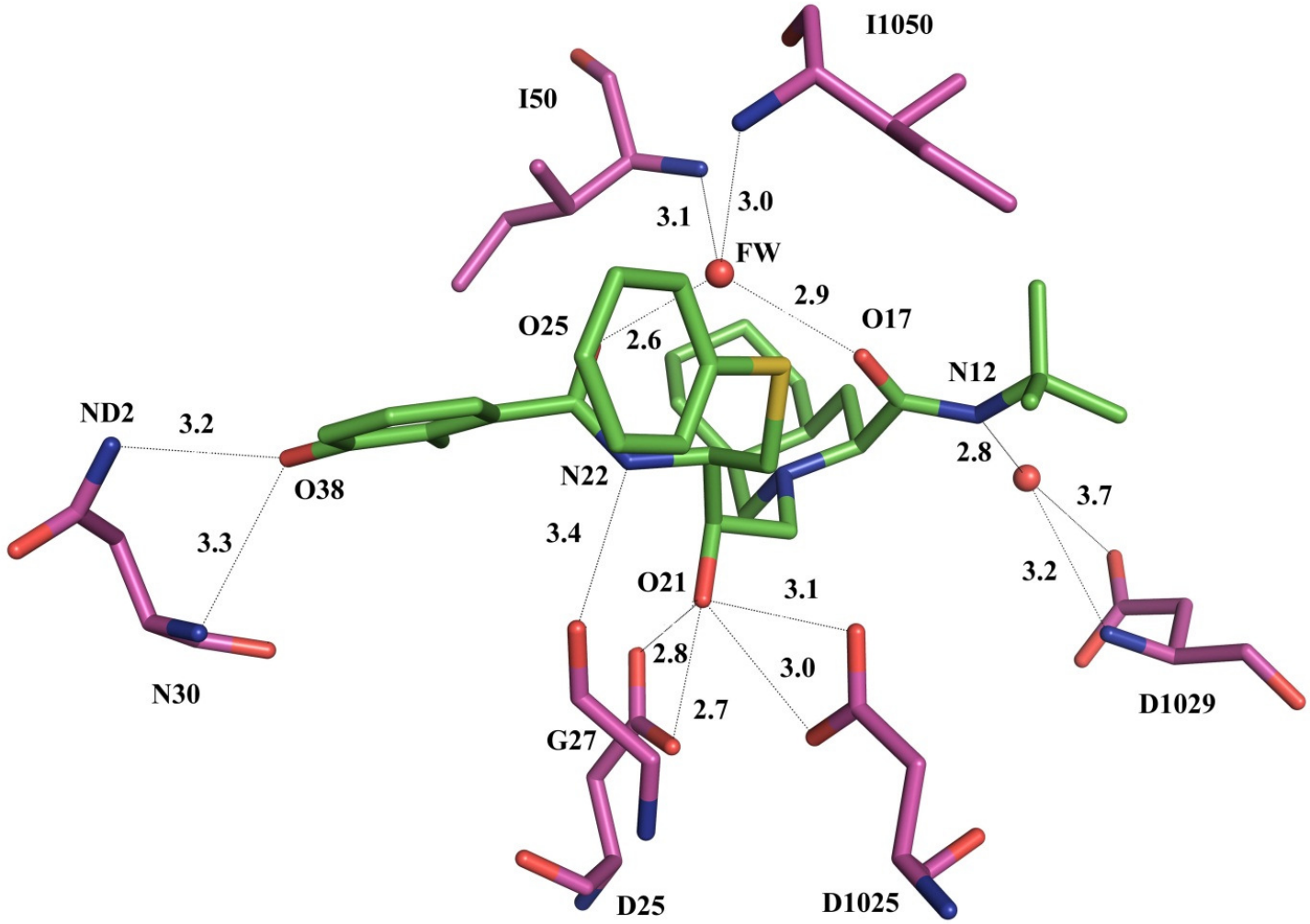

Figure 4.6: Hydrogen bonding interactions of nelfinavir in D30N HIV-1 protease. Only orientation M is shown here. Nelfinavir is shown in green and protein residues interacting with the drug are shown in magenta. Flap water is labeled as FW.

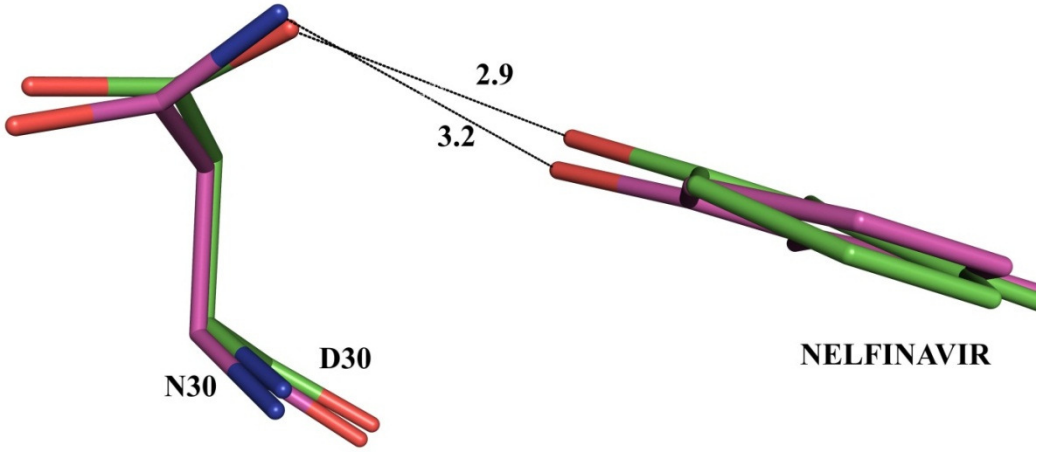

Figure 4.7: Weakening of hydrogen bond between nelfinavir and 30th side chain in D30N HIV-1 protease (magenta) as compared to wild type (green).

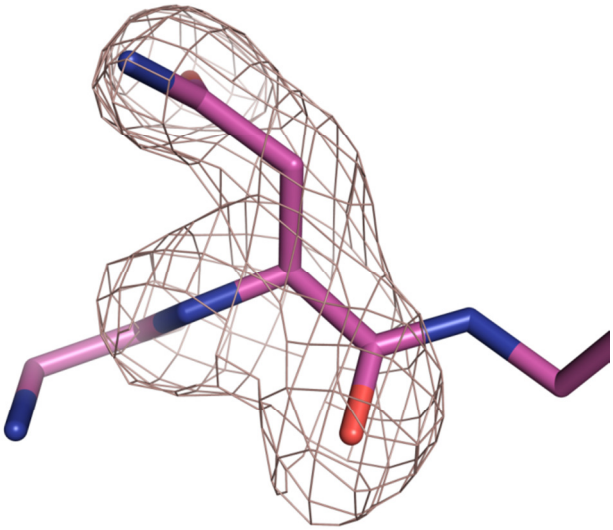

Figure 4.8: Simulated annealed omit map for N30 in D30N protease at 3σ level.

Discussion

Drug resistance against nelfinavir

Figures 4.8, 4.9 and 4.10, show the simulated annealed OMIT density calculated by omitting either the whole amino acid residue N30/1030 or only the ND2 atom in the side chain. It is clear that, position of the 30/1030 ND2/OD1 atom in the present structure have been determined very accurately. Therefore, the increase in hydrogen bond length with the drug appears to be genuine and significant. Present structure also shows that there can be two different possibilities of hydrogen bonding in D30N-nelfinavir complex. In one orientation, the drug is a donor in an O-H…O hydrogen bond, while in alternative arrangement; the drug is an acceptor in an O…H-N hydrogen bond. Our analysis based on rotamer preference, R-factors and hydrogen bonding interactions suggest that both the orientations are equally likely and may be present at different time. In both the scenarios, the nature of hydrogen bond will be different in the mutant structure as compared to that of the wild type. In wild type HIV-1 protease, a charged acceptor is hydrogen bonded to a neutral donor whereas in both scenarios of the D30N protease a neutral donor will be hydrogen bonded to a neutral acceptor. It has been calculated that the free energy change associated with the first situation is 16-23 kcal/mol while for second situation, it is much smaller 3-6 kcal/mol [reviewed in Rose, 1993]. Therefore, both the change in the nature of hydrogen bond and the increase in the distance are likely to result into a relatively weak hydrogen bond between nelfinavir and D30N HIV-1 protease.

Overall conformation of nelfinavir in the present structure is also slightly deviated from the wild type (Figure 4.5). All the four sub-sites are found to be affected by these

deviations; however, overall number of non-bonded interactions shorter than 4 Å, remained similar to that of the wild type-complex structure. Active site volume of the mutant HIV-1 protease is little higher than that of the wild type protease in unliganded conditions but smaller in the complex structures as compared to the wild type-nelfinavir complex structure. This shows that nelfinavir is not bound optimally in the mutant protease. Due to the change in conformation of the drug and change in the active site volume, overall normalized complementarity between nelfinavir and active site calculated using LPC software [Sobolev et al 1999], is reduced from 0.79 in the wild type protease to 0.67 in the D30N HIV-1 protease.

Binding of a competitive inhibitor is determined by the balance of affinity of drug and substrate to the protein. A subtle change in this balance of the relative affinities can lead to high level of drug resistance. Both the suboptimal binding in the active site and decrease in the strength of crucial hydrogen bond should result into loss of binding affinity of nelfinavir. This is likely to disrupt the relative balance of affinities of inhibitor and substrates and lead to the drug resistance seen in HIV/AIDS patients.

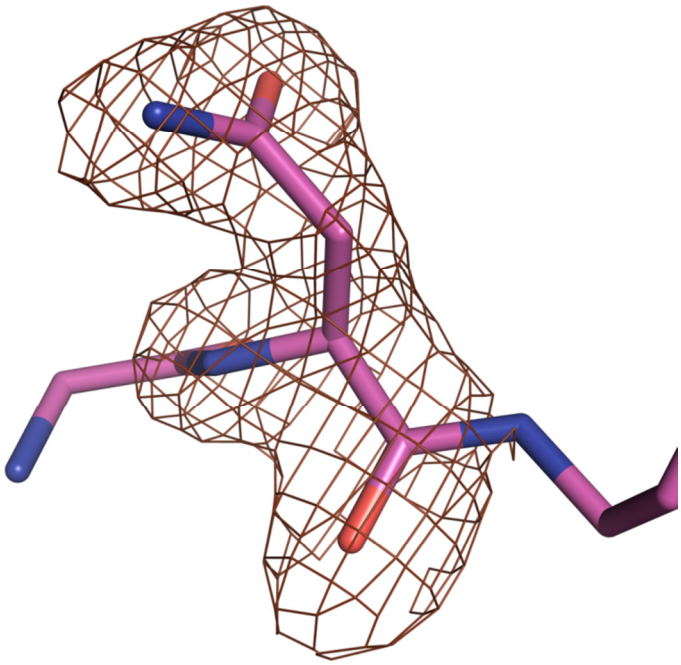

Figure 4.9: Simulated annealed omit map for N1030 in D30N protease at 3σ level

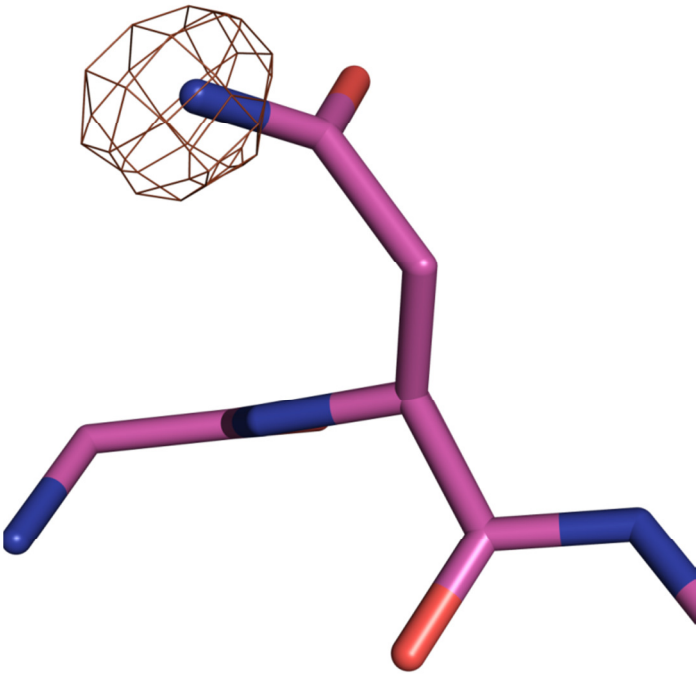

Figure 4.10: Simulated annealed omit map for N1030 ND2 in D30N protease at 3σ level

D30N mutation and cross resistance to other drugs

The mutation D30N is selected exclusively by nelfinavir and does not provide cross resistance to any other drug. Atazanavir, however shows some minor resistance by D30N mutation [Rhee et al 2003]. It has also been suggested that darunavir might also select for this mutation [Kovalevski et al 2006]. Table 4.2 lists interactions of different drugs with the residue 30 of the wild type HIV-1 protease. It can be seen that nelfinavir interacts strongly through a direct hydrogen bond with side chain of the D30 and mutation D30N drastically reduces its affinity to mutant protease. Ritonavir, indinavir, saquinavir, amprenavir, tipranavir and lopinavir do not form any hydrogen bond with the D30 side chain so they are not affected by the D30N substitution. Atazanavir forms water-mediated hydrogen bond, which should not be contributing significantly to the binding affinity of the drug. In accordance with this suggestion, D30N substitution only provides 2.4 ~fold resistance against the drug atazanavir [Rhee et al 2003]. Recently, Kovalevsky et al [2006] have shown that the D30N mutation gives resistance against the new inhibitor TMC-114 (darunavir). These authors attribute the loss in binding affinity to substitution of a direct hydrogen bond by a water-mediated hydrogen bond. Thus, there appears to be a correlation between efficacy of D30N as a drug resistance mutation and direct hydrogen bonding of the drug with D30.

Table 4.2: Interaction between different protease inhibitors and side chain of the residue 30 of HIV-1 protease

Drug	Interaction with D30	PDB Id
Nelfinavir	Direct strong hydrogen bond	1OHR
Saquinavir	No or very weak hydrogen bond	1HXB
Ritonavir	No hydrogen bond	1RL8
Indinavir	No hydrogen bond	1SDT
Amprenavir	No hydrogen bond	1T7J
Tipranavir	No or very weak hydrogen bond	2O4P
Darunavir	Direct strong hydrogen bond	2HS1
Atazanavir	Water mediated contact	2O4K
Lopinavir	No hydrogen bond	2O4S

Reduced cleavage activity of D30N protease

D30 is located in the substrate binding site in HIV-1 protease and has been shown by X-ray structural studies to interact with many substrates involving both main chain and side chain atoms from D30 [Prabhu-Jeyabalan et al 2002]. The D30N mutation will have two different effects on the direct interaction of substrate and side chain of 30^{th} residue. First, it removes the charge on the 30^{th} residue, which is likely to affect the catalytic activity of the protease with respect to those substrates where this charge interaction plays some role in cleavage efficiency. Second, deviation in the side chain seen in present structure is likely to increase the hydrogen bonding distance of this interaction. Both, the removal of charge and increase in hydrogen bonding distance will weaken the hydrogen bond strength and is likely to affect those substrates where the strength of this hydrogen bond is important to hold substrates in optimum position and orientation. Further, deviations in the active site loop, including catalytic aspartates led to an increase in the active site volume of D30N HIV-1 protease. HIV-1 protease active site is evolved to recognize substrates with different sequences primarily based on their shape. The shape complementarity between the substrates and HIV-1 protease active site is very high (0.65-0.7) and a total of 900-1000 $Å^2$ accessible substrate surface is buried upon binding [Prabhu-Jeyabalan et al 2002]. Increase of the active site volume as seen in the present structure will negatively affect this shape complementarity and hamper the proper positioning of substrates in the active site. These structural effects on the binding of substrates along with weakened hydrogen bond with 30^{th} residue are likely to affect the catalytic activity of the enzyme.

D30N and reduced replicative capacity of the virus

Thermal denaturation analysis of the wild type and D30N mutant HIV-1 protease carried out using CD spectropolarimeter showed that D30N mutant protease is less stable than the wild type counterpart is. Since decrease in T_m in mutant proteases is previously attributed to lower dimer stability [Mahalingam et al 1999], present structure was analyzed for its dimer interface properties using various methods. Gap volume analysis was done using ProtorP server [Reynolds et al 2009]. Gap volume index for wild type protease (PDB ID 1LV1) is 0.74, which decreases to 0.70 in D30N HIV-1 protease. The gap volume index is a measure of the complementarity between the two subunits and tight packing of the interface [Reynolds et al 2009]. The reduction of gap volume in D30N protease suggests that dimer is in fact getting stronger in mutant protease [Reynolds et al 2009]. Number of interactions (less than 4 Å) between the two monomers is also increased to 383 in D30N protease as compared to 325 observed in the wild type protease (PDB ID 1LV1). This dichotomy of low melting temperature and strong dimer interface can be explained by the effect of D30N mutation on folding. D30 is part of structurally conserved loop (24-34), which has been shown to be the part of HIV-1 protease folding core [Wallqvist et al 1998, Broglia et al 2008, Verkhivker et al 2008, Bonomi et al 2010]. Mutation D30N might affect the stability of this folding core, thereby prompting the unfolding of mutant protease at lower temperature. This deleterious effect of D30N mutation on the folding of the enzyme along with the lower catalytic activity may be the reason of reduced replicative capacity of virus carrying D30N HIV-1 protease.

Conclusion

Crystal structure of D30N mutant HIV-1 protease-nelfinavir complex shows that the hydrogen bonding interaction between O38 of nelfinavir and carboxyl O of residue 30 in the wild type protease is substantially weakened in the mutant. This weakening is due to increase in interatomic distance and loss of ionic nature of hydrogen bond. Displacements in the catalytic loop resulted into less than optimum binding of the drug to HIV-1 protease active site. Both these effects are likely to be the major reasons for the loss of affinity of nelfinavir against D30N HIV-1 protease. A similar weakening of the hydrogen bond during substrate binding and deviations in the catalytic loop are suggested to result into reduced cleavage activity of the D30N HIV-1 protease. Effect of D30N substitution on the stability of folding core along with decreased catalytic activity is proposed to be the reason behind reduced replicative capacity of the HIV harboring D30N HIV-1 protease.

Chapter - 4B

Structural Investigation into Exclusivity of Active Site D30N and Non-Active Site L90M Drug Resistance Mutations in HIV-1 Protease: Effect of Improper Folding?

The mutation D30N is the most frequent primary mutation against nelfinavir [Patick et al 1998]. The second primary mutation selected by nelfinavir is L90M, where leucine at 90th position is substituted by methionine. The L90M has been found in 12% of patients, as compared to D30N, which is observed in 36-38% of patients [Yerly 2001, Perrin and Mammano 2003]. In a smaller study, Saah et al [2003] found that 59% of the patients showing viral rebound during nelfinavir therapy had D30N substitution while 38% had L90M substitution. As discussed earlier, D30N is selected specifically by nelfinavir and has not been observed in patients showing resistance to other HIV-1 protease inhibitors [Perrin and Mammano 2003]. On the other hand, L90M mutation provides broad cross resistance to other protease inhibitors [Rhee et al 2003]. D30N provide higher level of resistance against nelfinavir than L90M [Gonzalez et al 2004]. A general trend in HIV-1 protease drug resistance is the accumulation of two or more drug resistance mutations resulting into even higher level of drug resistance. However, combination of D30N and L90M substitutions is only rarely found *in-vivo* [Sugiura et al 2002] suggesting that these substitutions are mutually exclusive [Saah et al 2003]. It is further interesting that in some rare *in-vivo* and *in-vitro* studies, combination of the D30N and L90M have been found to cause very high level of drug resistance against nelfinavir. Stanford database shows that D30N/L90M double mutant provides ~ 81-fold resistance against nelfinavir [Rhee et al 2003]. It has been shown by *in-vitro* studies that presence of both these substitutions in HIV-1 protease strongly impaired the functioning of the enzyme [Sugiura et al 2002]. Recombinant viruses carrying HIV-1 protease gene with both D30N and L90M mutations were less fit with decreased replicative capacity than the wild type HIV-1 [Martinez-Picado et al 1999, Gonzalez et al 2004]. However, presence of additional

mutations such as N88D can compensate for the loss of activity and allow the protease variants harboring both these mutations to appear [Martinez-Picado et al 1999, Sugiura et al 2002, and Saah et al 2003]. These observations suggest that in the patients showing resistance against nelfinavir, HIV-1 protease evolves along two mutually exclusive pathways characterized by the presence of either D30N or L90M mutations. Kozisek et al [2007] have determined medium resolution structure of D30N/L90M-nelfinavir complex and suggested that unfavorable interactions of L90M with the catalytic loop destabilize the dimer interface. To understand the mutual exclusivity of these two mutations and mechanism of high level of drug resistance caused by their combination, crystal structures of D30N/L90M double mutant HIV-1 protease both in unliganded form and in complex with nelfinavir were determined, and these structures are described in the present chapter.

Experimental Procedures

Site-directed mutagenesis

Mutation L90M was introduced on D30N HIV-1 protease backbone by Quick-change multi-site directed mutagenesis method [Stratagene, La Jolla, CA]. The primer was designed with the help of software primerX [Lapid 2003] and the sequence of the primer used for mutagenesis was CGGCCGTAACCTG*A*TGACTCAGATCGG. Incorporation of mutation in the protease gene was confirmed by DNA sequencing.

Expression and purification of mutant HIV-1 protease

Expression and purification of mutant protease were done as described in the previous chapters for wild type and D30N HIV-1 protease. The purified protein was concentrated

using centrifugal concentrator and exchanged into 50mM sodium acetate buffer (pH 4.5) containing 1mM DTT.

Thermal denaturation analysis

Far UV Circular Dichroism (CD) spectrum on D30N/L90M double mutant HIV-1 protease was recorded with protein concentration of 10 mM in 50 mM sodium acetate buffer, pH 4.5. Thermal denaturation experiments were carried out over a temperature range of 20-80 ºC with a scan rate of 2 ºC/min and CD signal was monitored at wavelength 215 nm.

Crystallization of D30N/L90M HIV-1 protease

Final protein concentration used for crystallization was in the range of 3-5 mg/ml in 50 mM sodium acetate buffer (pH 4.5) containing 1mM DTT. Both D30N/L90M double mutant HIV-1 protease in unliganded form and its nelfinavir complex could be successfully crystallized under conditions similar to that observed for wild type and D30N HIV-1 protease.

Diffraction data collection

X-ray diffraction data were collected under cryo-conditions, at ESRF with an oscillation angle of 1 degree. For D30N/L90M unliganded crystal, 120 frames were collected at ID29 beamline while for D30N/L90M-nelfinavir complex crystal 60 frames were collected at FIP beamline. The oscillation frames were indexed, integrated and scaled with program suite XDS [Kabsch 2010]. The crystals diffracted to resolutions of 1.8 Å and 2.0 Å for D30N/L90M unliganded crystal and D30N/L90M-nelfinavir complex crystal respectively. Data collection statistics are summarized in Table 4.3.

Structure refinement

The structures of the both unliganded D30N/L90M HIV-1 protease and nelfinavir complex were determined by difference Fourier methods, using phases from the unliganded tethered HIV-1 protease structure [PDB Id 1LV1]. Refinement was carried out using the software suite PHENIX [Adams et al 2010]. X-ray diffraction data was processed in $P6_1$ space group. Analysis of the diffraction data using Xtriage module of PHENIX suggested the possibility of either space group higher than $P6_1$ or the presence of twin. However, because of the presence of a pentapeptide tether used to join the two sub-units of HIV-1 protease in the present structure, the space group higher than $P6_1$ is ruled out. Possibility of twinning was further explored. Robbins et al [2010] have discussed the presence of twinning and its effect on the data statistics while determining the structure of HIV-1 protease from subtype A virus. In the first cycle, the molecular model was obtained from wild type structure by removing all the water from the pdb and it was subjected to rigid body refinement with standard simulated annealing refinement protocol**.** Twin refinement as implemented in PHENIX was used in subsequent refinement steps with h, -h-k, -l as twin parameter. Refinement with twinning led to a 2.5-3.5% drop in R_{free} when compared with an identical refinement cycle without twinning, which verify the presence of twinning [Robbins et al 2010]. About 5% of the total reflections were set aside to provide a test set for cross validation calculations. R_{free} set was generated considering highest lattice symmetry, which ensures that reflections related by twin operator are either in the test set or in the working set. Electron density maps produced by PHENIX were interpreted using the computer graphics program O [Jones et al 1991] and Coot [Emsley and Cowtan 2004, Emsley et al 2010]. Translational/

liberation/screw (TLS) refinement was performed at final stages [Painter and Merritt 2006a]. TLS groups were identified using TLSMD server [Painter and Merritt 2006b]. Water molecules were picked automatically using the program PHENIX. These water positions were edited by visual inspection of electron density maps.

Table 4.3: Crystallographic statistics for D30N/L90M unliganded and D30N/L90M-nelfinavir complex structures.

	D30N/L90M	D30N/L90M-Nelf
Wavelength of radiation used (Å)	0.949821	0.979734
Resolution range used in refinement (Å)	44.9 - 1.8	45.4 - 2.0
Space group	$P6_1$	$P6_1$
Unit cell dimensions (Å)	a=b= 61.98, c= 82.11	a=b= 62.79, c= 82.32
Asymmetric unit	One tethered dimer	One tethered dimer
Total number of reflections measured	119140	34508
Total number of unique reflections	16269	12276
R_{merge}-F (%)	8.6 (53.9)	15.6 (79.7)
Completeness (%)	97.9 (96.6)	98.3 (97.4)
$\langle I/\sigma(I) \rangle$	23.45 (3.24)	13.15 (2.09)
R_{cryst} (%)	18.09	16.61
R_{free} (%)	22.21	21.67
Number of water molecules	151	116
RMSD from ideal values		
Bond lengths (Å)	0.003	0.007
Bond angles (º)	0.729	0.853

Results

Determining the orientation of the side chain of asparagine 30

Hydrogen atoms were added using MolProbity server to facilitate the analysis of the hydrogen bonding interactions [Davis et al 2007, Chen et al 2010]. Similar to the D30N HIV-1 protease, in D30N/L90M also, correct assignments of the orientation of side chain amide group of N30/1030 were analyzed using the hydrogen bonding interactions. As in the D30N structure, here also the side chain orientations are described as Nas or Oas respectively, based on whether ND2 or OD1 of N30 is facing the drug molecule. In the D30N/L90M HIV-1 protease complex structure, R_{free} for Oas orientation was slightly lower (0.03%) than for Nas orientation. Based on hydrogen bonding analysis and overall local interaction analysis using MolProbity server, it could not be unambiguously decided which orientation is better, so it is likely that both the orientations are present in the protein at different times.

Comparison of D30N/L90M structures with wild type structures

Similarity or differences of D30N/L90M unliganded structure with the wild type protease were analyzed using least square superposition as implemented in O [Jones et al 1991]. The overall conformation of the protein was found to be similar to that of the wild type HIV-1 protease structure [PDB Id 1LV1] with RMSD of 0.383 Å for 198 Cα pairs. In the structure of D30N HIV-1 protease, it was found that D30N mutation does not have large structural effect. By comparing the D30N/L90M double mutant unliganded structures with the D30N HIV-1 protease structure, it will be possible to analyze the effect of L90M mutation onto the structure of HIV-1 protease. The two structures were superposed with RMSD of 0.307 Å for 198 Cα pairs. The tethered dimer in the D30N was much more

symmetric with RMSD of 0.152 Å for 99 Cα pairs when the two sub-units were superposed on each other while corresponding value for the D30N/L90M HIV-1 protease is 0.362 Å. In comparison, the two sub-units in the wild type HIV-1 protease superpose to RMSD of 0.199 Å. This suggests that effect of L90M substitution is different in the two sub-units and that it makes the two sub-units asymmetric. The D30N/L90M-nelfinavir complex structure was compared with the nelfinavir complex structures of wild type and D30N HIV-1 protease. The D30N/L90M-nelfinavir complex structure, superposed on the wild type complex structure with RMSD of 0.607 Å for 198 Cα pairs while the corresponding value for D30N complex structure was 0.255 Å. Higher RMSD for wild type complex structure is because of difference in the crystal packing due to different space groups.

Hydrogen bonding interactions of nelfinavir in D30N/L90M HIV-1 protease and comparison with wild type complex structure

The hydrogen bonding interactions of nelfinavir in the active site of D30N/L90M HIV-1 protease are shown in figure 4.11. In the present structure also, all the four oxygens of nelfinavir are involved in hydrogen bonding interactions with protease as seen with the wild type and D30N HIV-1 protease. Central secondary hydroxyl group (O21) is at distances of 2.7-3.1 Å from all four oxygens of catalytic aspartates. Flap water molecule mediates hydrogen bonds between I50 and I1050 of protease and carbonyl oxygens of nelfinavir molecule, in a manner similar to wild type-nelfinavir complex structure. However, in the present structure, hydrogen bonds of flap water with nelfinavir are asymmetric. Hydrogen bond to P1 carbonyl oxygen O25 is stronger with a distance of 2.5 Å, while hydrogen bond with P1' carbonyl oxygen O17 is weaker with a distance of 3.4

Å (Figure 3.11). Corresponding distances in the wild type complex structure are 2.8 Å and 2.9 Å. Similarly, in the D30N HIV-1 protease-nelfinavir complex structure these distances are 2.6 Å and 2.9 Å. In the wild type-nelfinavir complex structure hydrogen bonds of flap water with the amide N atoms of I50 and I1050 are also symmetric with distances 2.9 Å in both the cases. In the present structure, these interactions are symmetric but weaker than the wild type with the distances of 3.2 Å and 3.1 Å respectively. In the D30N-nelfinavir complex, these distances were 3.1 Å and 3.0 Å respectively. Interestingly, the crucial hydrogen bond involving the 30^{th} side chain and the hydroxyl oxygen O38 of the benzamide group from nelfinavir is the most affected interaction. In the wild type HIV-1 protease, this hydrogen bond is a strong interaction with a distance of 2.9 Å. In the present structure, this distance is increased to 3.4 Å with the N1030 ND2/OD1 (Figure 4.12). In the case of D30N mutation, this distance was increased to 3.1 Å/3.2 Å. The N22 atom of nelfinavir forms a hydrogen bond to main chain carbonyl O of G1027 with a distance of 3.3 Å in the wild type HIV-1 protease. In the D30N HIV-1 protease, this distance is 3.4 Å. In the present structure, this hydrogen bond becomes even weaker with the distance increased to 3.6 Å. In the wild type, a water molecule mediates interaction of side chain OD2 and main chain N atoms of the residue D29 with the N12 atom of nelfinavir. These interactions are slightly stronger in the present structure. Same water molecule also mediates interaction of the N12 with main chain nitrogen of G27. This distance is increased to 4 Å in the present structure as compared to 3.6 Å in the wild type complex and 3.9 Å in the D30N HIV-1 protease-nelfinavir complex.

Melting temperature

For the D30N/L90M double mutant, T_m further decreases to 60 ºC as compared to 68 ºC observed for the wild type HIV-1 protease and 63 ºC observed for the D30N HIV-1 protease.

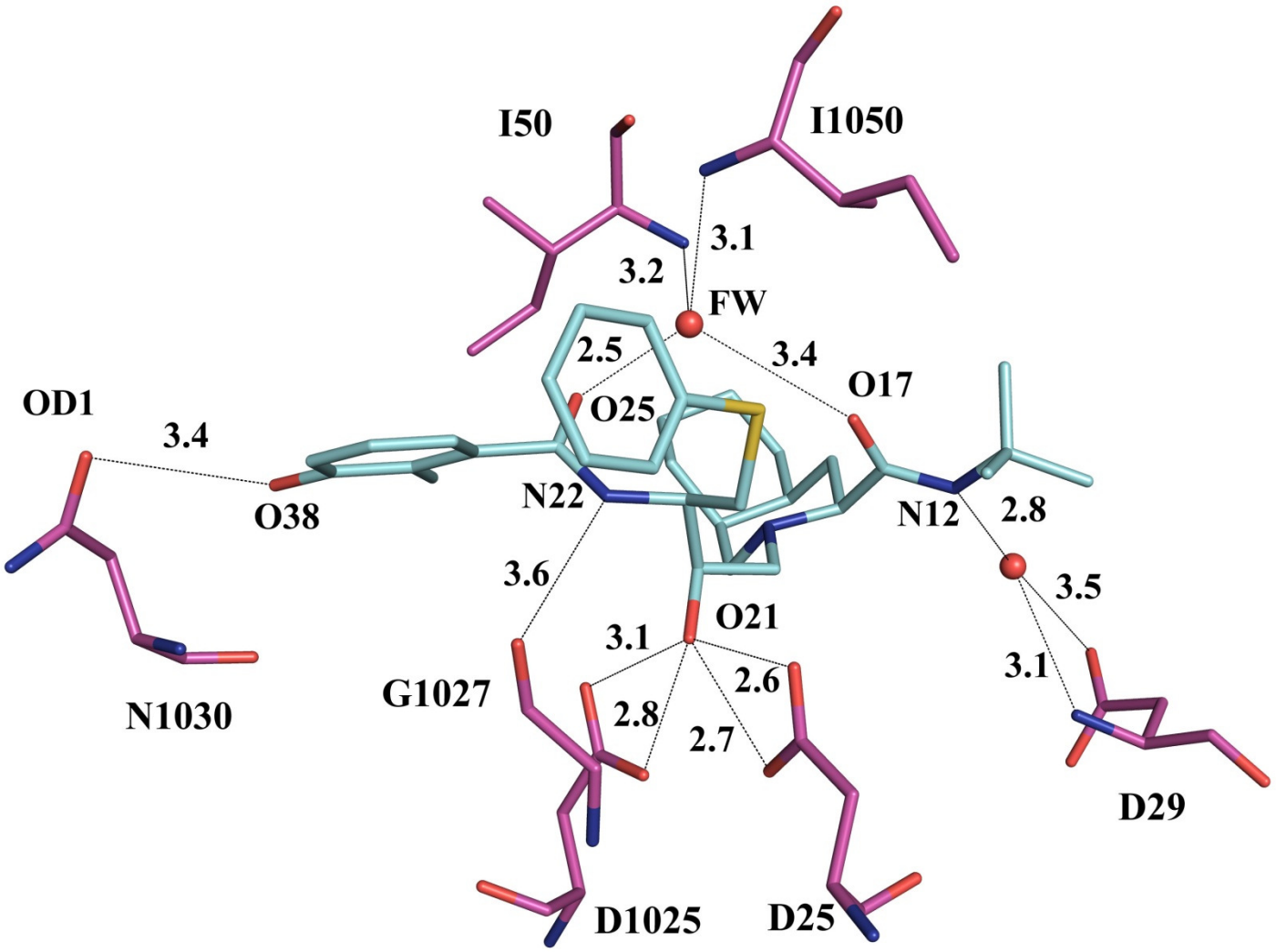

Figure 4.11: Hydrogen bonding interactions of nelfinavir in D30N/L90M HIV-1 protease. Drug is shown in cyan color while protein residues interacting with the drug are shown in magenta. Flap water is labeled as FW.

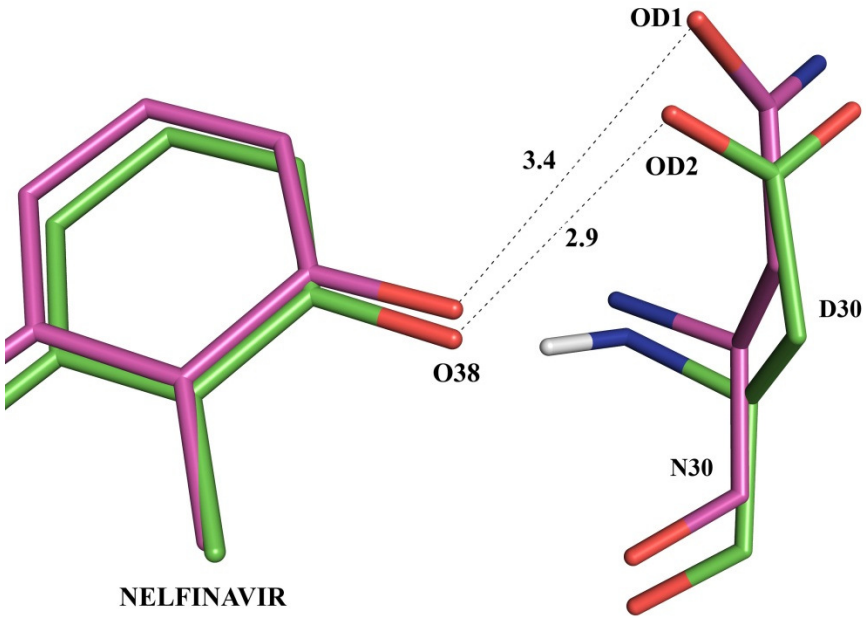

Figure 4.12: Weakening of hydrogen bond between nelfinavir and side chain of 30th residue in D30N/L90M HIV-1 protease (magenta) as compared to wild type (green). Only benzamide group of nelfinavir is shown.

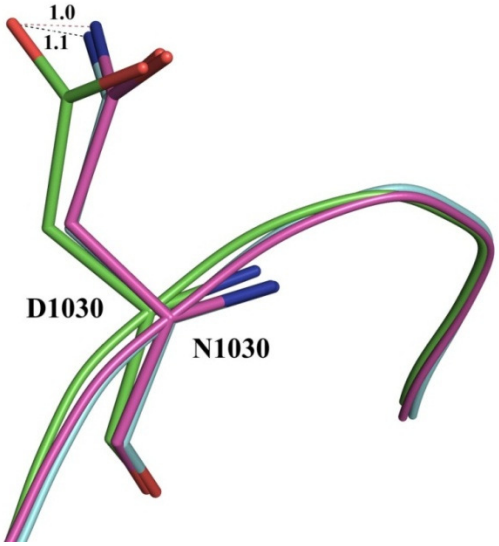

Figure 4.13: Deviations at residue 30 in D30N (cyan) and D30N/L90M HIV-1 protease (magenta) compared to wild type HIV-1 protease (green).

Discussion

D30N and reduced catalytic activity

Residue 30 is located in the active site and interacts with the substrates through direct hydrogen bonds. It was observed in D30N HIV-1 protease structure that mutation D30N, changes the charged aspartate at 30^{th} residue to a neutral asparagine without large structural changes. However, the active site loop and catalytic aspartates are displaced by amount in the range of 0.1 Å-0.4 Å. The side chain atoms of 30N show larger displacements from their positions in the wild type enzyme (Figure 4.13). These deviations were suggested to affect the catalytic efficiency of the protease primarily due to the weakening of direct hydrogen bond between side chain of 30^{th} residue and different substrates. In the D30N/L90M unliganded structure, the orientation and position of N30 is similar to the D30N unliganded structure (Figure 4.13). Therefore, the effect of D30N mutation on catalytic efficiency of the enzyme in double mutant will be similar to that of the D30N single mutant.

L90M and reduced catalytic activity

L90 is located in the middle of the only α-helix of HIV-1 protease situated just beneath the active site loop containing catalytic aspartates, and does not show any large structural effect when substituted by M. However, methionine is bigger than leucine and has an extra Cε atom, which allows it to make additional interactions with surrounding residues. The bigger methionine residue at 90^{th}/1090^{th} position in D30N/L90M is accommodated by the outward movement of 90^{th}/1090^{th} main chain and slight adjustments of catalytic loop (Figure 4.14). L90M mutation causes the gap below the catalytic loop to vanish and that should affect the flexibility of active site loop and catalytic aspartates. This loss of

flexibility may hamper any movement required to position properly the different substrates and should reduce catalytic activity of the enzyme.

Despite these movements in the structure of mutant protease to accommodate M90/1090, Sδ and Cε atoms of M1090, make interactions which were not present in the wild type protease. In wild type HIV-1 protease (PDB ID1LV1), side chain atoms of L1090 make 12 interactions shorter than 4 Å. This number is increased to 19 in the present structure primarily due to additional interactions of Cε atom of M1090. The Sδ atom of methionine introduces some new interactions in HIV-1 protease (Figure 4.15). It interacts with the main chain amide N of catalytic aspartate and carbonyl O of I1085. Cε atom makes new C-H…O interactions with main chain carbonyl O of D1025 and I1085 (Figure 4.16 and 4.17). Cε atom of M1090 has Van der Waals interactions within the α-helix with G1086 carbonyl C and R1087Cα (Figure 4.16). It also has Van der Waals interactions with carbonyl C of D1025, Cα and Cβ of T1026 in the active site loop (Figure 4.17). Catalytic activity of HIV-1 protease is dependent on the pKa and charge on the catalytic aspartates. pKa of amino acid residues are strongly affected by their environment. It is possible that these new interactions of 1090M with the catalytic loop alter the subtle balance of pKa and charge distribution on the two catalytic aspartates and therefore affect catalytic activity.

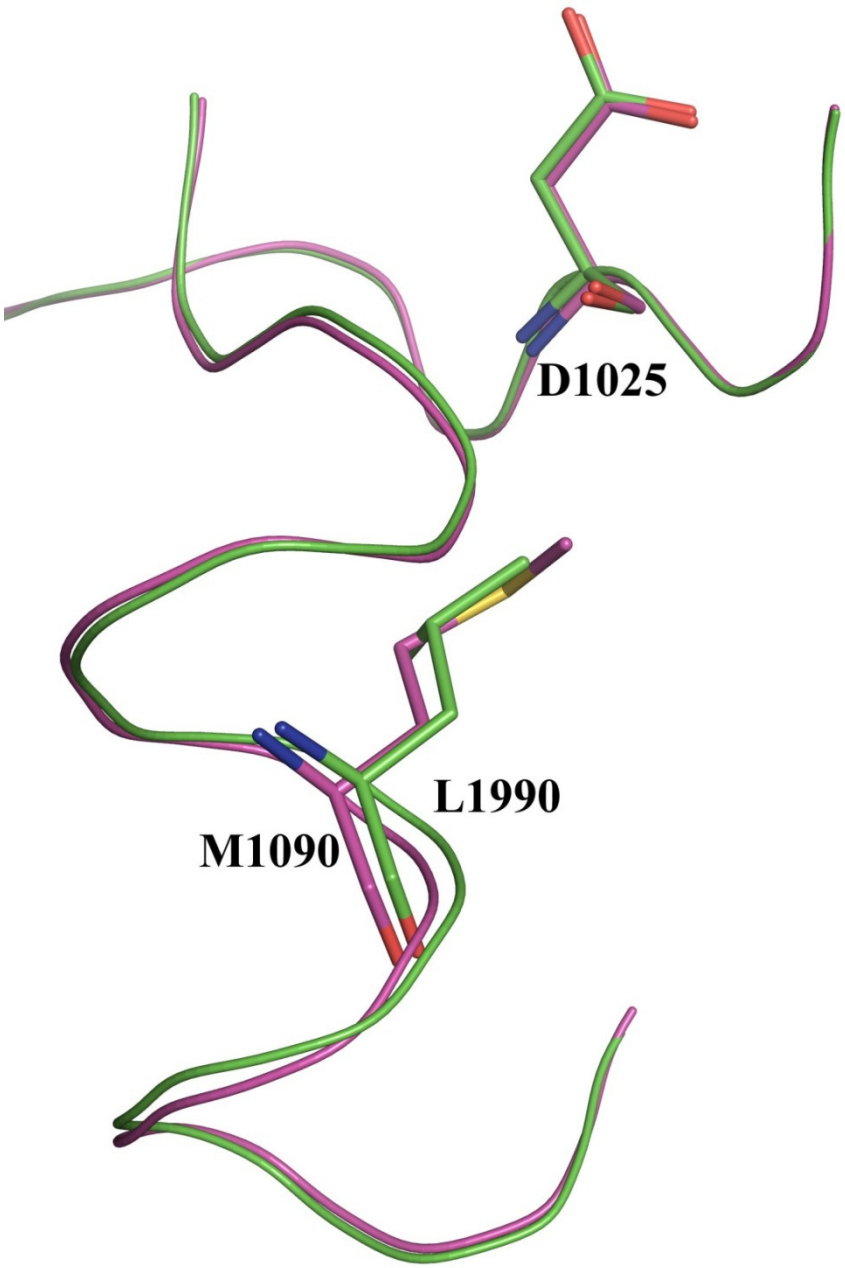

Figure 4.14: Accommodation of larger methionine residue by structural displacements: structural superposition of D30N/L90M HIV-1 protease (magenta) with wild type protease (green).

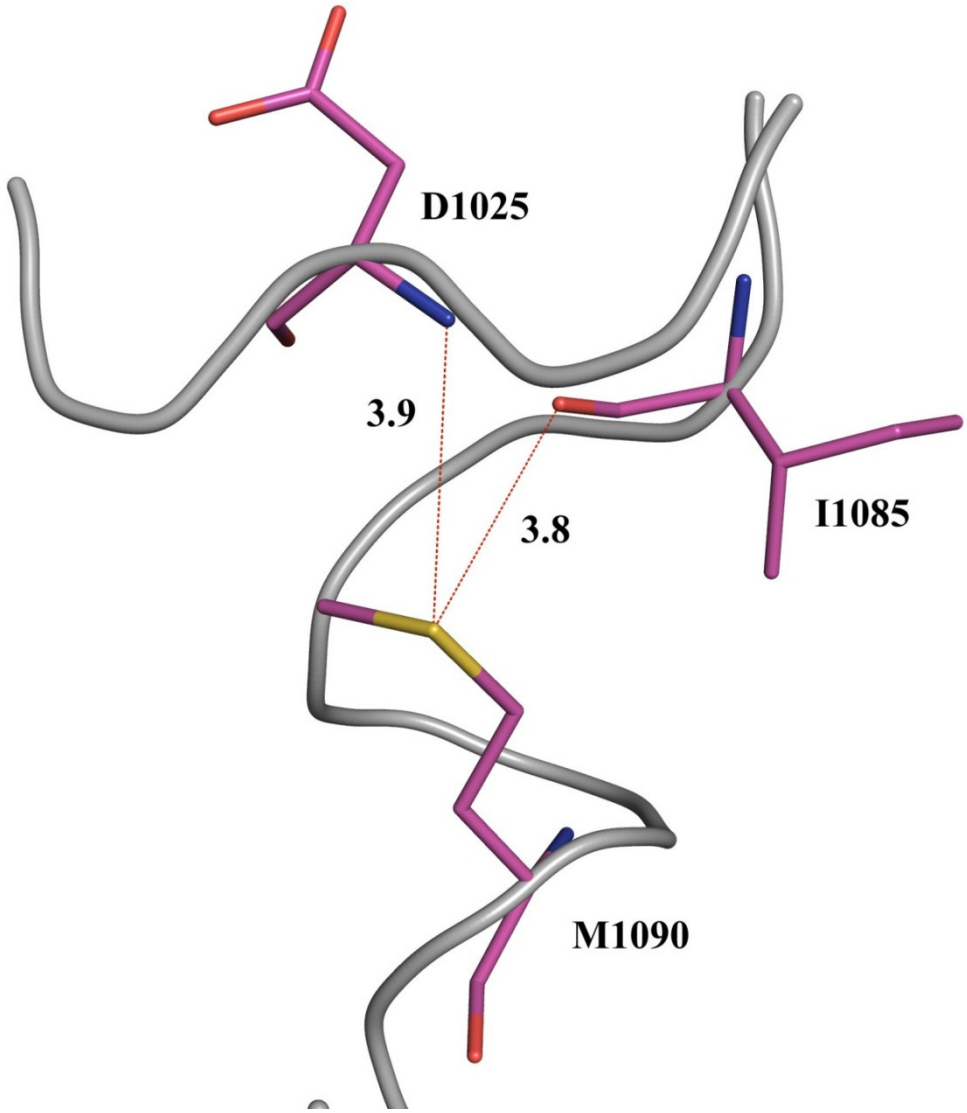

Figure 4.15: Interactions of Sδ atom of M1090.

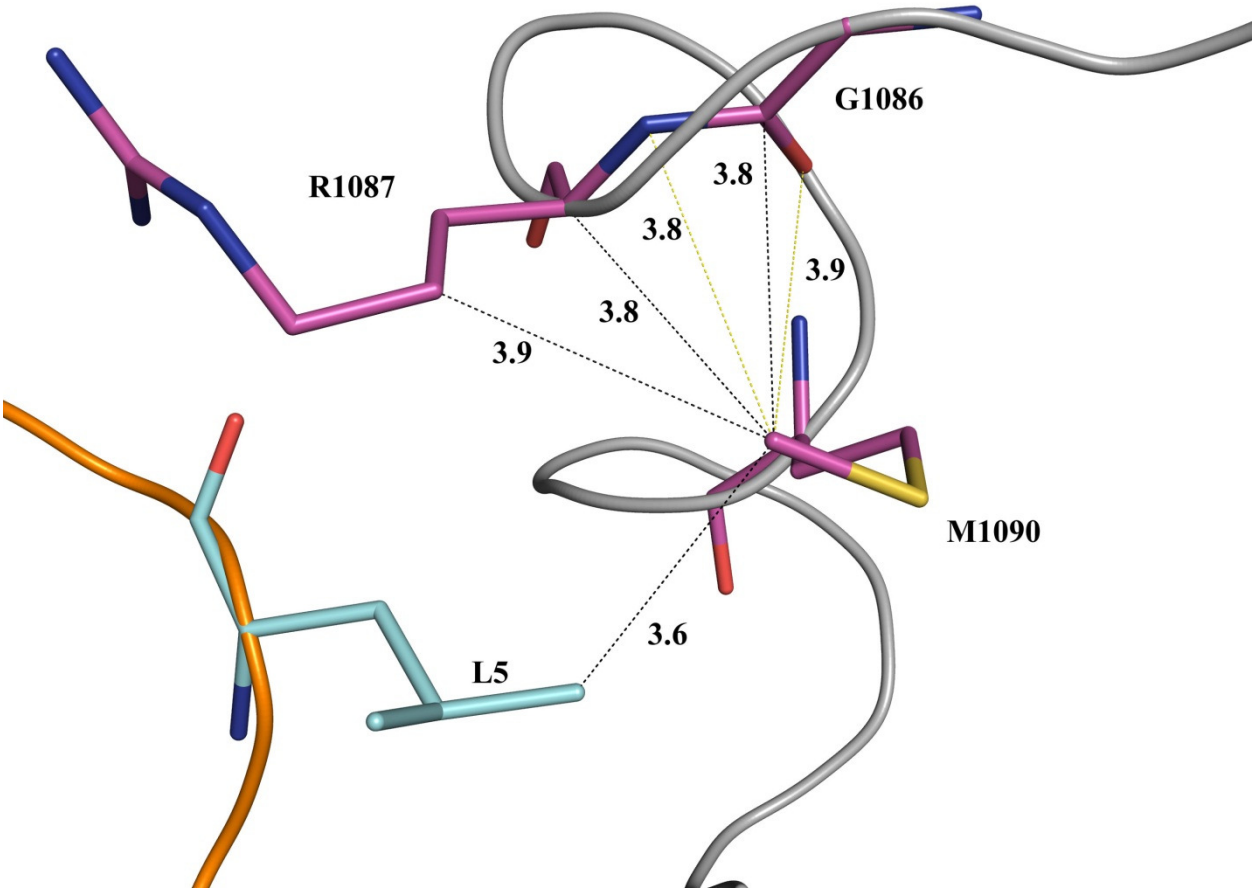

Figure 4.16: Interactions of Cε atom of M1090 with residues in the helix and L5 of other monomer. C-H…O and C-H…N interactions are shown by yellow dots while van der Waals interactions are represented by black dots.

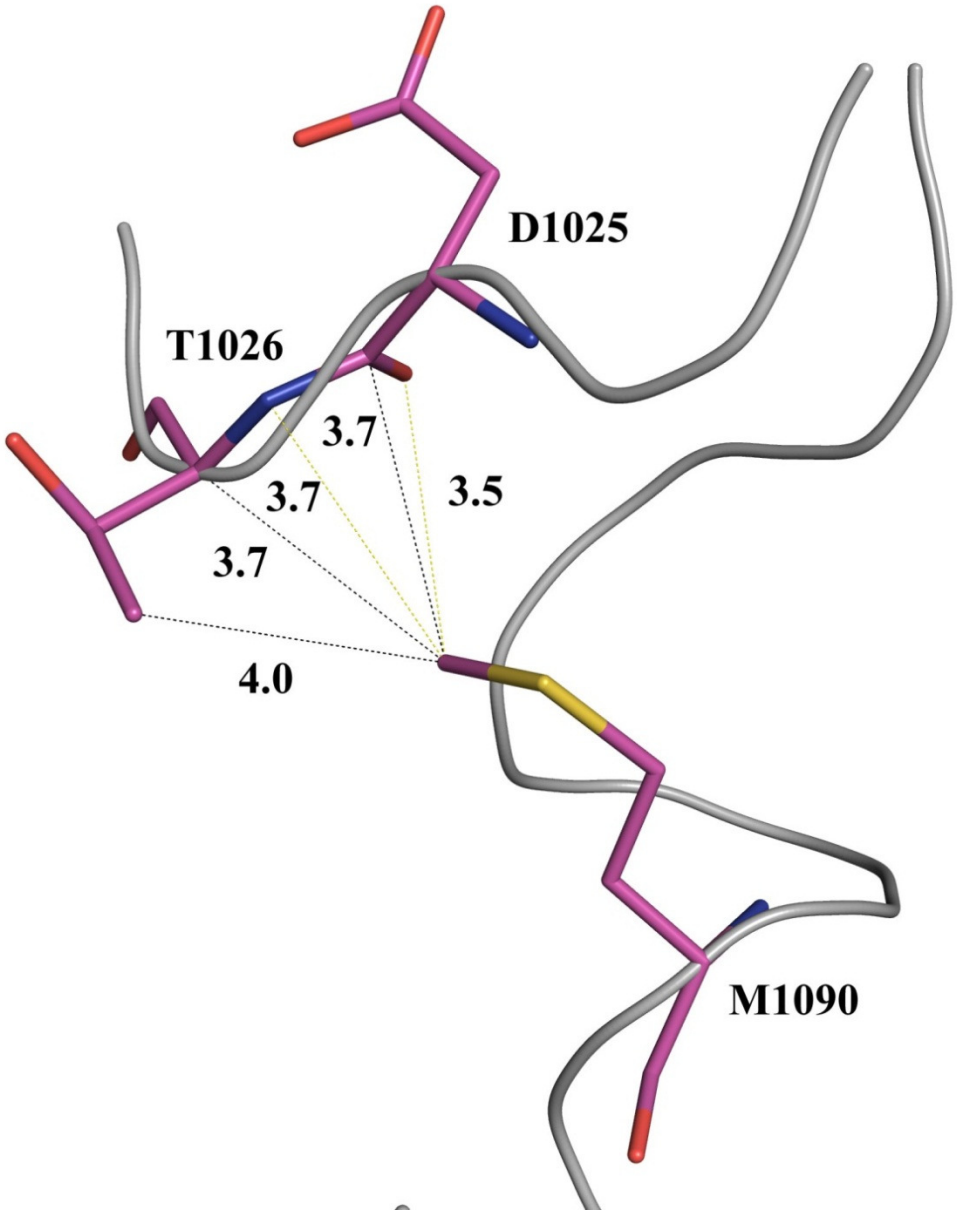

Figure 4.17: Interactions of Cε atom of M1090 with residues in the catalytic loop. C-H…O and C-H…N interactions are shown by yellow dots while van der Waals interactions are represented by black dots.

Effect of D30N and L90M on the folding and monomer stability of the enzyme

T_m values of the wild type, D30N, and D30N/L90M double mutant were determined and it was found that both mutant proteases are less stable than the wild type protease. Mahalingam et al [1999, 2004] also have shown that L90M mutation destabilizes the protein. However, contrary to our observation, Mahalingam et al [1999] found no effect on the stability of the D30N enzyme. Lower T_m values in mutant HIV-1 protease are generally attributed to the lower dimer stability [Mahalingam et al 1999]. However, D30N and D30N/L90M have no structural effect that can be rationalized in the form of reduced dimer stability. In fact, analysis of D30N HIV-1 protease structure suggested that dimer is getting stronger. Gap volume index (which is lower for better packing) calculated using ProtorP server decreases further to 0.66 in D30N/L90M HIV-1 protease from 0.74 in wild type protease and 0.70 in D30N protease [Reynolds et al 2009]. Number of inter sub-unit-interactions (below 4 Å) between the two monomers in D30N/L90M protease is 339, which is slightly higher than 325 observed in wild type protease, though it is less than 383 observed in D30N HIV-1 protease. For the D30N HIV-1 protease, it was suggested that destabilization of core and impairment of folding can be a reason for low T_m in the mutant protease. D30 is the part of structurally conserved loop (24-34, known as S2 region) and, L90 is located in the α-helix (87-93, known as S8 region), which is also one of the most conserved regions of HIV-1 protease. Both these regions have been shown to be part of HIV-1 protease folding core [Wallqvist 1998] and were suggested to form local elementary structures (LES). These are also shown to be the most stable region during unfolding of the HIV-1 protease [Wallqvist 1998, Levy et al 2004]. These LESs then interact with one another to form a stable

folding nucleus (FN) which leads to native folded HIV-1 protease [Broglia et al 2008, Verkhivker et al 2008, Bonomi et al 2010]. Interactions with S2 and S8 regions have been shown to be important for forming FN [Broglia et al 2008, Verkhivker et al 2008, Bonomi et al 2010]. In fact, a peptide based on sequence of S8 region has been shown to inhibit HIV-1 protease folding by binding to S2 region [Broglia et al 2008, Verkhivker et al 2008, Bonomi et al 2010]. Therefore, it is suggested that the mutations, D30N in S2 region and L90M in S8 region might disrupt the interaction between these regions during folding and lead to impaired folding. Both these mutations might destabilize their respective LES and thus lead to unfolding at lower temperature. Further, L90 in wild type HIV-1 protease is completely buried in hydrophobic environment. L to M mutation in such environment in the protein is estimated to cause a loss of stability of 1.4 Kcal/mol [Gassner et al 1996, Lipscomb et al 1998]. This loss of stability is due to the greater entropic penalty imposed for burial of an extra bond in methionine as compared to leucine. Therefore, L to M substitution in HIV-1 protease is likely to reduce the monomer stability.

L90M mutation also introduces several long-range interactions with residues 86 and 87, which are known to be crucial for the folding and stability of HIV-1 protease. Further, in wild type HIV-1 protease, 95 is Cys, S…S long range interactions might also play some role in destabilizing HIV-1 protease monomer.

Mutual exclusivity of D30N and L90M

Above discussion shows that both D30N and L90M, individually reduce the catalytic activity of the enzyme. Active site volume in D30N/L90M is reduced to 960 $Å^3$ from 996.6 $Å^3$ in wild type HIV-1 protease. This reduced volume may hamper the proper

binding of substrates and lead to reduced activity of the enzyme. The combined result of these deleterious effects will be a drastic reduction in the catalytic activity of the enzyme, which may not be able to support the growth of the virus. Both the mutations are likely to destabilize their respective LES formed during folding of the enzyme and as discussed earlier, the combination of these two mutations is likely to result into a folding impaired HIV-1 protease. Therefore, it is suggested that because of all these multiple effects, these two mutations have not been observed in drug resistant patient despite having selective advantage of high level of drug-resistance.

Nelfinavir resistance by D30N/L90M HIV-1 protease

The crucial hydrogen bond of O38 atom with charged donor, OD2 atom of D30 residue (2.9 Å) was changed to a weaker hydrogen bond in D30N (3.1 Å/3.2 Å) as the donor was changed to a neutral ND2 atom or carbonyl O. Both, the change in the nature of hydrogen bond and increase in distance were suggested to weaken the strength of this hydrogen bond and lead to reduced binding affinity of nelfinavir in D30N HIV-1 protease. In D30N/L90M HIV-1 protease, this hydrogen bond is further weakened as the distance is increased to 3.4 Å. Active site volume of the D30N/L90M unliganded protease is decreased as compared to the wild type protease due to the upward movement of catalytic loop. Presence of bigger methionine below the catalytic loop may impose rigidity in the loop and restrict proper movement during the binding of nelfinavir. This resulted into the asymmetric movement of the flap residues to accommodate the drug. Interactions of the flap water in the present structure are weaker as compared to wild type protease-nelfinavir complex. In addition, hydrogen bond of N22 with G1027 O is weakened. These changes in the interactions will lead to suboptimal binding. This suboptimal

binding is reflected in reduced complementarity between nelfinavir and the active site in the present structure. Overall normalized complementarity of nelfinavir calculated using LPC software [Sobolev et al 1999] is lower (0.74) in the D30N/L90M HIV-1 protease than that of the wild type protease (0.79). Binding of nelfinavir in the active site of HIV-1 protease is entropically driven. In such case, complementarity of the ligand to the binding site is very important. Any decrease, will reduce the effectiveness of the inhibitor, as it will not be able to compete with the substrates.

Conclusion

The D30N/L90M double mutant structures in unliganded form and in complex with the drug nelfinavir were determined to atomic resolutions. D30N and L90M both results into structural effects that lead to drastic reduction of catalytic activity of D30N/L90M double mutant HIV-1 protease. In addition, it is suggested that D30N/L90M mutation might destabilize folding core and impair the folding process. Combined effect of these two factors is suggested to make this combination lethal for HIV and is likely to be the reason of mutual exclusivity of D30N and L90M substitutions. Crucial hydrogen bond of nelfinavir with side chain of 30^{th} residue is further weakened in D30N/L90M HIV-1 protease-nelfinavir complex. In addition, presence of methionine at 90^{th} position limits the movement of catalytic loop leading to suboptimal binding of nelfinavir. Both of these effects will reduce the binding affinity of nelfinavir.

Chapter – 4C

Resistance Mechanism Revealed By Crystal Structures Of Unliganded Nelfinavir-Resistant HIV-1 Protease Non-Active Site Mutants N88D and N88S.

Two different nelfinavir resistant mutations were observed at position 88, one where asparagine is changed to aspartate (N88D) and other where asparagine is replaced by serine (N88S) [Patick et al 1998, Shafer et al 1999, Pellegrin et al 2002]. N88D alone is shown to be lethal to the virus, and reduces the catalytic efficiency of HIV-1 protease by 20-40% [Mahalingam et al 1999, Parera et al 2007]. However, when combined with the D30N mutation, N88D acts as a compensatory mutation for D30N and partially restores the cleavage efficiency loss by D30N [Parera et al 2007]. N88D in combination with D30N increases nelfinavir resistance by several folds [Kozisek et al 2007]. N88D is strongly correlated with D30N as it usually appears with D30N. N88S on the other hand, is negatively correlated with D30N [Mitsuya et al 2006]. N88S alone is capable of giving resistance to nelfinavir and atazanavir [Shafer et al 1999, Gong et al 2000, Shafer, 2002]. Other interesting aspect of N88S mutation is that it makes the HIV hyper susceptible to the drug amprenavir [Ziermann et al 2000]. N88S mutation also results in reduced cleavage efficiency [Resch et al 2002]. Unliganded structures of non-active site mutant proteases can help to understand fully the effect of these non-active site drug resistant mutations on the structure of HIV-1 protease. So far very few crystal structures of unliganded mutant HIV-1 protease have been published [Liu et al 2006]. This sub-chapter describes the crystal structures of unliganded N88D and N88S mutant HIV-1 proteases refined to resolutions of 1.65 Å and 1.8Å respectively. Interestingly, in both the structures, the residue most affected by mutation is not 88, but it is D30. There is no major structural change in N88D mutant protease but there are subtle changes in hydrogen bonding network connecting 30^{th} and 88^{th} residues in mutant protease, and these changes might be involved in reduced catalytic efficiency of N88D HIV-1 protease.

N88S mutation on the other hand has caused large structural change in the conformation of D30. The oxygen atom D30 OD2 has moved by about 2.8 Å to form a direct hydrogen bond with S88 OH. This movement of D30 will eliminate the direct hydrogen bond between nelfinavir hydroxyl O of 2-methyl-3-hydroxy benzamide group and D30 carboxyl O observed in the wild type-nelfinavir complex.

Materials and Methods

Site-directed mutagenesis and purification of HIV-1 protease

The tethered dimer of HIV-1 protease as described in chapter 2 was used in the present study. We have labeled two subunits of the tethered dimer as A and B. Residues are numbered from 1 to 99, in subunit A and from 1001 to 1099, in subunit B. Mutation N88D and N88S were introduced in HIV-1 protease construct by Quick-change multi-site directed mutagenesis method [Stratagene, La Jolla, CA]. Oligonucleotide primers were designed using Stratagene's primer design software and sequences of primers used for the mutagenesis are as follows:

N88D - 5' – CTGAGTCAGCAGGT*C*ACGGCCGATAATGTTAACCGG – 3'

N88S – 5' – TTAACATTATCGGCCGTA*G*CCTGCTGACTCAGATC – 3'

Incorporation of the mutations in the HIV-1 protease gene was confirmed by DNA sequencing. Expression and purification protocol for mutant protease were similar to that described for wild type HIV-1 protease. Purified mutant HIV-1 protease were concentrated using centrifugal concentrator, to 2-3 mg/ml and then exchanged to 50mM sodium acetate buffer (pH 4.5) containing 1mM DTT. CyBio-HTPC crystallization robot was used for initial screening of crystallization conditions. Initial hits were optimized manually using hanging-drop vapor diffusion method in 24-well crystallization plates.

Reservoir solution contained 1-6% saturated ammonium sulfate as precipitant and 200/100mM sodium dihydrogen phosphate/sodium citrate buffer at pH 6.2. Hexagonal rod shaped crystals appeared in 2-3 days and grew to a size of 0.4mm x 0.07mm x 0.07mm.

Diffraction data collection

X-ray diffraction data were collected by the oscillation method, under cryo-conditions, on the ID29 beamline at ESRF. Before exposing to X-rays, crystals were first equilibrated in cryoprotectant solution (25% glycerol and 75% reservoir solution) for a few seconds. 60 and 90 frames were collected for N88D and N88S respectively, each frame for an oscillation angle of 1 degree and exposure time of 1 second. The oscillation frames were indexed, integrated and scaled with program suite XDS [Kabsch 2010].

Structure refinement

Crystal structures of mutant proteases were solved by difference-Fourier method, and the protein model of the unliganded tethered HIV-1 protease dimer (PDB ID 1LV1) was directly used for phasing. The molecular model was subjected first to rigid body and later to individual atom refinement using standard simulated annealing refinement protocol, as implemented in the crystallography software suite PHENIX [Adams et al 2010]. Water molecules were picked automatically using the program PHENIX [Adams et al 2010]. These water positions were then edited by visual inspection of both 2mFo-DFc and mFo-DFc electron density maps. During later stages of refinement, only energy minimization without simulated annealing was employed. Translation/Libration/Screw refinement was performed during final refinement cycles [Painter and Merritt 2006a, b]. Likelihood weighted electron density maps were used in interpreting the molecular structures using

the computer graphics programs O [Jones et al 1991] and Coot [Emsley and Cowtan 2004, Emsley et al 2010]. Molecular superpositions in three dimensions were carried out using software O [Jones et al 1991] and CCP4 suite. All figures have been prepared using the software PyMOL [Schrodinger, LLC]. The accessible surface areas and volumes of cavities were calculated using the CASTp server with a probe radius of 1.4 Å [Binkowski et al 2003].

Results

Crystal structures

The mutants N88D and N88S of tethered HIV-1 protease crystallized in the $P6_1$ space group, and the crystal contained one tethered dimer in the asymmetric unit. The crystals diffracted to 1.65 Å and 1.8 Å resolution respectively. N88D structure was refined to R_{cryst} = 0.1859 and R_{free} = 0.2076 whereas N88S structure was refined to R_{cryst} = 0.1781 and R_{free} = 0.2252. In the interpretation of electron density, alternate conformations were modeled for a total of 14 amino acids. In N88D protease, these residues are 7, 66, 67, 68, 69, 82, 84, 92, 97 in subunit A and 1007, 1032, 1084, 1092 and 1097 in subunit B. In N88S protease structure these residues are 8, 61, 64, 82, 89, 92 in subunit A and 1003, 1007, 1012, 1032, 1082, 1087, 1092 and 1097 in subunit B. Data collection and refinement statistics are given in table 4.4. Following discussion is applicable to both subunits even though subunit B residue numbers are used in the figures.

Structural changes in N88D protease

To understand the structural effects of the mutation N88D (Figure 4.18), we compared the structure of unliganded N88D HIV-1 protease with wild type closed flap unliganded HIV-1 protease structure (PDB ID 1LV1). The overall conformation of the mutant

protein N88D protease was very similar to that of the native unliganded structure (PDB ID 1LV1) with RMSD of 0.339 Å for 198 Cα atom pairs. Figure 4.19 shows structural comparison around the mutation site. While the deviation of Cα atom of 1088/88 is 0.2 Å/0.3 Å, the largest deviation is of side chain oxygen atom at 1.4 Å/1.3 Å. In wild type HIV-1 protease (PDB ID 1LV1), there is an attractive interaction between ND2 of N1088/88 and carbonyl oxygen of T1074/74 (Figure 4.19). When ND2 is changed to a partially charged oxygen atom in N88D protease, this interaction is changed to a repulsive interaction, and to minimize this repulsion D1088/88 side chain orientation is changed to increase the separation between T1074/74 O and D1088/88 OD2. Interestingly, a large change is also observed in the position of the OD1 atom of residue D1030/30 (Figure 4.19). As a result of these changes the distance between closest oxygen atoms from residues 1030/30 and 1088/88 is reduced from 5.2 Å/5.7 Å in wild type protease to 4.5 Å/4.5 Å in N88D protease.

Table 4.4: Data collection and refinement statistics for unliganded N88D and N88S HIV-1 protease structures.

	N88D	N88S
Wavelength of radiation used	0.976180	0.976180
Resolution range used in refinement (Å)	26.8-1.65	32.8-1.8
Space group	$P6_1$	$P6_1$
Unit cell dimensions (Å)	a=b=62.05, c= 83.7	a=b=62.17, c= 83.0
Molecules in asymmetric unit	One tethered dimer	One tethered dimer
Total number of reflections measured	80933	91442
Total number of unique reflections	21868	16806
Rmerge-F (%)	10.4 (60.5)	8.7 (46.2)
Completeness (%)	99.4 (99)	99.4 (97.7)
$\langle I/\sigma (I)\rangle$	14.15 (2.82)	17.05 (3.62)
R_{cryst} (%)	0.1859	0.1780
R_{free} (%)	0.2076	0.2247
Number of water molecules	290	240
RMSD from ideal values		
Bond lengths (Å)	0.006	0.007
Bond angles (°)	1.058	1.118

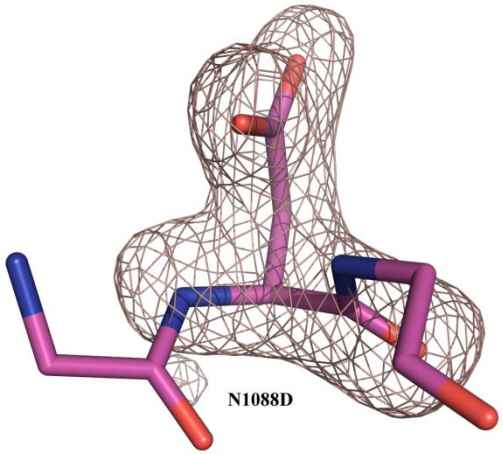

Figure 4.18: Simulated annealed omit map around residue D1088 in N88D HIV-1 protease contoured at 3σ level.

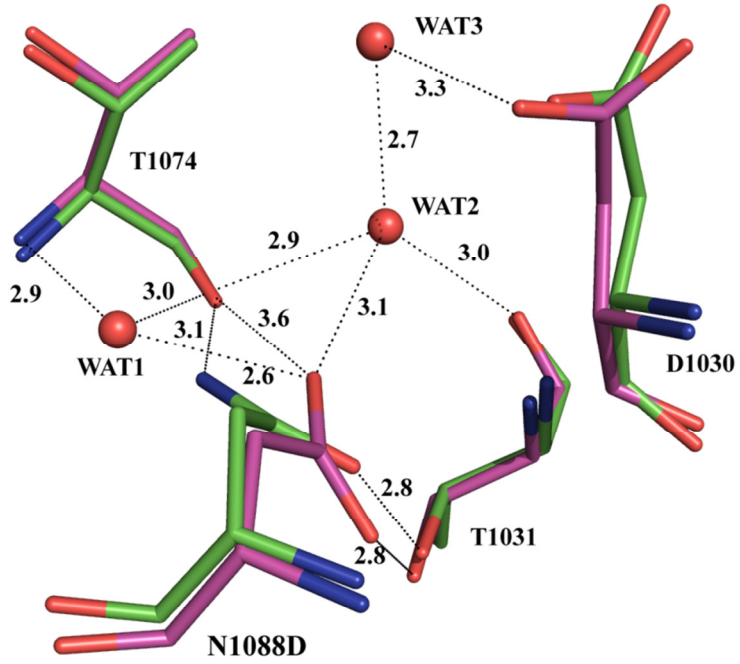

Figure 4.19: Hydrogen bonding around residue 88 in wild type [PDB ID 1LV1] and N88D HIV-1 protease (green 1LV1 and magenta N88D).

Changes in hydrogen bonding interactions in N88D

The mutation N88D has resulted in altered hydrogen bonding pattern around residue 1088/88. Interactions involving main chain atoms from residue 1088/88 are similar in both structures. Among interactions involving side-chain atoms hydrogen bond between N1088/88 OD1 and T1031/31 OG1, present in wild type protease is maintained by D1088/88 OD1 in N88D protease (Figure 4.19). In wild type HIV-1 protease, N1088/88 also makes a hydrogen bond with T1074/74 O through ND2. This hydrogen bond is not formed in N88D. Due to change in side-chain orientation of D1088/88, distance between D1088/88 OD2 and T1074/74 O is increased to 3.6 Å/3.7 Å and this configuration is stabilized by water molecules (WAT 1 and WAT 2) which hydrogen bond to both residues as shown in figure 4.19.

Structural changes in N88S protease

To understand the structural effect of N88S mutation (Figure 4.20), structure of N88S protease was compared with the structure of wild type protease (PDB ID 1LV1). The two structures superposed with RMSD of 0.337Å for 198 Cα atom pairs. The side-chain conformation of residue 1088/88 has not shown any change, with the S1088/88 OG overlapping onto the CG of N1088/88 (Figure 4.21). Interestingly, there is significant change in the position of D30/D1030. Cα of D1030/30 is shifted by 0.7 Å/0.7 Å, while the side chain of D1030/30 has moved much more, with the shift in the OD1 is as much as 2.3 Å/ 2.8 Å. This shift brings D30/D1030 OD2 to be within hydrogen bonding distance from S1088/88 OG1 (Figure 4.21). This shift of D1030/30 also increases the active site volume of N88S protease from 996.6 $Å^3$ in 1LV1 to 1091.8 $Å^3$.

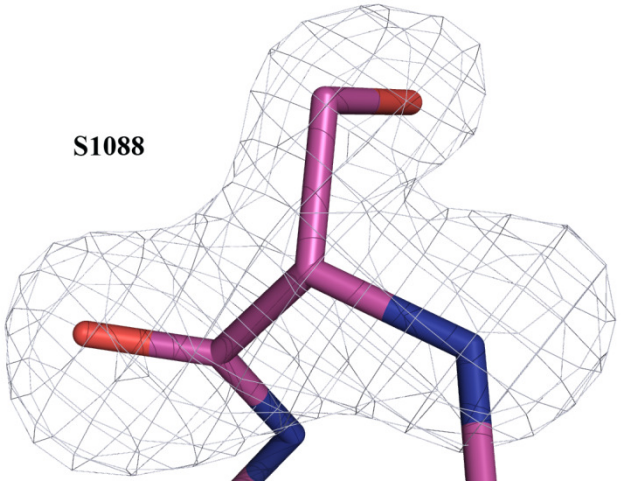

Figure 4.20: Simulated annealed omit map around residue S1088 in N88S HIV-1 protease contoured at 3 σ level.

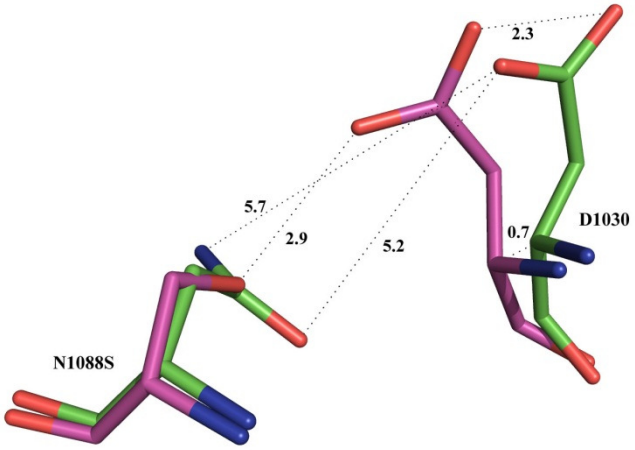

Figure 4.21: Hydrogen bond between D30 and 88S in N88S HIV-1 protease: structural superposition of present structure (magenta) with wild type HIV-1 protease (green) [PDB ID 1LV1]. Deviations at Cα and side chain atoms between 30th residues of the two structures are also shown.

Changes in hydrogen bonding interactions of N88S protease

Hydrogen bonding interactions of N88S protease are also altered (Figure 4.22). In wild type HIV-1 protease, N1088/88 is involved in hydrogen bonding interactions through both its main chain and side chain atoms. Main chain interactions are almost similar in both structures but due to the smaller size, S1088/88 cannot form the same hydrogen bonds that are made by side chain atoms of N1088/88. The direct hydrogen bonds with T1031/31 OG1 and T1074/74 O are lost. Since S occupies approximately 33 $Å^3$ less volume than N, N88S mutation creates a small cavity [Hebert et al 1998]. In the present structure, this cavity is filled by a well-ordered water molecule (WAT 4 in figure 4.22). This water connects S1088/88 to T1031/31 OG1 and T1074/74 main chain O (Figure 4.22) thereby fulfilling the role of side chain N atom in the native molecule. So, in N88S protease direct hydrogen bonds from the side chain of residue 1088/88 with residues 1031/31 and 1074/74 are replaced by water-mediated hydrogen bonds. In 1LV1 N1088/88 does not interact with D1030/30 while in N88S protease, S1088/88 forms a direct hydrogen bond with D1030/30 OD2 (Figure 4.22).

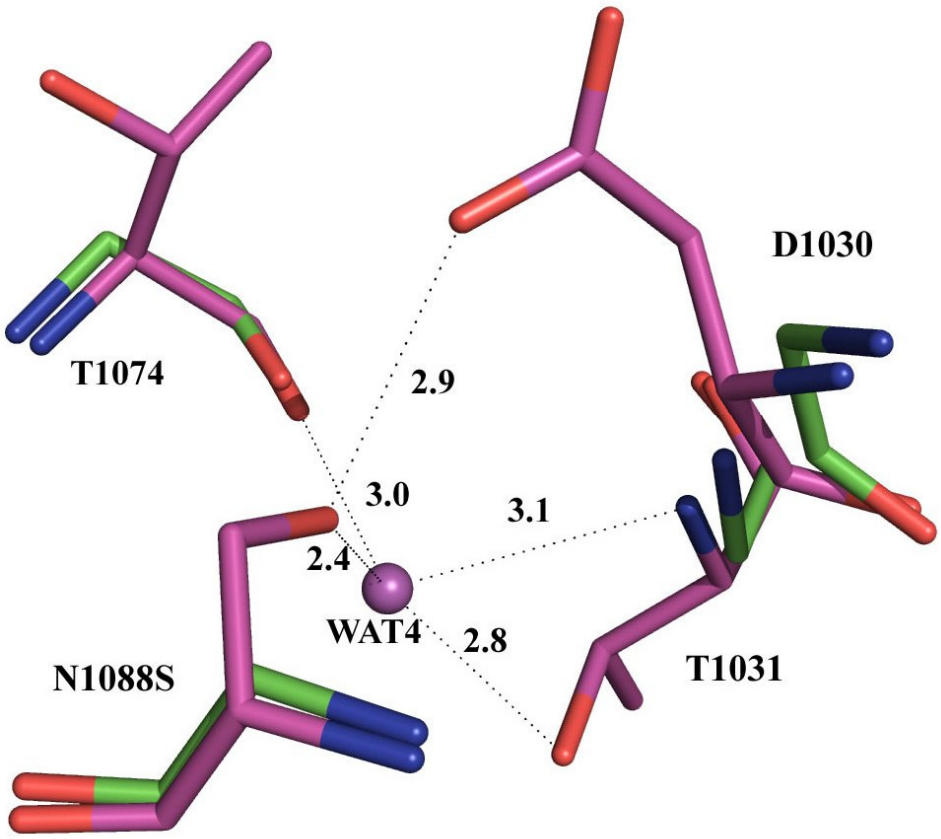

Figure 4.22: Hydrogen bonding interactions of S88 in N88S: structural superposition of present structure (magenta) with wild type HIV-1 protease (green) [PDB ID 1LV1]. For clarity, only main chain atoms of wild type structures are shown.

Discussion

Coupling between residues 88 and 30

Three-dimensional structure of HIV-1 protease is very similar to that of other retroviral proteases. Some of the critical residues are fairly conserved throughout the retroviral family. The C-terminal helix region in HIV-1 protease containing G86-R87-D/N88 residues is unique and well conserved among retroviral proteases [Ishima et al 2001, Bagossi et al 2005]. However, there are some variations at residue 88 in different retroviral proteases. In HIV and SIV proteases 88th position is N whereas in most other retroviral proteases this position is occupied by D [Bagossi et al 2005]. Similarly, only HIV and SIV proteases have D at 30 whereas most other proteases contain either I or L at position 30. In FIV protease, mutation I35D (equivalent to 30th residue in HIV-1 protease) results into inactive enzyme [Lin et al 2003]. Addition of D105N (equivalent to 88 in HIV-1 protease) to I35D FIV protease makes the double mutant FIV protease capable of cleaving HIV-1 protease substrates whereas FIV protease activity is only marginally restored. These observations suggest that residues 30 and 88 function in tandem and N at position 88 with D at 30 is involved in determining substrate specificity and affect catalytic activity. Interestingly, in the two structures presented here, the mutation at position 88 is significantly affecting the position of residue 30 indicating a structural coupling between the two residues.

Reduced cleavage efficiency of N88D protease and N88S protease

N88D mutation alone significantly reduces cleavage efficiency of HIV-1 protease [Mahalingam et al 1999, Parera et al 2007]. In N88D HIV-1 protease, the active site volume is increased from 996.6 Å^3 in 1LV1 to 1118.6 Å^3. The increased size of the active

site may result in sub-optimal binding of substrate leading to higher K_m values. In N88D protease, the presence of an extra negative charge in the vicinity of the active site may also be inhibiting substrate binding and catalysis.

Structural comparisons have shown that catalytic aspartates form one of the most rigid regions in HIV-1 protease [Zoete et al 2002, Kumar and Hosur, 2003] suggesting that any deviation in the position of the aspartates may affect enzyme activity. For example, in the case of L90M HIV-1 protease, impaired activity is attributed to deviation of aspartates to the tune of 0.17-0.19 Å [Kovalevsky et al 2006]. In the present N88D structure, deviations in the position of the aspartates are of the order of 0.28 Å-0.3 Å, and this could therefore contribute to reduced activity (Figure 4.23).

N88S protease was reported to reduce cleavage efficiency of HIV-1 protease and process GAG polyprotein aberrantly [Resch et al 2002]. Like in N88D, in N88S the volume of the active site has increased, though not to the same extent. This increased volume may result in sub-optimal binding thereby affecting cleavage efficiency. In N88S protease structure, we have observed a large shift in position of D1030/30 which moves its side-chain away from the active site. Residue 1030/30 is directly involved in binding of several substrates [Prabhu-Jeyabalan et al 2002]. Change in the position of D1030/30 may affect these interactions thereby affecting the catalytic efficiency of N88S protease.

Nelfinavir resistance by N88S

Experimental structure of N88S HIV-1 protease/nelfinavir complex is not yet available so we modeled the position of nelfinavir in the active site of N88S HIV-1 protease by superposition of wild type HIV-1 protease-nelfinavir complex with the present N88S HIV-1 protease structure. The two structures superposed with overall RMSD of 0.786 Å

for 198 Cα atom pairs. Figure 4.24 shows the position of nelfinavir ring obtained from such superposition. In N88S protease, D30 is moved away from the active site towards S88 to make a hydrogen bond with the later. Due to change in position of D30, distance between nelfinavir hydroxyl O of benzamide group and D30 OD2 is increased from 2.9 Å in wild type-nelfinavir complex to 5.3 Å in N88S protease-modeled nelfinavir complex (Figure 4.24) which is too large to form a direct hydrogen bond. Instead, nelfinavir may now interact with D30 OD2 through a water-mediated hydrogen bond as shown in figure 4.24. If this is indeed the case, replacement of a direct hydrogen bond with a water-mediated hydrogen bond may be the likely reason for nelfinavir resistance in N88S protease. A similar change in hydrogen bonding interaction was implicated in resistance against the drug TMC-114 [Kovalevsky et al 2006].

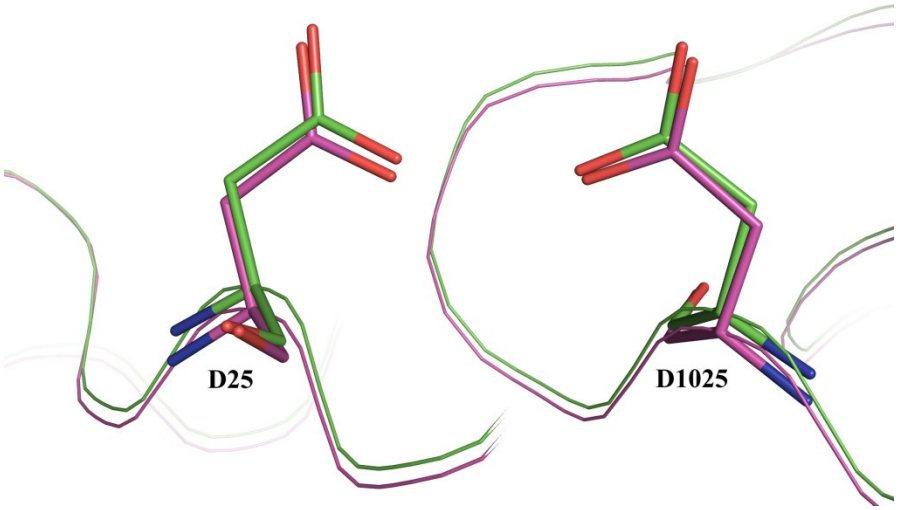

Figure 4.23: Deviations at catalytic loop due to N88D mutation: structural superposition of N88D HIV-1 protease (magenta) with wild type (green) [PDB ID 1LV1]. Catalytic aspartates are also shown.

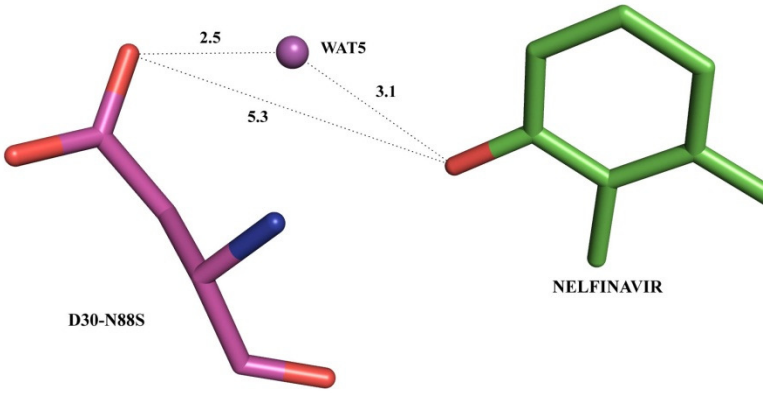

Figure 4.24: Proposed water mediated interaction of nelfinavir with D30 side chain in N88S HIV-1 protease: nelfinavir (green) position is obtained from structural superposition of wild type-nelfinavir complex structure [PDB ID 1OHR] on present N88S structure (magenta). WAT 5 is part of N88S HIV-1 protease structure.

Hyper susceptibility of N88S protease against amprenavir

Mutation N88S is also associated with hypersusceptibility against the drug amprenavir [Ziermann et al 2000]. N88S mutation alone was observed to give 25-fold increase in amprenavir susceptibility. The structure of HIV-1 protease-amprenavir complex determined earlier (PDB ID 1T7J) superposed well on N88S protease with RMSD of 0.583 Å for 198 Cα atom pairs. Figures 4.25 and 4.26 show derived position of parts of amprenavir molecule relative to the mutated residue at position 88. In N88S protease, D30 and D1030 are hydrogen bonded with two water molecules (WAT 5 & WAT 6) which are not present in wild type HIV-1 protease. If we assume that binding mode of amprenavir does not change in N88S protease, amprenavir will make two new hydrogen bonds with N88S through these water molecules (Figures 4.25 & 4.26). These new interactions will possibly increase the binding strength of the drug amprenavir to N88S protease and are likely reasons of hyper susceptibility of N88S protease to amprenavir.

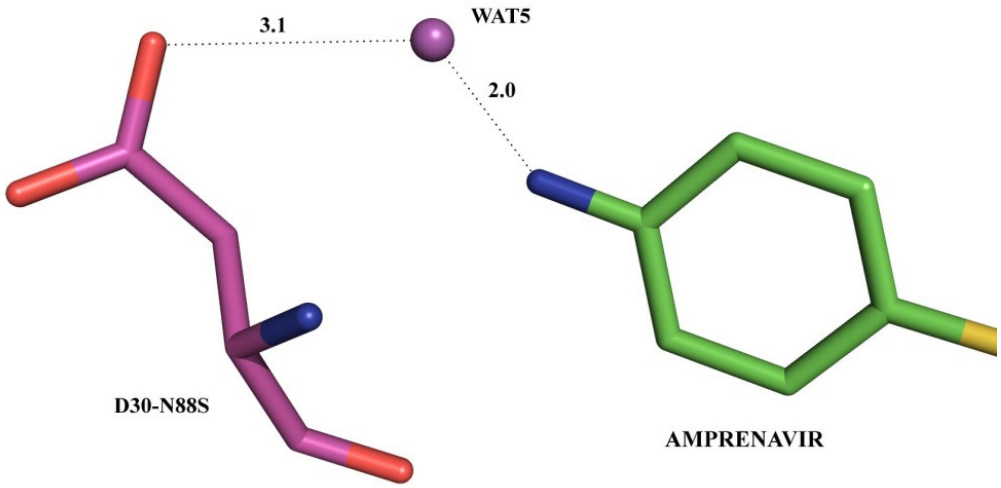

Figure 4.25: Modeling of amprenavir (green) in binding site of N88S protease (magenta) through structural superposition: Gain of water mediated hydrogen bond with D30 side chain. WAT5 is part of N88S HIV-1 protease structure.

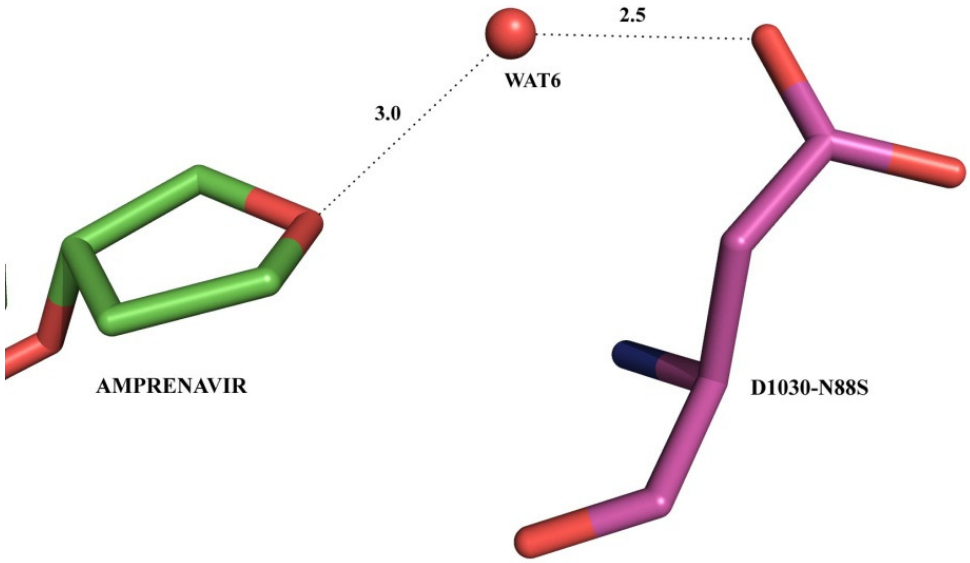

Figure 4.26: Modeling of amprenavir (green) in binding site of N88S protease (magenta) through structural superposition: Gain of water mediated hydrogen bond with D30 side chain. WAT6 is part of N88S HIV-1 protease structure.

Conclusion

High-resolution structures of two unliganded drug resistant mutants of HIV-1 protease have been reported. In both mutants, active site volume is increased thereby contributing to reduction of enzymatic activity. N88D mutation resulted in higher RMSD at catalytic triad, which may reduce cleavage efficiency of N88D protease. N88S mutation resulted in a large shift at D1030/30, which may change interactions of D1030/30 with substrates thereby reducing catalytic efficiency of N88S protease. The large shift in the position of D30 may result into loss of direct hydrogen bond between nelfinavir 2-methyl-3-hydroxyl benzamide group and D30 side chain and this loss could reduce binding affinity of the mutant protease toward nelfinavir. The large deviations of D30/1030 side chains and presence of two additional water molecules in N88S mutant are likely to result into gain of two water-mediated hydrogen bonds during binding of amprenavir to N88S protease and are suggested to be the reason for higher susceptibility of N88S toward amprenavir. This study has shown a unique mechanism of drug resistance by non-active site mutations, where non-active site mutations have caused changes in the interactions involving residues present in the active site. This calls for further study of role of non-active site residues in influencing binding of natural ligands and inhibitors. This may help in designing better and improved drugs, which will have higher genetic barrier to drug resistant mutations.

Chapter - V

Summary and Future Directions

Acquired Immuno Deficiency Syndrome (AIDS) is one of the most devastating diseases in the world claiming more than 30 million lives since its discovery in 1981. Even now, approximately 33 million people are infected with the virus and the number is increasing every year. AIDS is caused by Human Immunodeficiency Virus (HIV). Between two serotypes, HIV-1 and HIV-2, HIV-1 is more prevalent than HIV-1. Although there is yet no permanent cure or a protective vaccine available, several drugs have been developed for HIV/AIDS treatment. These drugs inhibit different stages of the viral life cycle and slow down the progression of HIV infection into fully developed AIDS. Currently, a cocktail of many of these drugs is used in Highly Active Anti-Retroviral Therapy (HAART). Inhibitors of HIV-1 protease are an important ingredient of HAART. HIV-1 produces its proteins in the form of polyproteins, which need to be cleaved, during maturation step, by HIV-1 protease to produce structural proteins and active enzymes. Inhibition of this step results into immature & non-infectious virus particles. Therefore, HIV-1 protease is a prime target for development of drugs against HIV/AIDS. As of now, 9 inhibitors of HIV-1 protease, (Saquinavir, Indinavir, Ritonavir, Nelfinavir, Amprenavir, Lopinavir, Atazanavir, Tipranavir and Darunavir) are approved by the Food and Drug Administration, USA, for the treatment of HIV/AIDS patients. These drugs have helped in improving the life standard and lifetime of HIV/AIDS patients. However, drug resistance mutations within HIV-1 protease are now limiting the use of these antiretroviral drugs, thereby necessitating continuous development of novel and better drugs. A clear understanding of the catalytic mechanism of HIV-1 protease will be helpful in this regard. Structural studies of HIV-1 protease in complex with true substrates can contribute to our knowledge of cleavage mechanism. A complementary

approach is to understand the molecular basis of drug resistance in HIV-1 protease. Crystal structures of drug resistant HIV-1 proteases and of their drug complexes will help in elucidating the molecular mechanism of drug resistance. This thesis reports few results obtained from both these approaches. HIV-1 protease cuts at nine different cleavage sites in HIV Gag and Gag-Pol polyproteins. One of the unique features of HIV-1 protease cleavage sites is that, a proline residue is present at C-terminus of scissile peptide bond in three of them. However, according to the proposed cleavage mechanisms in the literature, there is no distinction between proline containing substrates and non-proline substrates. To gain further insight into the cleavage mechanism, structural study of HIV-1 protease in complex with an oligo-peptide substrate representing one of the proline containing cleavage sites has been undertaken. Drug resistance mutations in HIV-1 protease are observed against all currently approved drugs. For the current study, I have chosen nelfinavir resistance mutations D30N, L90M, N88D, and N88S, which are located both in the active site and in non-active site region of the enzyme.

Site Directed Mutagenesis

HIV-1 protease is a homodimeric aspartyl protease consisting of two non-covalently associated chemically identical subunits of 99 amino acid residues. Due to the proximity in the three dimensional structure, of carboxy terminus of first monomer to amino terminus of second monomer, two copies of HIV-1 protease gene can be joined by a small linker DNA to produce a tethered dimer. For present work, a tethered dimer linked by a GGSSG linker is used. Chosen drug resistant mutations were introduced in HIV-1 protease gene using Stratagene's quick-change multi-site-directed mutagenesis method.

Protein Expression and Purification

Genes for both native and mutant HIV-1 protease inserted in PET-11A vector were over-expressed in *E.coli.* HIV-1 protease is accumulated as inclusion bodies, which were solubilized with 67% ice-cooled acetic acid. Pure, active HIV-1 protease was obtained by refolding of solubilized inclusion bodies by dialysis against refolding buffer at pH 6.5, containing 20mM PIPES, 100mM NaCl, 1mM DTT and 10% glycerol. The protein was concentrated and exchanged to a concentration of 1-4 mg/ml in 50mM sodium acetate buffer (pH 4.5) containing 1 mM DTT.

Crystallization and Structure Determination

HIV-1 protease was crystallized by hanging-drop vapor-diffusion method at room temperature. Hexagonal rod shaped crystals were obtained at pH 6.2 by using as precipitant, ammonium sulfate (1-6% of saturated solution) in sodium citrate/ sodium dihydrogen phosphate buffer (200/100mM) at pH 6.2. Crystals of complexes between nelfinavir and drug resistant HIV-1 protease were obtained by the method of co-crystallization under conditions similar to those for unliganded HIV-1 proteases. In an alternate condition for complex crystals, NaCl (0.5-1.25 M) is used as a precipitant with sodium acetate buffer (50 mM, pH 4.5) as reservoir. DMSO (1-10 %) was used as additive while growing HIV-1 protease-drug complex crystals. Diffraction data in the form of 1º oscillation frames were collected at European Synchrotron Radiation Facility (ESRF), Grenoble, France. Oscillation frames were indexed, integrated, and scaled using the software package XDS. All the crystal structures were solved by difference-Fourier method using native HIV-1 protease structure (PDB ID 1LV1) as template. Molecular

models were refined using computer programs, CNS and PHENIX. Electron density map interpretation and model building was done using software packages O and Coot.

Crystal Structure of HIV-1 Protease-Substrate Complex: Implications for Cleavage Mechanism

Crystal structure of HIV-1 protease in complex with substrate oligo-peptide of sequence SQNY*PIV representing the MA-CA cleavage site is determined at 2 Å resolution and refined to R_{cryst} = 20.54% and R_{free} = 23.75% for 12070 reflections. This is the first structure of a complex between proline containing oligo-peptide substrate and an active enzyme. Comparison of present structure with the structure containing products derived from non-proline substrate shows lateral shift of both products across the catalytic aspartates in the present complex. This lateral shift suggests, differential positioning of the P1' proline containing substrate in the active site. Another interesting difference is the absence of the hydrogen bond between scissile nitrogen and aspartate outer oxygen atoms. Since this hydrogen bond is suggested to act as a conduit for the essential protonation of scissile N atom, its absence in the present structure raises doubts about the suggested mechanisms. Interestingly, scissile N is hydrogen bonded to carboxyl O from N terminal product. This inter-product hydrogen bond was also observed in the complexes with products of non-proline substrates. These observations suggest that the proton transfer to scissile N atom may not happen through catalytic aspartate, either in all substrates or in substrates containing proline at P1' position. In fact, it supports the view of gem-hydroxyl being the source in this proton transfer step.

Crystal Structure of Unliganded D30N HIV-1 Protease: Reduced Cleavage Activity

D30N is a major resistance mutation against nelfinavir, and it drastically reduces cleavage efficiency of HIV-1 protease. Structure of D30N HIV-1 protease in unliganded form was solved to 1.65 Å resolution. The structure has been refined to R_{cryst} = 18.17% and R_{free} = 22.43% for 21751 reflections. The structure shows deviations in conserved catalytic loop (residues 24-34) including catalytic aspartates. The position of 30^{th} residue has shown deviations of 0.25/0.26 Å for Cα atoms and higher displacements of 0.78/1.05 Å in side chain atoms. D30 in wild type HIV-1 protease forms direct hydrogen bonding interactions with substrates through its side chain atoms. Structural displacements due to D30N mutation are likely to weaken these interactions by increasing the interatomic distances and by eliminating coulombic part of original interactions. These factors are suggested to be the reason for reduced cleavage activity of D30N protease.

Crystal Structure of D30N HIV-1 Protease-Nelfinavir Complex: Nelfinavir Resistance

Crystal of D30N HIV-1 protease in complex with nelfinavir was also determined to a resolution of 1.91 Å, and the structure was refined to R_{cryst}=21.87% with 13778 unique reflections. The structure shows that a crucial OH…O ionic hydrogen bond involving the 2-methyl-3-hydroxyl benzamide group from nelfinavir and carboxyl oxygen from the D30 of the wild-type protein is weakened due to increase in distance from 2.9 Å in wild type HIV-1 protease to 3.2 Å in D30N protease, and change in nature of hydrogen bond to a neutral OH…N/ O…HN hydrogen bond for the D30N mutant. This change is likely to reduce the strength of this hydrogen bond due to loss of coulombic interaction part of the original interaction. Therefore, it is proposed that reduction in the strength of this

crucial hydrogen bond with nelfinavir in D30N HIV-1 protease reduces the binding affinity of nelfinavir.

Crystal Structure of D30N/L90M HIV-1 Protease: Mutual Exclusivity of Active Site and Non-Active Site Mutations

In addition to D30N, L90M is another major mutation against nelfinavir, but both are mutually exclusive *in-vivo*. However, *in-vitro* experiments have shown that this combination provides high level of resistance against nelfinavir. In order to investigate this mutual exclusivity despite having selective advantage, crystal structure of D30N/L90M double mutant HIV-1 protease is determined in unliganded form. The structure shows that both individual and combined effects of these two mutations make them a lethal combination. D30N is already shown to reduce the catalytic activity of enzyme by its effect on substrate binding. Conformation of 30^{th} residue in D30N/L90M is similar to that in D30N structure suggesting that in double mutant also this mutation will have same effects. L90 is located just beneath the catalytic loop. L90M mutation introduces several new interactions of Cε atom of M90 with residues in catalytic loop. This is likely to affect the charge balance and pKa of catalytic aspartates, which critically affect the activity of HIV-1 protease. In addition, L90M mutation will affect the binding of substrates in the active site by restricting the movement of catalytic loop. Due to these individual effects, the catalytic activity of D30N/L90M will be drastically reduced to a level where it cannot support the growth of virus.

D30 and L90 are both located in conserved regions, which are part of the folding core of HIV-1 protease. Both these regions form local structural elements (LES) which interact to form a stable folding nucleus. It is possible that D30N and L90M mutations destabilize

their respective LES and impair folding thereby drastically affecting catalysis. The melting temperature measured by circular dichroism spectropolarimetry was found to be lower by 8ºC compared to that of wild type HIV-1 protease. This lower melting temperature could not be explained by reduced dimer stability, as there is not much change in the dimer interface. In-fact, in both D30N and D30N/L90M HIV-1 protease, interactions at the dimer interface are becoming stronger as measured by lower gap volume index and increased number of inter subunit interactions. Effect of these mutations on the stability of LES can explain this behavior. Further, L to M mutation in the core region of the protein makes protein inherently less stable. Immobilization of methionine imposes a greater entropic penalty than leucine due to burial of an extra bond in methionine. Therefore, it is suggested that, all these individual effects on catalytic activity and combined effect on the folding are likely reasons for the mutual exclusivity of D30N and L90M mutations.

D30N/L90M HIV-1 Protease-Nelfinavir Complex Structure: Drug Resistance

The structure of D30N/L90M in complex with nelfinavir shows that there are deviations in the binding of nelfinavir in case of D30N/L90M. All the four subsites are affected by these mutations resulting into sub-optimal binding. Also, the crucial hydrogen bond between nelfinavir and 30^{th} residue side chain is further weakened due to increase in distance to 3.4 Å from 2.9 Å in wild type. Flap water mediated interactions are also affected. These changes should be the reason behind reduced affinity toward nelfinavir.

Crystal structure of N88D HIV-1 protease: reduced cleavage activity

N88D alone reduces the catalytic efficiency of HIV-1 protease by 20–40% and is shown to be lethal for the virus. However, it also acts as a compensatory mutation for the D30N

mutation and restores loss of cleavage efficiency caused by D30N. Crystal structure of unliganded N88D mutant HIV-1 protease refined to resolution of 1.65 Å (comprising 21868 measure unique reflections) shows higher RMSD (0.28–0.33 Å) at catalytic loop. The active site volume is also increased to 1118.6 $Å^3$ from 996.6 $Å^3$ in wild type HIV-1 protease, which may result in sub-optimal binding of substrates. Both these effects are likely to reduce cleavage efficiency of N88D HIV-1 protease. Interestingly, D30 also show displacement in N88S protease suggesting possible structural linkage between them.

Crystal Structure of N88S HIV-1 Protease: Nelfinavir Resistance and Amprenavir Hypersusceptibility

Crystal structure of unliganded N88S mutant HIV-1 protease is determined to resolution of 1.8 Å. The structure revealed a large structural change in the conformation of D30. The oxygen atom D30 OD2 moves by about 2.8 Å to form a hydrogen bond with S88 OH. This large shift in the position of D30 will lead to loss of above-mentioned direct hydrogen bond between nelfinavir and side chain of D30. Through modeling of nelfinavir in present structure, it is suggested that this direct hydrogen bond may be replaced by a water-mediated one in N88S HIV-1 protease, thereby reducing binding affinity toward nelfinavir. N88S mutation also shows hypersusceptibility against amprenavir. It is suggested based on modeling studies, that amprenavir is likely to gain two water mediated hydrogen bonds in N88S protease and thereby is more effective against N88S mutant protease.

Future Directions

Crystal structure of HIV-1 protease in complex with proline containing oligo-peptide substrate shows that there could be a different mechanism to cleave proline-containing substrate. Further work with other proline containing substrates is required to confirm the same. An interesting observation about the source of protonation during cleavage mechanism has been made. Further studies in that direction may help elucidate detailed molecular mechanism for cleavage. Molecular mechanism of resistance caused by some of the drug resistance mutations against nelfinavir have been described in present work. This information can be used to modify nelfinavir to make it more effective against these drug resistant variants. New drug development can also use the information while optimizing interactions in the active site of HIV-1 protease. One key observation is that the drug should minimize its interactions with side chain of D30. Effect of D30N and L90M on the folding of HIV-1 protease is suggested to be one of the reasons of their mutual exclusivity. Further work is needed to support the hypothesis. NMR experiments and computational studies using these mutant proteases will be helpful. For N88S protease, loss of direct hydrogen bond is suggested to reduce the affinity of nelfinavir. To validate this hypothesis, N88S-nelfinavir complex structure needs to be solved. The structure can provide some more information about these non-active site mutations. N88S-amprenavir complex structure can be determined which could validate the hypothesis given for explaining amprenavir hypersusceptibility.

References

Adams P D, Afonine P V, Bunkóczi G et al (2010) PHENIX: a comprehensive python-based system for macromolecular structure solution. Acta Cryst D 66: 213-221.

Adamson C S and Freed E O (2007) Human immunodeficiency virus type 1 assembly, release, and maturation. Adv Pharmacol 55: 347-387.

Alexander M, Bor Y C, Ravichandran K et al (2004) Human immunodeficiency virus type 1 Nef associates with lipid rafts to downmodulates cell surface CD4 and class I major histocompatibility complex expression and to increase viral infectivity. J Virol 78: 1685-1696.

Alkhatib G and Berger E A (2007) HIV coreceptors: from discovery and designation to new paradigms and promise. Eur J Med Res 12: 375-384.

Bagossi P, Sperka T, Feher A et al (2005) Amino acid preferences for a critical substrate binding subsite of retroviral protease in type 1 cleavage sites. J Virol 79: 4213-4218.

Barre-Sinoussi F, Chermann J C, Rey F et al (1983) Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). Science 220: 868-871.

Bebenek K, Abbotts J, Wilson S H and Kunkel T A (1993) Error-prone polymerization by HIV-1 reverse transcriptase.Contribution of template-primer misalignment, miscoding, and termination probability to mutational hot spots. J Biol Chem 268: 10324-10334.

Bhat T N (1988) Calculation of an OMIT map. J Appl Cryst 21: 279-281.

Bhat T N, Baldwin E T, Liu B et al (1994) Crystal structure of a tethered dimer of HIV-1 proteinase complexed with an inhibitor. Nat Struct Biol 1: 552-556.

Binkowski T A, Naghibzadeg S and Liang J (2003) CASTp: Computed atlas of surface topography of proteins. Nucleic Acid Res 31: 3352-3355.

Bizub D, Weber I T, Cameron C E et al (1991) A range of catalytic efficiencies with avian retroviral protease subunits genetically linked to form single polypeptide chains. J Biol Chem 266: 4951-4958.

Boden D and Markowitz M (1998) Resistance to human immunodeficiency virus Type 1 protease inhibitors. Antimicrob Agents Chemother 42: 2775-2783.

Bonfanti P, Capetti A and Rizzardini G (1999) HIV disease treatment in the era of HAART. Biomed Pharmacother 53: 93-105.

Bonomi M, Barducci A, Gervasio F L and Parrinello M (2010) Multiple routes and milestones in the folding of HIV-1 protease monomer. PLoS one 5: e13208.

Bour S and Strebel K (2003) The HIV-1 Vpu protein: a multifunctional enhancer of viral particle release. Microbes and infection / Institut Pasteur 5: 1029-1039.

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

Broglia R A, Levy Y and Tiana G (2008) HIV-1 protease folding and the design of drugs which do not create resistance. Curr Opin Struct Biol 18: 60-66.

Brunger A T (1992) Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. Nature 30: 472-475.

Brunger A T, Adams P D and Rice L M (1998a) Recent developments for the efficient crystallographic refinement of macromolecular structures. Curr Op Struct Biol 8: 606-611.

Brunger A T, Adams P D, Clore G M et al (1998b) Crystallography & NMR System: A new software suite for macromolecular structure determination. Acta Cryst D54: 905-921.

Chan D C and Kim P S (1998) HIV entry and its inhibition. Cell 93: 681-684.

Chan D C, Fass D, Berger J M and Kim P S (1997) Core structure of gp41 from the HIV envelope glycoprotein. Cell 89: 263–73.

Charneau P, Borman A M, Quillent C et al (1994) Isolation and envelope sequence of a highly divergent HIV-1 isolate: definition of a new HIV-1 group. Virology 205: 247-253.

Chen V B, Arendall III W B, Headd J J et al (2010) MolProbity: all-atom structure validation for macromolecular crystallography. Acta Cryst D 66: 12-21.

Cheng Y S, Yin F H, Foundling S et al (1999) 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 24: 9660-9664.

Chertova E, Bess J W, JrCrise B J et al (2002) Envelope glycoprotein incorporation, not shedding of surface of glycoprotein (gp120/SU), is the primary determinant of SU content of purified human immunodeficiency virus type 1 and simian immunodeficiency virus. J Virol 76: 5315-5325.

Clavel F, guetard D, Brun-Vezinet F et al (1986) isolation of a new human retrovirus from West African patients with AIDS. Science 233: 343-346.

Cleland W W, Frey P A and Gerlt J A (1998) The low barrier hydrogen bond in enzymatic catalysis. J Biol Chem 273: 25529–25532.

Coakley E, Petropoulos C J and Whitcomb J M (2005) Assessing chemokine co-receptor usage in HIV. Curr Opin Infect Dis18: 9-15.

Coffin J M (1995) HIV population dynamics in vivo: implications for genetic variation, pathogenesis, and therapy. Science 267: 483-489.

Coffin J, Haase A, Levy J A et al (1986) Human immunodeficiency viruses. Science 232: 697-697.

Coiras M, Lopez-Huertas M R, Perez-Olmeda M and Alcami J (2009) Understanding HIV-1 latency provides clues for the eradication of long-term reservoirs. Nat Microbiol Rev 7: 798-812.

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

Darke P L, Leu C T, Davis L J et al (1989) Human immunodeficiency virus protease. Bacterial expression and characterization of the purified aspartic protease. J Biol Chem 264: 2307-2312.

Das A, Prashar V, Mahale S et al (2006) Crystal structure of HIV-1 protease in situ productcomplex and observation of a low-barrier hydrogen bond between catalytic aspartates. Proc Natl Acad Sci USA 103: 18464-18469.

Davies D R (1990) The Structure and function of the aspartic proteinases. Ann Rev Biophys Biophys Chem 19: 189-215.

Davis W D, Leaver-Fay A, Chen V B et al (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. Nucleic Acids Res 35: W375-W383.

De Gruttola V, Dix L, D'Aquila R et al (2000) The relation between baseline HIV drug resistance and response to antiretroviral therapy: re-analysis of retrospective and prospective studies using a standardized data analysis plan. Antivir Ther 5: 41-48.

De Leys R, Vanderborght B, Vanden Haesevelde M et al (1990) Isolation and partial characterization of an unusual human immunodeficiency retrovirus from two persons of west-central African origin. J Virol 64: 1207-1216.

De Meyer S, Azijn H, Surleraux D et al (2005) TMC114, a novel human immunodeficiency virus type 1 protease inhibitor active against protease inhibitor-resistant viruses, including a broad range of clinical isolates. Antimicrob Agents Chemother 2005 49: 23142321.

Diederichs K and Karplus A P (1997) Improved R-factors for diffraction data analysis in macromolecular crystallography. Nat Struct Biol 4: 269 – 275.

Dilanni C L, Davis L J, Holloway M K et al (1990) Characterization of an active single polypeptide form of the human immunodeficiency virus type 1 protease. J Biol Chem 265: 17348-17354.

Dimitrov D S (1997) How do viruses enter cells? The HIV coreceptors teach us a lesson of complexity. Cell 91: 721-730.

Douck D C, Brenchley J M, Betts M R et al (2002) HIV preferentially infects HIV-specific $CD4^+$ T-cells. Nature 417: 95-98.

Dunn B M (2002) Structure and mechanism of the pepsin-like family of aspartic peptidases. Chem Rev 102: 4431-4458.

Emerman M and Malim M H (1998) HIV-1 regulatory/accessory genes: keys to unraveling viral and host cell biology. Science 280: 1880-1884.

Emsley L P and Cowtan K (2004) Coot: Model – building tools for molecular graphics, Acta Cryst D 60: 2126-32.

Emsley L P, Lohkamp B, Scott W G and Cowtan K (2010) Features and development of Coot. Acta Cryst D 66: 486-501.

Erickson J, Neidhart D J, Van Drie J et al (1990) Design, activity, and 2.8 Å crystal structure of a C2 symmetric inhibitor complexed to HIV-1 protease. Science 249: 527-533.

Evans P (2006) Scaling and assessment of data quality. Acta Cryst D 66: 72-82.

Fanales-Belasio E, Raimondo M, Suligoi B and Butto S (2010) HIV virology and pathogenic mechanisms of infection: a brief overview. Ann Ist Super Sanita 46: 5-14.

Freed E O (1998) HIV-1 gag proteins: diverse functions in the virus life cycle. Virology 251: 1-15.

Gallo R C, Sarin P S, Gelmann E P et al (1983) Isolation of Human T-cell leukemia virus in acquired immune deficiency syndrome (AIDS). Science 220: 865-867.

Ganser-Pornillos B K, Yeager M and Sundquist W (2008) The structural biology of HIV assembly. Curr Opin Struct Biol 18: 203-217.

Gao F, Bailes E, Robertson D L et al (1999) Origin of HIV-1 in the chimpanzee Pan troglodytestroglodytes. Nature 397: 436-441.

Garcia J V and Miller A D (1991) Serine phosphorylation-independent downregulation of cell-surface CD4 by nef. Nature 350: 508–511.

Gassner N C, Basse W A and Matthews B W (1996) A test of the "jigsaw puzzle" model for protein folding by multiple methionine substitutions within the core of T4 lysozyme. Proc Natl Acad Sci 93: 12155-12158.

Gehlhaar D K, Moerder K E, Zichi D et al (1995) De novo design of enzyme inhibitors by Monte Carlo ligand generation. J Med Chem 38: 466-472.

Gilbert P B, McKeague I W, Eisen G et al (2003) Comparison of HIV-1 and HIV-2 infectivity from a prospective cohort study in Senegal. Stat Med 22: 573-593.

Goh W C, Rogel M E, Kinsey C M et al (1998) HIV-1 Vpr increases viral expression by manipulation of the cell cycle: a mechanism for selection of Vpr in vivo. Nat Med 4: 65-71.

Gong Y F, Robinson B S, Rose R E et al (2000) In vitro resistance profile of the human immunodeficiency virus type 1 protease inhibitor BMS-232632. Antimicrob Agents Chemother 44: 2319-2326.

Gonzalez L M F, Brindeiro R M, Aguiar R S et al (2004) Impact of nelfinavir resistance mutations on in vitro phenotype fitness and replication capacity of human immunodeficiency virus type 1 with subtype B and C protease. Antimicrob Agents Chemother 48: 3552-3555.

Gottlieb M S, Schroff R, Schanker H M, and Saxon A (1981) Pneumocystis pneumonia-Los Angeles. Morb Mortal Wkly Rep 30: 250-252.

Gougeon M L (2005) To kill or to be killed: how HIV exhausts the immune system. Cell Death Differ 12: 845-854.

Griffiths J T, Tomchak L A, Mills J S et al (1994) Interactions of substrates and inhibitors with a family of tethered HIV-1 and HIV-2 homo- and heterodimeric proteinases. J Biol Chem 269: 4787-4793.

Gustchina A and Weber I T (1990) Comparison of inhibitor binding in HIV-1 protease and in non-viral aspartic proteases: the role of the flap. FEBS Lett 269: 269-272.

Guyader M, Emerman M, Sonigo P et al (1987) Genome organization and transactivation of the human immunodeficiency virus type 2. Nature 326: 662-669.

Hansen J, Billich S, Schulze T et al (1988) Partial purification and substrate analysis of bacterially expressed HIV protease by means of monoclonal antibody. EMBO J 6: 1785-1791

Hebert E J, Giletto A, Sevcik J et al (1998) Contribution of a conserved asparagine to the conformational stability of ribonucleases Sa, Ba and T1. Biochemistry 37: 16192-16200.

Hemelaar J, Gouws E, Ghys P D and Osmanov S (2011) Global trends in molecular epidemiology of HIV-1 during 2000-2007. AIDS 25: 679-689.

Ho D D (1997) Perspectives series: host/pathogen interactions. Dynamics of HIV-1 replication *in vivo.* J Clin Invest 99: 2565-2567.

Hope T J (1999) The ins and outs of HIV Rev. Arch Biochem Biophys 365: 186-191.

Hyland L J, Tomaszek Jr T A and Meek T D (1991) Human immunodeficiency virus-1 protease. 2. Use of pH rate studies and solvent kinetic isotope effects to elucidate details of chemical mechanism. Biochemistry 30: 8454–8463.

International Committee on Taxonomy of Viruses (2002) 61.0.6. Lentivirus. (http://www.ncbi.nlm.nih.gov/ICTVdb/ICTVdB/61060000.htm). National Institutes of Health.

Ishima R, Ghirlando R, Toser J et al (2001) Folded monomer of HIV-1 protease. J Biol Chem 276: 49110-49116.

Jaskolski M, Tomasselli A G, Sawyer T K et al (1991) Structure at 2.5Å resolution of chemically synthesized human immunodeficiency type-1 protease complexed with a hydroxyethylene-based inhibitor. Biochemistry 30: 1600-1609.

Johnson E C B, Malito E, Shen Y et al (2007) Insights from atomic-resolution x-ray structures of chemically synthesized HIV-1 protease in complex with inhibitors. J Mol Biol 373: 573-586

Jones T A, Zou J Y, Cowan S W and 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.

Kabsch W (1993) Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. J App Cryst 26: 795-800.

Kabsch W (2010) XDS. Acta Cryst D 66: 125-132.

Kaldor S W, Kalish V J, Davie J F 2nd et al (1997) Viracept (nelfinavir mesylate, AG1343): a potent, orally bioavailable inhibitor of HIV-1 protease. J Med Chem 40: 3979-85.

Katoh I, Yoshinaka Y, Rein A et al (1985) Murine leukemia virus maturation: protease region required for conversion from immature to mature form and for virus infectivity. Virology 145: 280-292.

Kempf D J, Marsh K C, Denissen J F et al (1995) ABT-538 is a potent inhibitor of human immunodeficiency virus protease and has high oral bioavailability in humans. Proc Natl Acad Sci USA 92: 2484-2488.

Kempf D J, Marsh K C, Kumar G et al (1997) Pharmacokinetic enhancement of inhibitors of the human immunodeficiency virus protease by co administration with ritonavir. Antimicrob Agents Chemother 41: 654–660.

Kendrew J C, Bodo G, Dintzis H M et al (1958) A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. Nature 181: 662-666.

Klabe R M, Bacheler L T, Ala P J et al (1998) Resistance to HIV protease inhibitors: a comparison of enzyme inhibition and antiviral potency. Biochemistry 37: 8735-8742.

Kleywegt G J (2007) Crystallographic refinement of ligand complexes. Acta Cryst D 63: 94-100.

Kleywegt G J and Jones T A (1998) Databases in Protein Crystallography. Acta Cryst D 54: 1119-1131.

Koh Y, Nakata H, Maeda K et al (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: 31233129.

Kohl N E, Emini E A, Schleif W A et al (1988) Active human immunodeficiency virus protease is required for viral infectivity. Proc Natl Acad Sci USA 85: 4686-4690.

Korber B, Gaschen B, Yusim k et al (2001) Evolutionary and immunological implications of contemporary HIV-1 variation. Br Med Bull 58: 19-42.

Kovalevsky A V, Chumanevich A A, Liu F et al (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.

Kovalevsky A Y, Tie Y, Liu F et al (2006) Effectiveness of nonpeptide clinical inhibitor TMC-114 on HIV-1 protease with highly drug resistant mutations D30N, I50V and L90M. J Med Chem 49: 1379-1387.

Kozisek M, Bray J, Rezacova P et al (2007) Molecular analysis of the HIV-1 resistance development: enzymatic activities, crystal structures and thermodynamics of nelfinavir-resistant HIV protease mutants. J Mol Biol 374: 1005-1016.

Krohn A, Redshaw S, Ritchie J C et al (1991) Novel biding mode of highly potent HIV-proteinase inhibitors incorporating the (R)-hydroxyethylamineisostere. J Med Chem 34: 3340-3342.

Kumar M and Hosur M V (2003) Adaptability and flexibility of HIV-1 protease. Eur J Biochem 270: 1231–1239.

Kumar M, Kannan K K, Hosur M V et al (2002) Effects of remote mutation on the autolysis of HIV-1 PR: X-ray and NMR investigations. Biochem Biophys Res Commun 294: 395-401.

Kumar M, Prashar V, Mahale S and Hosur M V (2005) Observation of a tetrahedral reaction intermediate in HIV-1 protease substrate complex. Biochem J 389: 365-371.

Lapatto R, Blundell T, Hemmings A F et al (1989) X-ray analysis of HIV-1 proteinase at 2.7Å resolution confirms structural homology among retroviral enzymes. Nature 342: 299-302.

Lapid C (2003) PrimerX - Automated design of mutagenic primers for site-directed mutagenesis. http://www.bioinformatics.org/primerx/index.htm.

Larder B A and Kemp S D (1989)) Multiple mutations in HIV-1 reverse transcriptase confer high-level resistance to zidovudine (AZT). Science 246: 1155-1158.

Ledergerber B, Egger M, Erard V et al (1999) AIDS-related opportunistic illnesses occurring after initiation of potent antiretroviral therapy: the Swiss HIV Cohort Study. JAMA 282: 2220-2226.

Lee H, Darden T A and Pedersen L G (1996) An ab initio quantum mechanical model for the catalytic mechanism of HIV-1 protease. J Am Chem Soc 118: 3946–3950.

Levy J A (1993) HIV pathogenesis and long-term survival. AIDS 7: 1401-1410.

Levy J A (2009) HIV pathogenesis: 25 years of progress and persistent challenges. AIDS 23: 147-160.

Levy Y, Caflisch A, Onuchic J N and Wolynes P G (2004) The folding and dimerization of HIV-1 protease: evidence for a stable monomer from simulations. J Mol Biol 340: 67-79.

Li S, Juarez J, Alali M et al (1999) Persistent CCR5 utilization and enhanced macrophage tropism by primary blood human immunodeficiency virus type 1 isolates from advanced stages of disease and comparison to tissue-derived isolates. J Virol 73: 9741-9755.

Lin Y C, Beck Z, Morris G M et al (2003) Structural basis for distinctions between substrate and inhibitor specificities for feline immunodeficiency virus and human immunodeficiency virus protease. J Virol 77: 6589-6600.

Lipscomb L A, Gassner N C, Snow S F et al (1998) Context-dependent protein stabilization by methionine-to-leucine substitution shown in T4 lysozyme. Prot Sci 7: 765-773.

Liu F, Boross P I, Wang Y F et al (2005) Kinetic, stability, and structural changes in high-resolution crystal structures of HIV-1 protease with drug-resistant mutations

L24I, I50V, and G73S. J Mol Biol 354: 789-800.

Liu F, Kovalevsky A Y, Louis J M et al (2006) Mechanism of drug resistance revealed by the crystal structure of the unliganded HIV-1 protease with F53L mutation. J Mol Biol 358: 1191-1199.

Liu H, Müller-Plathe F and van Gunsteren W F (1996) A combined quantum/classical molecular dynamics study of the catalytic mechanism of HIV protease. J Mol Biol 261: 454–469.

Liu J, bartesaghi A, Borgnia M J et al (2008) Molecular architecture of native HIV-1 gp120 trimers. Nature 455: 109-113.

Loeb D D, Hutchinson C A, Edgell M H et al (1989) Mutational analysis of human immunodeficiency virus type 1 protease suggests functional homology with aspartic proteinases. J Virol 63: 111-121.

Lohkamp B, Emsley P & Cowtan K (2005) Coot News. CCP4 Newsletter, 42, Contribution 7.

Mahalingam B, Louis J M, Reed C C et al (1999) Structural and kinetic analysis of drug resistant mutants of HIV-1 protease. Eur J Biochem 263: 238-245.

Mahalingam B, Wang Y-F, Boross P I et al (2004) Crystal structures of HIV protease V82A and L90M mutants reveal changes in the indinavir-binding site. Eur J Biochem 271: 1516-1524.

Marin M, Rose K M, Kozak S L and Kabat D (2003) HIV-1 Vif protein binds the editing enzyme APOBEC3G and induces its degradation. Nat Med 9: 1398-1403.

Martinez-Picado J, Savara A V, Sutton L et al (1999) Replicative fitness of protease inhibitor-resistant mutants of human immunodeficiency virus type 1. J Virol 73: 3744-3752.

McCutchan F E (2000) Understanding the genetic diversity of HIV-1. AIDS 14: 531-544.

Miller M, Schneider J, Sathyanarayana B K et al (1989) Structure of complex of synthetic HIV-1 protease with a substrate-based inhibitor at 2.3Å resolution. Science 246: 1149-1152.

Mitsuya Y, Winters M A, Fessel W J et al (2006) N88D facilitates the co-occurrence of D30N and L90M and the development of multidrug resistance in HIV type 1protease following nelfinavir treatment failure. AIDS Res Human Ret 22: 1300-1305.

MMWR C (1981) Kaposi's sarcoma and Pneumocystis pneumonia among homosexual men. New York City and California. Morb Mortal Wkly Rep 30: 305-308.

Morgan D, Mahe C, Mayanja B et al (2002) HIV-1 infection in rural Africa: is there a difference in median time to AIDS and survival compared with that in industrialized countries? AIDS 16: 597-632.

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

Navia M A, Fitzgerald P M, McKeever B M et al (1989) Three-dimensional structure of aspartyl protease from human immunodeficiency virus HIV-1. Nature 337: 615-620.

Northrop D B (2001) Follow the protons: A low-barrier hydrogen bond unifies the mechanisms of the aspartic proteases. Acc Chem Res 34: 790–797.

Ode H, Matsuyama S, Hata M et al (2007) Computational characterization of structural role of the non-active site mutation M36I of human immunodeficiency virus type 1 protease. J Mol Biol 370: 598-607.

Ott D E (2008) Cellular proteins detected in HIV-1. Rev Med Virol 18: 159-175.

Painter J and Merritt E A (2006a) Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. Acta Cryt D 62: 439-450.

Painter J and Merritt E A (2006b) TLSMD web server for the generation of multi-group TLS models. J Appl Cryst 39: 109-111.

Pal D and Chakrabarti P (2001) Non-hydrogen bond interactions involving the methionine sulfur atom. J Biomol Struct Dyn 19: 115-128.

Palella F J, Delaney K M, Moorman A C et al (1998) Declining morbidity and mortality among patients with advanced human immunodeficiency virus infection. New Engl J Med 338: 853-860.

Parera M, Fernandez G, Clotet B and Martinez M A (2007) HIV-1 protease catalytic efficiency effects caused by random single amino acid substitutions. Mol Biol Evol

24: 382-387.

Park H, Suh J and Lee S (2000) Ab initio studies on the catalytic mechanism of aspartic proteinases: Nucleophilic versus general acid/general base mechanism. J Am Chem Soc 122: 3901–3908.

Patick A K, Duran M, Cao Y et al (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.

Pearl L H and Taylor W R (1987) A structural model for the retroviral protease. Nature 329: 351-354.

Pechik I V, Gustchina A E, Andreeva N S and Fedorov A A (1989) Possible role of some groups in the structure and function of HIV-1 protease as revealed by molecular modeling studies. FEBS Lett 247: 118-122.

Pellegrin I, Breilh D, Montestruc F et al (2002) Virologic response to nelfinavir-based regimens: pharmacokinetics and drug resistance mutations (VIRAPHAR study). AIDS 16: 1331-1340.

Perrin V and Mammano F (2003) Parameters driving the selection of nelfinavir-resistant human immunodeficiency virus type 1 variants. J Virol 77: 10172 – 10175.

Perutz M F, Rossmann M G, Cullis A F et al (1960) A three-dimensional Fourier synthesis at 5.5-Å resolution, obtained by X-ray analysis. Nature 185: 416–422.

Piana S and Carloni P (2000) Conformational flexibility of the catalytic asp dyad in HIV-1 protease: An ab initio study on the free enzyme. Proteins: Struct Funct Genet 39: 26-36.

Piana S, Carloni P and Rothlisberger U (2002) Drug resistance in HIV-1 protease: Flexibility- assisted mechanism of compensatory mutations. Protein Sci 11: 2393-2402.

Pillai B, Bhat S V, Kannan K K and Hosur M V (2004) Rapid screening for HIV-1 protease inhibitor leads through X-ray diffraction. Acta Cryst D 60: 594-596.

Pillai B, Kannan K K and Hosur M V (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.

Plantier J C, Leoz M, Dickerson J E et al (2009) A new human immunodeficiency virus derived from gorillas. Nat Med 15: 871-872.

Pollard V W and Malim M H (1998) The HIV-1 Rev protein. Annu Rev Microbiol 52: 491-532.

Pomerantz R J, Seshamma T and Trono D (1992) Efficient replication of human immunodeficiency virus type 1 requires a threshold level of Rev: potential implications for latency. J Virol 66: 1809-1813.

Prabhu-Jeyabalan M, Nalivaika E and Schiffer C A (2002) Substrate shape determines specificity of recognition for HIV-1 protease: analysis of crystal structures of six

substrate complexes. Structure 10: 369-381.

Prabu-Jeyabalan M, Nalivaika E A, King N M and Schiffer C A (2003) Viability of a drug-resistant human immunodeficiency virus type 1 protease variant: structural insights for better antiviral therapy. J Virol 77: 1306–1315.

Preston B D, Poiesz B J, and Loeb L A (1988) Fidelity of HIV-1 reverse transcriptase. Science 242: 1168-1171.

Ratner L, Haseltine W, Patarca R et al (1985) Complete nucleotide sequence of AIDS virus, HTLV-III. Nature 313: 277-284.

Reeves J D and Doms R W (2002) Human immunodeficiency virus type 2. J Gen Virol 83: 1253–65.

Resch W, Ziermann R, Parkin N et al (2002) Nelfinavir- resistant, amprenavir-hypersusceptible strains of human immunodeficiency virus type 1 carrying an N88S mutation in protease have reduced infectivity, reduced replication capacity, and reduced fitness and process the gag polyprotein aberrantly. J Virology 76: 8659-8666.

Reynolds C, Damerell D and Jones S (2009) ProtorP: a protein-protein interaction analysis server. Bioinformatics 25: 413-414.

Rhee S-Y, Gonzales M J, Kantor R et al (2003) Human immunodeficiency virus reverse transcriptase and protease sequence database. *Nucleic Acids Res* 31: 298-303.

Richards A D, Roberts R, Dunn B M et al (1989) Effective blocking of HIV-1 proteinase activity by characteristic inhibitors of aspartic proteinases. FEBS Lett 247: 113-117.

Robbins A H, Coman R M, Bracho-Sanchez E et al (2010) Structure of the unbound form of HIV-1 protease: comparison with unbound forms of proteases from other HIV subtypes. Acta Cryst D 66: 233-242.

Roberts J D, Bebenek K, and Kunkel T A (1988) The accuracy of reverse transcriptase from HIV-1.Science 242: 1171-1173.

Roberts N A, Martin J A, Kinchington D et al (1990) Rational design of peptide based HIV proteinase inhibitors. Science 248: 358-361.

Robertson D L, Hahn B H and Sharp P M (1995) Recombination in AIDS viruses. J Mol Evol 40: 249–59.

Robinson B S, Riccardi K A, Gong Y F et al (2000) BMS-232632, a highly potent human immunodeficiency virus protease inhibitor that can be used in combination with other available antiretroviral agents. Antimicrob Agents Chemother 44: 2093-2099.

Rodriguez E J, Angeles T S and Meek T D (1993) Use of nitrogen-15 kinetic isotope effects to elucidate details of the chemical mechanism of human immunodeficiency virus 1 protease. Biochemistry 32: 12380–12385.

Rose G D (1993) Hydrogen bonding, hydrophobicity, packing and protein engineering. Annu Rev Biophys Biomol Struct 22: 381-415.

Rose R B, Craik C S, Douglas N L and Stroud R M (1996) Three-dimensional structures of HIV-1 and SIV protease product complexes. Biochemistry 35: 12933-44.

Roth M, Carpentier P, Kaikati O et al (2002) FIP: a highly automated beamline for multiwavelength anomalous diffraction experiments. Acta Cryst D 58: 805-814.

Saah A J, Haas D W, DiNubile M J et al (2003) Treatment with Indinavir, Efavirenz, and Adefovir after failure of nelfinavir therapy. The J Infect Dis 187: 1157-1162.

Sarafianos S G, Marchand B, Das K et al (2009) Structure and function of HIV-1 Reverse Transcriptase: molecular mechanisms of polymerization and inhibition. J Mol Biol 385: 693-713.

Schackman B R, Gebo K A, Walensky R P et al (2006) The lifetime cost of current HIV care in the United States. Med Care 44: 990-997.

Schechter I and Berger A (1967) On the size of the active site in protease. I. Papain. Biochem Biophys Res Comm 27: 157-162.

Schwartz O, Maréchal V, Le Gall S et al (1996) Endocytosis of major histocompatibility complex class I molecules is induced by the HIV-1 Nef protein. Nat Med 2: 338–342.

Seelmeier S, Schmidt H, Turk V and von der Helm K (1988) Human immunodeficiency virus has an aspartic-type protease that can be inhibited by pepstatin A. Proc Acad Sci USA 85: 6612-6616.

Shafer R W (2002) Genotypic testing for human immunodeficiency virus type -1 drug resistance. Clin Microbiol Rev 15: 247-277.

Shafer R W and Vuitton D A (1999) Highly active antiretroviral therapy (HAART) for the treatment of infection with human immunodeficiency virus type 1. Biomed Pharmacother 53: 73-86.

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

Sham H L, Kempf D J, Molla A et al (1998) ABT-378, a highly potent inhibitor of the human immunodeficiency virus protease. Antimicrob Agents Chemother 42: 3218-3224.

Sharp P M and Hahn B H (2010) The evolution of HIV-1 and the origin of AIDS. Philos Trans R Soc Lond B 365: 2487-2494.

Sharp P M, Robertson D L and Hahn B H (1995) Cross-species transmission and recombination of 'AIDS' viruses. Philos Trans R Soc Lond 349: 41-47.

Silva A M, Cachau R E, Sham H L and Erickson J W (1996) Inhibition and catalytic mechanism of HIV-1 aspartic protease. J Mol Biol 255: 321–346.

Simoes E A, Babu P G, John T J et al (1987) Evidence for HTLV-III infection in prostitutes in Tamil Nadu (India). Indian J Med Res 85: 335-338.

Simon F, Mauclere P, Roques P et al (1998) Identification of a new human immunodeficiency virus type 1 distinct from group M and group O. Nat Med 4: 1032-1037.

Sobolev V, Sorokine A, Prilusky J et al (1999) Automated analysis of interatomic contacts in proteins. Bioinformatics 15, 327-332.

Spinelli S, Liu Q Z, Alzari P M et al (1991) The three-dimensional structure of the aspartyl protease from the HIV-1 isolate BRU. Biochemie 73: 1391-1396.

Stumptner-Cuvelette P, Morchoisne S, Dugast M et al (2001) HIV-1 Nef impairs MHC class II antigen presentation and surface expression. Proc Natl Acad Sci USA 98: 12144-12149.

Sugiura W, Matsuda Z, Yokomaku Y et al (2002) Interference between D30N and L90M in selection and development of protease inhibitor-resistant human immunodeficiency virus type 1. Antimicrob Agents Chemother 46: 708-715.

Surleraux D L, Tahri A, Verschueren W G et al (2005) Discovery and selection of TMC114, a next generation HIV-1 protease inhibitor. J Med Chem 48: 18131822.

Swanson C M and Malim M H (2008) SnapShot: HIV-1 proteins. Cell 133: 742e1.

The PyMOL Molecular Graphics System, Version 1.4.1, Schrödinger, LLC.

Toh H, Ono M, Saigo K and Miyata T (1985) Retroviral protease-like sequence in yeast transposon Ty1. Nature 315: 691-691.

Tozser J (2010) Comparative Studies on Retroviral Proteases: Substrate Specificity Viruses 2: 147-165.

Tozser J, Yin F H, Cheng Y S et al (1997) Activity of tethered human immunodeficiency virus 1 protease containing mutations in the flap region of one subunit. Eur J Biochem 244: 235-241.

Turner B G and Summers M F (1999) Structural biology of HIV. J Mol Biol 285: 1-32.

Turner S R, Strohbach J W, Tommasi R A et al (1998) Tipranavir (PNU-140690): A potent, orally bioavailable nonpeptidic HIV protease inhibitor of the 5, 6-dihydro-4-hydroxy-2-pyrone sulfonamide class. J Med Chem 41: 3467-3476.

UNAIDS Global report (2010): UNAIDS report on the global AIDS epidemic 2010.

Verkhivker G, Tiana G, Camilloni C et al (2008) Atomistic simulations of HIV-1 protease folding inhibition. Biophys Journal 95: 550-562.

Wallqvist A, Smythers G W and Covell D G (1998) A cooperative folding unit in HIV-1 protease. Implications for protein stability and occurrence of drug-induced mutations. Prot Eng 11: 999-1005.

Weber I T, Miller M, Jaskolski M et al (1989) Molecular modeling of the HIV-1 protease and its substrate binding site. Science 243: 928-931.

Weiss M S (2001) Global indicators of X-ray data quality. J Appl Cryst 34: 130-135.

Whittle H, Morris J, todd J et al (1994) HIV-2 infected patients survive longer than HIV-1 infected patients. AIDS 8: 1617-1620.

Wlodawer A and Vondrasek J (1998) Inhibitors of HIV-1 protease: A major success of structure-assisted drug design. Annu Rev Biophys Biomol Struct 27: 249-284.

Wlodawer A, Miller M, Jaskolski M et al (1989) Conserved folding in retroviral proteases: Crystal structure of a synthetic HIV-1 protease. Science 245: 616–621.

Word J M, Lovell S C, Richardson J S, and Richardson D C (1999) Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation. J Mol Biol 285:1735–1747.

Wyatt R, and Sodroski J (1998) The HIV-1 envelope glycoproteins: fusogens, antigens, and immunogens. Science 280: 1884-1888.

Xie D, Gulnik S, Gustchina E et al (1999) Drug resistance mutations can affect dimer stability of HIV-1 protease at neutral pH. Protein Sci 8: 1702-1707.

Yerly S, Rickenbach M, Popescu M et al (2001) Drug resistance mutations in HIV-1-infected subjects during protease inhibitor-containing highly active antiretroviral therapy with nelfinavir or indinavir. Antivir Ther 6: 185-189.

Zhu P, chertova E, Bess J et al (2003) Electron tomography analysis of envelope glycoprotein trimers on HIV and simian immunodeficiency virus virions. Proc Natl Acad Sci USA 100: 15812-15817.

Zhu P, Liu J, Bess Jr J et al (2006) Distribution and three-dimensional structure of AIDS virus envelope spikes. Nature 441: 847-852.

Ziermann R, Limoli K, Das K et al (2000) A mutation in Human Immunodeficiency Virus type 1 protease, N88S, that caused in vitro hypersensitivity to amprenavir. J Virol 74: 4414-4419.

Zoete V, Michielin O and Karplus M (2002) Relation between sequence and structure of HIV-1 protease inhibitor complexes: a model system for the analysis of protein flexibility. J Mol Biol 315: 21–52.

Publications from thesis work

1. Bihani S C, Das A, Prashar V, Ferrer J-L and Hosur M V (2009) Resistance mechanism revealed by crystal structures of unliganded nelfinavir-resistant HIV-1 protease non-active site mutants N88D and N88S. Biochem Biophys Res Comm 389: 295–300.

2. Bihani S C, Das A, Prashar V, Ferrer J-L and Hosur M V (2008) X-ray structures of HIV-1 protease-in-situ product complex. Proteins: Struct Func Bioinfo, 74: 594-602.

3. Hosur M V, Prashar V, Das A, and Bihani S (2008) Structural investigations of HIV-1 protease complex with substrates and inhibitors; In 37th National Seminar on Crystallography at Kolkata (India), IT4.

4. Bihani S C, Ferrer J-L and Hosur M V. Molecular basis for mutual exclusivity of D30N and L90M in HIV-1 protease. Manuscript under preparation.

5. Bihani S C, Ferrer J-L and Hosur M V. Effect of drug resistance mutation D30N on the binding of drug nelfinavir. Manuscript under preparation.