Synopsis
Human Immunodeficiency Virus (HIV) is the causative agent of Acquired ImmunoDeficiency Syndrome (AIDS), and has killed more than 27 million people around the world (UNAIDS report on the global AIDS epidemic, 2010). The HIV type-1 protease (HIV-1 PR) is the virally encoded enzyme that processes, in a temporally regulated manner, the viral polyproteins, gag and gag-pol, which are precursor to functional enzymes and structural proteins of the AIDS virus. HIV-1 PR is a unique endopeptidase with abilities to cleave the peptide bond between amino acid residues X and Y, even when Y is a proline residue. The cleavage sites are classified into two types: type-1 where amino acid residue Y is a proline, and type-2 where both X and Y are non-proline amino acids. From amongst the nine cleavage sequences in viral polyproteins of HIV, three are type-1 sites and six are type-2 sites. Inhibition of HIV-1 PR leads to the production of immature & non-infectious viral particles and therefore HIV-1 PR is an important target for the design of anti-AIDS drugs. Ten drugs approved by United States Food and Drug Administration (FDA) are currently in the market. However, there is a need to develop new drugs because, emergence of drug-resistant mutant enzymes have reduced the efficacy of existing drugs. Critical inputs to development of more effective inhibitors are likely to come from knowing correct enzyme mechanism, enzyme-substrate and enzyme-inhibitor interactions. The candidate has therefore used the technique of single crystal X-ray diffraction to obtain accurate three dimensional structures of: 1) unliganded HIV-1 PR 2) HIV-1 PR/substrate complexes, and 3) drug-resistant HIV-1 PR mutant complexed with cognate drug ritonavir. Results of these investigations are reported in this thesis titled, X-ray crystallographic studies on HIV-1 protease.

**Thesis highlights:**

The unique contribution of the candidate is structural mapping of the peptide bond cleavage reaction catalyzed by HIV-1 PR. The candidate has obtained the first crystal structures of complexes between an active HIV-1 PR enzyme and natural cleavable substrates of both types. In the complexes with a type-2 substrate, the substrate is captured in the active site at three different stages of the cleavage reaction: when bound as a regular peptide, after in-situ modification into a tetrahedral reaction intermediate, and when cleaved into product peptides, which are still bound in the active site of the enzyme. He has discovered an inter enzyme-substrate short ionic hydrogen bond (SIHB) in the tetrahedral intermediate complex, and an intraenzyme low barrier hydrogen bond (LBHB) just after the substrate is cleaved into product peptides. In the complex with type-1 substrate, the substrate is cleaved and the C-terminal peptide bond about the proline residue displays cis conformation. These are very novel and atomic level descriptions of substrate recognition and processing by HIV-1 protease. Based on these structural inputs he has given a detailed mechanism of the peptide bond hydrolysis by HIV-1 PR. Through very high resolution structures of unliganded HIV-1 PR, the candidate has found that the interaspartate hydrogen bond is not a LBHB, in contradiction to the latest mechanistic proposal. The candidate has also determined the structures of the complexes of FDA
approved drug, ritonavir with native as well as HIV-1 PR mutants that are resistant to ritonavir. These structures reveal that, in V82F and M36I mutants, loss of hydrophobic interactions contribute significantly toward development of resistance. A chapter-wise summary of the thesis work is given below.

Chapter 1 is a brief introduction to HIV-1, the causative agent of AIDS, and to HIV-1 PR, the virally encoded protease enzyme, which is essential for the survival and propagation of HIV-1 [1]. HIV-1 is a complex virus, and has a life cycle consisting of the following steps: a) internalization through the host cell receptor mediated endocytosis, b) disassembly of the viral coat proteins, reverse transcription and integration of the viral genome to the host chromosome, c) transcriptional activation and synthesis of the viral polyproteins using host resources and d) maturation and infectivity of the assembled virion through the action of viral protease, HIV-1 PR [2]. Intervention into each of these steps is being pursued in the development of drugs against AIDS [3]. HIV-1 PR is an enzyme that belongs to the family of aspartyl proteases characterized by the highly conserved amino acid sequence Asp-Thr-Gly at their active site. The viral enzyme is functional as a homodimer cleaving the natural gag and gag-pol polyproteins (see figure below).

![Ribbon diagram of HIV-1 PR dimer. Each monomer is coloured differently. The catalytic aspartates and the attacking water molecule near it, are shown as yellow sticks and red sphere. The structural water molecule, flap water, is coloured grey.](image)

The active site cavity of the enzyme is hydrophobic, and is demarcated into substrate-binding pockets designated S1(S1’), S2(S2’), S3(S3’), …. the primed pockets being on the C-terminal side of the scissile peptide bond of the substrate. The substrate residues binding into these pockets are correspondingly designated P1(P1’), P2(P2’), P3(P3’) …. etc. The N-terminal peptide product after cleavage is described as the P-product while the C-terminal product is described as the Q-product. The pH-activity profile is bell shaped over a wide range of pH from 2.0 to 7.0, with the pH optimum being 5.5 [4]. In the three dimensional structure of HIV-1 PR homodimer, the C-terminus of subunit one is spatially very close to the N-terminus of the second subunit. A single chain enzyme formed by tethering the two subunits through an oligopeptide linker is shown to have enzyme activities very similar to the native protein [5]. A catalog of mutations that arise in AIDS patients being treated with protease inhibitor drugs has been compiled, and it is found that mutations occur almost at 40 % residues in the protease sequence. A general overview of enzyme catalysis will also be described in brief.
Chapter 2 of the thesis is a description of the X-ray crystallographic techniques adopted for protein structure determination, with reference to the work reported in the thesis. X-rays produced in in-house generator and also at synchrotrons are used to characterize the crystals and collect diffraction data for structure analysis. X-ray crystallography is a standalone and powerful technique providing three dimensional structures of molecules to atomic detail. Three dimensional structures of macromolecules and their assemblies having molecular weights of even a few million Daltons can be studied by this method, once suitable crystals are obtained. Recent developments in the technologies of production and detection of X-rays are presented in this chapter along with a brief description of the steps involved in crystal structure determination. The oscillation data collection strategy and data processing procedures, namely indexing, scaling and merging of diffraction data are discussed. Methodologies for solving the ‘phase problem’ and procedures for refinement of macromolecular structures are also described.

Chapter 3 of the thesis begins with the biochemical methodologies used in site-directed mutagenesis, expression, purification and crystallization of HIV-1 PR. *E. coli* bacteria engineered for overexpressing tethered HIV-1 PR enzyme was used for protein production. This clone was also used to prepare site-directed mutant enzyme using tools of genetic engineering. In this tethered enzyme the C-terminus of one subunit was linked to the N-terminus of the second subunit of the dimer via the linker peptide of sequence GGSSG, and the N-terminus of the first subunit contained an overhang of 57 extra amino acids. The overhang has a Phe-Pro cleavage site at the very beginning thereby giving a built-in assay to assess the activity of the overexpressed protein. The overexpressed protein of MW 29 kD was extracted from inclusion bodies in denatured form and then refolded into active conformer of MW 22 kD by step-wise slow dialysis. Purity of the protein was checked through SDS-PAGE, and enzyme activity was checked by UV-VIS spectroscopic assay that uses a chromogenic substrate. Typical yields were 15 mgs of pure protein per litre of *E.coli* culture. Crystallization was attempted manually through the hanging-drop vapour diffusion method, and also by sitting-drop vapour diffusion method using the in-house CyBio protein crystallization robot. Hexagonal rod shaped crystals appeared within a few days and continued to grow till few months. The maximum thickness was 0.05 mm (see figure below).

This chapter also reports first glimpses of active HIV-1 PR complexed to a natural decapeptide substrate, \( \text{NH}_2\text{-AETF*YVDGAA-COOH} \), where the ‘*’ represents the peptide bond that is cleaved. The amino acid sequence is a type-2 cleavage site corresponding to reverse transcriptase-RNAse-H
junction in the viral polyprotein. This is the first successful study on a complex involving an active protease enzyme and a true substrate molecule. All earlier attempts by researchers elsewhere to prepare such complexes through co-crystallization had failed [6]. The unliganded-closed flap conformation of HIV-1 PR discovered by earlier researchers in the laboratory of the candidate has enabled the candidate to prepare the complex by using the soaking method. The pH of the soak solution and the duration of the soak have been systematically varied with a view to sample as many reaction coordinates as possible.

The structure at pH 6.2 and 3 days soak has been solved to a resolution of 1.65 Å. The crystallographic R- and R-free values are 21.4 % and 25.9 % respectively. The unit cell parameters are a= b= 62.03 Å and c= 81.78 Å. The difference electron density map shows that the decapeptide substrate is cleaved \textit{in situ}, with the two product peptides (AETF \[P\text{-product}\] and YVDGAA \[Q\text{-product}\]) still held in the active site. This is the first report of an \textit{in situ} bi-product complex of any proteolytic enzyme. Another discovery of this work was the low barrier hydrogen bond (LBHB) of length 2.3 Å between the inner oxygens of the two catalytic aspartates (see figure below). (Coordinates deposited under PDB code: 2NPH).

The structure at pH 2.5 and a soak-time of 1 day has been solved to a resolution of 1.76 Å. The refined reliability factors are R=22.28 % and R-free=25.00 %. The unit cell parameters are a= b= 62.56 Å and c= 81.86 Å. The crystal structure of the complex shows that the substrate is trapped as a tetrahedral \textit{gem}-diol intermediate inside the active site of the enzyme, after scissile peptide carbonyl is attacked by the water molecule. The refined coordinates of the tetrahedral intermediate have been deposited in the Protein Data Bank (PDB id: 3MIM). The structure shows there is a short ionic hydrogen bond (SIHB) between one \textit{gem}-diol oxygen and outer oxygen of a catalytic aspartate, at this stage of the reaction. This H-bond is important for proper positioning of the substrate scissile peptide bond, and for polarization of the carbonyl group in preparation for the attack by catalytic water. This structure also disproves the suggestion that enzyme-substrate covalent intermediate is formed in the catalysis by HIV-1 PR. The structure of this tetrahedral intermediate, which is a mimic of the transition state, offers a unique opportunity to design tight-binding-inhibitors.

The structure of the complex at pH 8.0 and a soak time of 24 hours has been refined to a resolution of 1.89 Å having R=15 % and R-free=18 %. The structure of the complex shows that the substrate is uncleaved in the active site cavity. Interestingly there is no catalytic water molecule in the active site to act as a nucleophile.

Comparison of the structures described above shows the re-organisation of the hydrogen bonds at the catalytic center as the reaction progresses. A molecular mechanism for the cleavage reaction based on these X-ray structures is proposed (see figure below).
Chapter 4 of the thesis describes structures of complexes with a type-1 oligopeptide substrate of sequence NH$_2$-VSFN*PQITC-COOH. This sequence corresponds to the transframe-protease cleavage site in the gag-pol polyprotein.

The structure at pH 6.2 and a soak time of 3 days has been solved to a resolution of 1.70 Å. The structure is refined to R=18 % and R-free=23 %.

The substrate is modeled as a peptide cleaved in situ at the scissile peptide bond. The most interesting finding is that the peptide bond between Proline and Glutamine residues of the Q-product peptide adopts cis conformation, which has not been observed before.

The structure at pH 6.2 and soak time 1 day has been solved to resolution of 1.74 Å. The structure is refined to R=18 % and R-free=22 %. The substrate is modeled as a cleaved peptide. The mode of binding of the two fragments generated in situ is similar but the occupancy is less. Here also the peptide bond between Proline and Glutamine residues is in cis conformation.

The increase in time of soaking depicts the increase in the population of the substrate in the active site cavity suggesting that diffusion of substrate into active sites of the enzyme molecules in the crystal is the rate-determining step. Observation of a cis conformation about a peptide bond in the product peptide is a very novel result from these series of structures. Comparison of complexes with type-1 and type-2 substrates shows that the ring structure of proline causes the Q-product to be released first in contrast to the P-product release in type-2 substrates.

The recent kinetic iso-mechanism proposal is critically dependent on the presence of a LBHB between inner oxygen atoms of catalytic aspartates in the free enzyme [7]. To investigate into this, the candidate has determined very high-resolution structures of unliganded HIV-1 PR over a range of pH values spanning enzyme activity (2.0, 6.2 and 7.5). Chapter 5 of the thesis describes results of these structural studies. The crystals of free enzyme were first grown at pH 6.2, and then were soaked in the buffers having the respective pH’s for 24 hours before data collection. The unliganded HIV-1 PR structures at pH values of 6.2, 2.0 and 7.5 are refined to resolutions of 1.35 Å, 1.63 Å and 1.72 Å respectively. The crystallographic R-factors are: R=18.00 % (R-free=20.24 %), R=15.72 % (R-free=19.59 %) and R=19.1 % (R-free=22.75 %) respectively. While the rest of the structure remains the same (average RMSD of Ca atoms = 0.23 Å), there is a difference in the water structure at the catalytic center, and this difference could contribute to differences in enzyme activity. At pH 6.2,
there is one water molecule, hydrogen bonding to catalytic aspartates. However, in the structures at pH 2.0 and at pH 7.5 there are two water molecules, which are hydrogen bonded to each other at 2.6 Å, and each one, in turn, hydrogen bonds separately to the outer oxygen atoms of the two catalytic aspartates.

The pair of water molecules could cause steric hindrance to the approaching scissile peptide bond during catalysis, thereby drastically reducing the activity of the enzyme at these pH values. Further, proper positioning of a water molecule for activation may not happen when two water molecules are bound in the active site, and this also could contribute to loss of activity. The inter aspartate hydrogen bond between inner oxygens is longer at pH 2.0 than at both pH 7.5 and pH 6.2. At none of these pH values, the interaspartate hydrogen bond is short enough to be classified as a LBHB. This observation is contrary to what was assumed in the kinetic iso-mechanism proposal [7].

Identification of the protonation state of the catalytic aspartates is crucial to understanding the molecular mechanism. The protonation state of the aspartates in native HIV-1 PR is unknown since ultrahigh resolution X-rays data are not yet available. Since even medium resolution neutron diffraction study will determine proton positions, deuterated protein was prepared and large crystals measuring upto a size of 0.3 mm$^3$ have been grown. Attempts to further increase their size to 1 mm$^3$ volume, required for neutron diffraction study, are being made by using the seeding technique.

Chapter 6 of the thesis describes the structures of complexes between FDA approved drug, ritonavir, with both native and drug-resistant mutants of HIV-1 PR.

Ritonavir is one of the essential drugs used for the treatment of AIDS, and mutation V82F confers resistance against ritonavir [8]. The mutant protein was prepared by using the standard procedures of site-directed mutagenesis, protein expression and purification. The mutant protein was co-crystallized with ritonavir by vapour diffusion in hanging drops. Diffraction data was collected and processed to a resolution of 1.90 Å. The structure is refined to R= 17.6 % and R-free= 21 %. The electron density for the mutation residue Phe and the drug ritonavir, are clearly visible in the OMIT map. The central hydroxyl group of ritonavir makes strong hydrogen bonds with carboxyl oxygens of the catalytic aspartates.

Crystals of ritonavir complexed to native HIV-1 PR were prepared by co-crystallization, and X-ray diffraction data were collected to a resolution of 1.60 Å. The reliability factors are R= 19.6 % and R-free= 21.9 %. The central hydroxyl group makes strong hydrogen bonds with the oxygens of two catalytic aspartates.

Comparing the V82F mutant and native complex structures show that the drug binds in similar fashion with most of the interactions being hydrophobic in nature. The conformation of the drug has altered to optimize hydrophobic interactions with the mutant protein. The hydrogen bonds with catalytic aspartates are very similar, but the interactions near the mutation site are different. In the mutant structure, the backbone of the protein near the 80’s loop is shifted away as a result of steric clash with the P1 benzyl side chain of the drug. The mechanism of drug resistance is due to loss of favourable van der Waals contacts.
The mutation M36I is known to be a minor non-active site drug-resistant mutation against ritonavir [9]. The mutant protein was prepared by using the standard procedures of site-directed mutagenesis, protein expression and purification. The mutant protein complex with ritonavir was co-crystallized by vapour diffusion in hanging drops. This structure, refined to a resolution of 1.60 Å with R= 20 % and R-free= 22.5 %, is also described in this chapter. The electron density for isoleucine residue is visible in the OMIT map.

Comparison with the native-ritonavir complex structure reveals shifts in the backbone near the 80’s loop in the active site, and also near the 36th residue, resulting in a decrease in the non-polar interactions with the drug. This structure shows the effect at the active site of a mutation away from the active site of HIV-1 PR.

REFERENCES:

Publications in peer reviewed international journals and symposia/conference presentations are given below.

**List of journal publications:**

1. **Amit Das**, S.Bihani, V.Prashar, J.-L.Ferrer and M.V.Hosur;
   “Crystal structures of complexes of wild-type and V82F mutant HIV-1 Protease with ritonavir - insights into drug resistance.”
   Manuscript submitted for publication.

2. **Amit Das**, J.-L.Ferrer and M.V.Hosur;
   “X-ray snapshots of HIV-1 protease catalysis and substrate recognition”;

   “X-ray snapshot of HIV-1 protease in action: observation of tetrahedral intermediate and its SIHB with catalytic aspartate”; 
   *Journal of American Chemical Society* (2010), 132, 6366-6373. *This paper has been selected to be published in JACS Select, 2011.*

4. **Amit Das**, D.R.Rao and M.V.Hosur; 
   “X-ray Structure of HIV-1 Protease Tethered Dimer Complexed to Ritonavir”;

   “X-ray structure of insitu HIV-1 protease-product complex: observation of a LBHB between catalytic aspartates”;
   *Proceedings of National Academy of Sciences, USA* (2006), 103 (49), 18464-18469.

6. S.Bihani, **Amit Das**, V.Prashar, J.-L.Ferrer and M.V.Hosur; 
   “X-ray structure of in-situ HIV-1 protease-product complex”; 

7. V.Prashar, **Amit Das**, S.C.Bihani, J.-L.Ferrer and M.V.Hosur; 
   “Catalytic water co-existing with product peptide in the active site of HIV-1 Protease: Insights into the reaction mechanism”;

**List of symposia presentations, conferences, etc.:**

1. M.V. Hosur, S.Bihani, V.Prashar and **Amit Das**;
   "X-ray structures of drug resistant HIV-1 protease and drug complexes";


