X-RAY CRYSTALLOGRAPHIC STUDIES ON HIV-1 PROTEASE

By

AMIT DAS

Bhabha Atomic Research Centre
Trombay, Mumbai – 400085

A thesis submitted to the
Board of Studies in Chemical Sciences

In partial fulfillment of requirements

For the Degree of

DOCTOR OF PHILOSOPHY

of

HOMI BHABHA NATIONAL INSTITUTE

December, 2011

Recommendations of the Viva Voce Board

As members of the Viva Voce Board, we certify that we have read the dissertation prepared by **Amit Das** entitled "**X-ray Crystallographic Studies on HIV-1 Protease**" and recommend that it may be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy.

Seshadri **Date:** 9/3/2012

Chairman – Prof. M. Seshadri

M V Hosur **Date:** 9/3/2012

Guide / Convener – Prof. M.V. Hosur

K. Suguna **Date:** 9-3-2012

External Examiner – Prof. K. Suguna, Indian Institute of Science, Bangalore.

A. K. Sharma **Date:** 9.3.2012

Member 1 – Prof. A. K. Sharma

Date: 9/3/2012

Member 2 – Prof. A.K. Tyagi

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.

Signature of Guide: M V Hosur

(Prof. M.V. Hosur)

Date: 9/3/2012

Place: Trombay, BARC, Mumbai.

STATEMENT BY THE 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.

AMIT DAS

DECLARATION

I, hereby declare that the investigation presented in the thesis has 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.

AMIT DAS

DEDICATIONS

This thesis is dedicated to my beloved family who has been a constant source of love, inspiration, encouragement and strength in the troughs and crests of my dissertation tenure. I am grateful to them for their long patience and moral support. I would also like to express my heart-felt gratitude to my parents.

AMIT DAS

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my research guide, Dr. M. V. Hosur, for his guidance, constructive criticisms and supporting my scientific endeavour during the entire study. His valuable scientific advice helped me sail through, during my research work. It was my privilege to have been associated with Dr. M. V. Hosur during the years of my research career.

I would also like to acknowledge the National Facility for Macromolecular Crystallography at Protein Crystallography Section, Solid State Physics Division, Bhabha Atomic Research Center for X-ray crystallographic and laboratory facilities.

I am thankful to my doctoral committee members Dr. S. K. Apte, Chairman, Dr. M. Seshadri and Dr. A.K. Tyagi for their valuable criticisms, suggestions and recommendations. I also thank my collaborators Dr. J.-L.Ferrer, GSY-IBS and FIP,ESRF at Grenoble, France for synchrotron access and Dr. S. Mahale, NIRRH, Mumbai, India for initial peptide synthesis.

My sincere thanks to Dr. K.K.Kannan, Dr. Mukesh Kumar, Mr.Vishal Prashar and Mr. S. Bihani, for their support in day to day laboratory work. I am thankful to Mr. S.R. Jadhav for laboratory help.

AMIT DAS

Contents **Page**

****

Imagination is more important than knowledge

- Albert Einstein

****

Synopsis

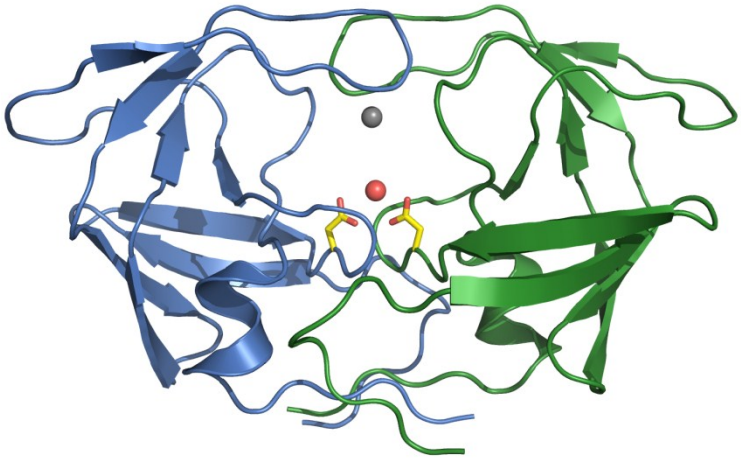
Human Immunodeficiency Virus (HIV) is the causative agent of Acquired ImmunoDeficiency Syndrome (AIDS), and has killed more than 27 million people around the world (UNAIDS report on the global AIDS epidemic, 2010). The HIV type-1 protease (HIV-1 PR) is the virally encoded enzyme that processes, in a 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-I 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.

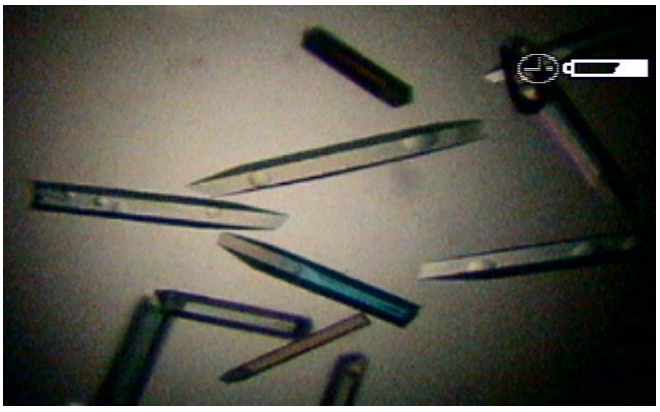
The active site cavity of the enzyme is hydrophobic, and is demarcated into substrate-binding pockets designated S1(S1'), S2(S2'), S3(S3'), …. the primed pockets being on the C-terminal side of the scissile peptide bond of the substrate. The substrate residues binding into these pockets are correspondingly 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).

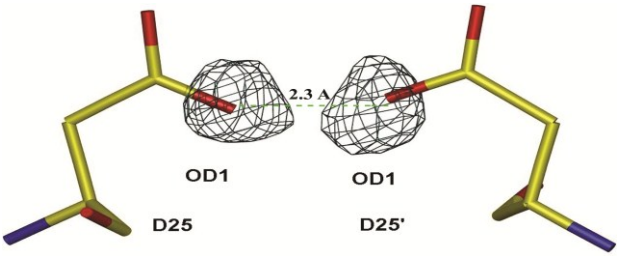


Hexagonal rod-shaped crystals of HIV-1 PR in the crystallization droplet, grown under the optimized conditions. Many such crystals were used for soaking experiments.

This chapter also reports first glimpses of active HIV-1 PR complexed to a natural decapeptide substrate, NH_2-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 *in situ*, with the two product peptides (AETF [P-product] and YVDGAA [Q-product]) still held in the active site. This is the first report of an *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).

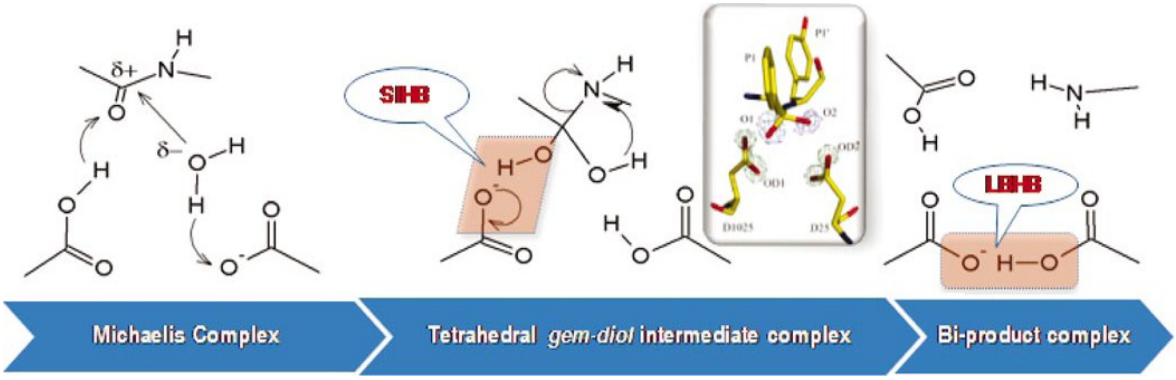


Simulated–annealed OMIT electron density map of the inner oxygens (OD1s) of catalytic aspartate dyad (D25 and D25') involved in LBHB (2.3Å). The modeled OD1s' are inside the OMIT map.

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 *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 *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-VSFNF*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 Cα 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 36^{th} 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:

1. Debouck C., Navia M. A., Fitzgerald P. M. D., McKeever B. M., Leu C-T., Heimbach J. C., Herber W. K., Sigal, I. S., Darke P. L. and Springer J. P. (1987). *Proc. Natl. Acad. Sci. U.S.A.,* 84: 8903–8906.
2. Weiss R.A. (1993). *Science,* 260 (5112): 1273–9; Peng, C.; Ho, B. K.; Chang, T. W. and Chang, N. T. (1989). *J. Virol.,* 63:2550–2556.
3. Brik A. and Wong C.H. (2003). *Org. Biomol. Chem.,* 1 (1): 5–14; Wlodawer A. and Vondrasek J. (1998). *Annu. Rev. Biophys. Biomol. Struct.,* 27: 249–284.
4. Hyland L. J., Tomaszek T. A., Roberts G. D., Carr S. A., Magaard V. W., Bryan H. L., Fakhoury, S. A., Moore M. L., Minnich M. D., Culp J. S., DesJarlias R. L. and Meek T. D. (1991). *Biochemistry,* 30: 8441–8453.
5. Cheng Y.S.E., Yin F.H., Foundling S., Blomstrom D. and Kettner C.A. (1990). *Proc. Natl. Acad. Sci. U.S.A.,* 87: 9660–9664.
6. Rose R.B., Craik C.S., Douglas N.L. and Stroud R.M. (1996). *Biochemistry,* 35:12933–12944.
7. Northrop D.B. (2001). *Acc. Chem. Res.,* 34:790–797.
8. Molla A., Korneyeva M., Gao Q., Vasavanonda S., Schipper P. J., Mo H-M., Markowitz M., Chernyavskiy T., Niu P., Lyons N., Hsu A., Granneman G. R., Ho D. D., Boucher C. A. B., Leonard J. M., Norbeck D. W. and Kempf D. J. (1996). *Nature Medicine,* 2:760–766.
9. Baharuddin A., Fischer M., van Maarseveen N., Nijhuis M., Boucher C.A.B., Hilfenfeld R. and Mesters J.R. (2006). *Antivir Ther.*, 11:S71.

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";
***Acta Crystallographica* (2011), A67, C425.**

3. Amit Das, S.Mahale, V.Prashar, S.Bihani, J.-L.Ferrer and M.V.Hosur;
"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";
***Protein and Peptide Letters* (2007), 14, 565-568.**

5. Amit Das, V.Prashar, S.Mahale, J.-L.Ferrer, L.Serre and M.V.Hosur;
"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*";*
***Proteins: Structure, Function and Bioinformatics* (2008), 74(3), 594-602.**

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";
***PLoS One* (2009), 4(11), e7860, 1-8.**

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";
Indo-US symposium/Workshop, Modern Trends in Macromolecular Structures, IL-29, Indian Institute of Technology, Mumbai, India (2011).

2. **Amit Das**, M.V.Hosur, V.Prashar and S.Bihani; European Molecular Biology Symposium, Recent advances in Macromolecular Crystallography, National Chemical Laboratory, Pune, India (2008).

3. M.V.Hosur, **Amit Das**, V.Prashar and S.Bihani; "Structural investigations of HIV-1 Protease complex with substrates and inhibitors"; 37th National Symposium on Crystallography, Jadavpur University, Kolkata, India (2008).

4. M.V.Hosur, **Amit Das** and V.Prashar; "Structure of a tethered dimer of HIV-1 protease complexed to a decapeptide corresponding to a RT-RH junction"; Joint Conference of the Asian Crystallographic Association and the Crystallographic Society of Japan, Epochal Tsukuba, Japan (2006).

5. M.V.Hosur, B.Pillai, M.Kumar, V.Prashar, **Amit Das**, S.C.Bihani and S. Mahale; "X-ray structures of HIV-1 protease substrate complexes"; International Symposium on Emerging Trends in Genomic and Proteomic Sciences, National Institute of Reproductive Research and Health (ICMR), Mumbai, India (2006).

6. **Amit Das**, V.Prashar and M.V.Hosur; "Structure of a tethered dimer of HIV-1 protease complexed to a decapeptide corresponding RT-RH junction"; National Seminar on Crystallography, National Physical Laboratory, New Delhi, India (2006).

7. M.V. Hosur, B.Pillai, M.Kumar, V.Prashar, **Amit Das** and S.Mahale;
"Structures of HIV-1 protease complexed to peptide substrates"; Indian Biophysical Society symposium, Kolkata, India (2006).

List of Figures

b) Standard nomenclature P1, _ _ _ , Pn and P1',_ _ _ _ , Pn' is used to designate amino acid residues of peptide substrates. The corresponding binding sites on the HIV-1 protease are referred to as S1, _ _ _ Sn and S1', _ _ _ _ Sn' subsites.

Figure 20: a) A cryoprotected HIV-1 protease crystal on a litholoop during data collection at 100K. **b)** The diffraction frame of a HIV-1 protease crystal.

Figure 21: The *β*–sheet region after structure solution using 1LV1 model along with the *2mFo-DFc* map, after one cycle of restrained refinement.

Figure 22: The protein backbone cartoon diagram is shown, where the two subunits are coloured green and pink. Simulated annealed *mFo-DFc* difference electron density in the active-site cavity of HIV-1 protease is shown as blue mesh, where the substrate model is omitted. The refined substrate model is shown along with catalytic aspartates as yellow sticks.

Figure 23: a) Stereo diagram of *2mFo-DFc* map in blue and *mFo–DFc* map in red when the substrate model refined is of a regular peptide. The peptide bond is in negative density of *mFo–DFc* map. **b)** Stereo diagram showing SA omit map overlaid with the three models: regular peptide model (cyan), tetrahedral hydrated peptide model (brown) and cleaved peptide model (yellow). Electron density for P1, P3, P1' and P3' residues defines the single orientation of the substrate.

Figure 24: SA omit map to show positions of inner oxygens of catalytic aspartates. The difference density in the *mFo–DFc* map is contoured at 3.0 σ level.

Figure 25: The superposed structures of apo-enzyme with the bi-product complex (yellow stick). The apo-enzyme catalytic water molecule (green sphere) is only shown. Note that the catalytic water is closer to the O2 oxygen of the P-product.

Figure 26: Hydrogen bonding interactions in the active site of the cleaved substrate/HIV-1 protease complex at pH 6.2. Side chain atoms of the substrate, except P3 Glu and P3' Asp, are not shown for clarity. Five structural water molecules are also shown as red spheres.

Figure 27: Hydrogen bonding interactions at the catalytic center are shown by dotted lines. O1 and O2 are the carboxylate oxygens of the Phe residue (P1) and the N is the amino nitrogen of the Tyr residue (P1'). Inner oxygens of the catalytic aspartates (D25 and D1025) are labeled as OD1 and the outer oxygens facing the substrates are labeled OD2.

Figure 28: Superposition of the cleaved peptide/active HIV-1 protease structure (yellow) with structures (rmsd 0.75 Å) representing **a)** the regular peptide/inactive HIV-1 protease D25N mutant complex (blue) and **b)** product release stages of the cleavage reaction with the P-product before release (green) where only the P1 residue of the P-product peptide is shown.

Figure 29: Stereo view of SA omit map contoured at 2.0 σ, when the substrate model is not included in map calculation. The refined tetrahedral intermediate, TI, (yellow sticks) and the regular peptide (green sticks) models are shown for comparison. Note that the scissile nitrogen atom of the regular peptide model is not inside electron density. Only the major orientation is shown for the sake of clarity.

Figure 30: Stereo view of SA omit map (blue) contoured at 2.4 σ when the two oxygen atoms in the major occupancy TI are not included in the calculation of structure factors. Also shown is the stereo view of SA omit map (green) contoured at 5.2 σ when the four oxygen atoms of the catalytic aspartates are omitted from structure factor calculations.

Figure 31: Stereo view of SA omit map contoured at 2.0 σ along with the refined TI model shown in the major orientation.

Figure 32: Comparison of catalytic centre from three superposed structures: 1) inactive HIV-1 protease/regular peptide complex, [PDB ID: 1KJG] (light green), 2) present structure (yellow) and 3) unliganded HIV-1 protease [PDB ID: 2G69] (orange). Note the proximity of O2 to the putative catalytic water in unliganded structure, and of O1 to the peptidyl oxygen in the inactive D25N complex.

Figure 33: Hydrogen bonding network in the active site cavity of TI/HIV-1 protease complex. The interactions of the *gem*-diol hydroxyls with the catalytic aspartates are shown in green. Only main chain atoms are shown for clarity, except for side-chain atoms of P3', D25/1025, R8/1008 and R87/1087.

Figure 34: Hydrogen bonds between aspartates and *gem*-diol of the TI at the catalytic center. This hydrogen bonding pattern has not been observed before.

Figure 35: Superposition of the present structure [PDB ID: 3MIM] (yellow) with that of the product complex [PDB ID: 2NPH] (purple). The flaps are shown as cartoon diagram.

Figure 36: Superposition of the TI complex [PDB ID: 3MIM] (green) with that of the product complex [PDB ID: 2NPH] (yellow) showing the P1-P1' residues and the catalytic aspartates only. Note the change in the planarity of the carboxylate of the catalytic aspartates.

Figure 37: a) The cartoon diagram of HIV-1 protease with the SA omit map and substrate model occupying the active site cavity. A phosphate ion is also shown as orange sticks. **b)** SA omit with the refined substrate model. **c)** SA omit map in the P2-P2' region of the substrate. The maps are contoured at 1.5 σ. The refined nine residues of uncleaved substrate model are shown as yellow sticks where the substrate atoms are excluded from structure factor calculations.

Figure 38: a) The refined tetrahedral intermediate, TI, (cyan sticks) is shown superposed on the regular peptide (yellow sticks) model. SA omit map contoured at 1.5 σ is shown, where the atoms of scissile peptide bond were excluded from map calculation. Note that the two hydroxyl groups of the tetrahedral peptide model are not well inside the electron density b) SA omit maps are shown where the scissile peptide bond (green, contoured at 1.5 σ) and the four oxygens of the catalytic aspartates (red, contoured at 2.7 σ) are omitted during map calculation.

Figure 39: Hydrogen bonding interactions of the uncleaved peptide in the active-site of HIV-1 protease.

Figure 40: Hydrogen bonding at the catalytic center of regular peptide/HIV-1 protease complex.

Figure 41: The regular peptide model (yellow sticks) superposed onto the FDA approved hydroxy inhibitors, ritonavir (blue sticks) and DMP323 (green sticks). Note the peptide carbonyl oxygen, O1 is near to the hydroxyl oxygens of the two different inhibitors.

Figure 42: The regular peptide model (yellow sticks) superposed onto the TI/HIV-1 protease complex [PDB ID: 3MIM] (cyan sticks). The flap water molecules are shown as red spheres.

Figure 43: The regular peptide model (yellow) superposed onto the same peptide/D25N inactive HIV-1 protease complex (PDB ID: 1KJG) (brown) and apo-HIV-1 protease. The catalytic water molecule present in apo-enzyme is shown as pink sphere, which is near the scissile carbonyl oxygen of the regular peptide.

Figure 44: Schematic diagram of the proposed reaction mechanism based on the TI [PDB ID: 3MIM] and product [PDB ID: 2NPH] complexes of HIV-1 protease involving the same oligopeptide substrate. **a)** Shows the SIHB and LBHB in TI and product complexes respectively. **b)** Shows the proton shifts in Michaelis, TI and product complexes.

Figure 45: Molecular structure of the 10-residue type-1 peptide substrate (H_2N-VSFNF*PQITC-COOH) of HIV-1 protease that was used for soaking experiments. The scissile peptide bond between the Phe (P1) and Pro (P1') residues is indicated by the wavy line.

Figure 46: a) SA omit *mFo-DFc* map (light blue), contoured at 2.0 σ, level showing the refined substrate model (yellow sticks) which is cleaved between Phenylalanine and Proline residues in 1d. **b)** SA omit *mFo-DFc* map (blue), contoured at 2.0 σ, showing the refined substrate model (orange sticks) which is cleaved between Phe and Pro residues in 3d. The atoms of the substrate were not included during map calculations.

Figure 47: Hydrogen bonding in the active site of HIV-1 protease in the **a)** 1d; the partial water molecules are shown in green spheres, and **b)** 3d. Some side-chain atoms are not shown for clarity. Hydrogen bonding interactions are shown as black dotted lines. Note the changes in orientations of P1 Phe carboxylates and P1' prolines in the two complexes.

Figure 48: a) SA *mFo-DFc* map (blue) with the peptide bond between P1' and P2' of the refined substrate modeled as *cis* peptide (yellow sticks) and *trans* peptide (green sticks) in 1d. **b)** SA *mFo-DFc* map (green) with the peptide bond between P1' and P2' of the refined substrate modeled as *cis* peptide (yellow sticks) and *trans* peptide (magenta sticks) in 3d. The maps were contoured at 2.4 σ.

Figure 49: Superposition of refined coordinates of 1d (yellow sticks) and 3d (cyan sticks) structures in the active site of HIV-1 protease. Note that in 3d structure, P1' residue has moved away from and P1 has moved closer to the plane of the catalytic aspartates as compared to the 1d structure.

Figure 50: Hydrogen bonding in the catalytic center of HIV-1 protease/type-1 substrate complex for **a)** 1d (yellow sticks), and **b)** 3d (cyan sticks) structures. Note that there is no hydrogen bond between P1' nitrogen atom with the enzyme.

Figure 51: Superposition of refined coordinates of the peptide soaked in 1d (yellow sticks), 3d (cyan sticks) and uncleavable reduced substrate (pink sticks; PDB ID: 2AOH) complexed to active HIV-1 protease.

Figure 52: Superposition of refined coordinates of bi-product complexes of type-1 peptide, 1d (yellow sticks), 3d (cyan sticks) and the type-2 peptide substrate (orange sticks; PDB ID: 2NPH) at the catalytic centre of HIV-1 protease. Note the differences in the Cα positions between the type-1 and type-2 substrates.

Figure 53: Superposition of refined coordinates of bi-product complexes of type-1 substrate peptide, 3d having Phe-Pro (cyan sticks) and Tyr-Pro (purple sticks; PDB ID: 3DOX) at the catalytic centre of HIV-1 protease.

Figure 54: The crystals of wild-type HIV-1 protease grown in **a)** H_2O having 2 mg/ml protein, and **b)** D_2O having 0.89 mg/ml protein.

Figure 55: The SA omit map of the two catalytic aspartates and associated water molecule/s (W1 and W2) in the refined structures of unliganded HIV-1 protease at **a)** pH 2.0 (green sticks) contoured at 2.1 σ, **b)** pH 6.2 (yellow sticks) contoured at 2.5 σ and, **c)** pH 7.5 (salmon sticks) contoured at 2.0 σ. The atoms of D25, D1025 and the water molecules shown were not included in the structure factor calculations.

Figure 56: The hydrogen bonds in the active site of unliganded HIV-1 protease at **a)** pH 2.0, protein atoms are shown as green sticks, **b)** pH 6.2, protein atoms are shown as yellow sticks and catalytic water having major occupancy is shown, and **c)** pH 7.5, protein atoms are shown in salmon sticks. Waters at the catalytic site are shown as green spheres, fully occupied water molecules as red spheres and partial waters as cyan or light blue spheres.

Figure 57: The hydrogen bonds at the catalytic center of unliganded HIV-1 protease at **a)** pH 2.0 (green sticks), note the two water molecules (W1 and W2) at pH 2.0 are coplanar to the catalytic carboxylates, **b)** pH 6.2 (yellow sticks), and **c)** pH 7.5 (salmon sticks). **d)** Superposed refined structures at the three pHs. Note that the water pairs (W1 and W2) at pH 2.0 and pH 7.5 are almost perpendicular to each other.

Figure 58: The protonation states at the catalytic center of unliganded HIV-1 protease at pH 2.0, pH 6.2 and pH 7.5.

Figure 59: The trajectory for nucleophilic attack on sp^3, sp^2, and sp carbons are shown. In HIV-1 protease catalysis, the catalytic water molecule approaches the δ^+ polarized sp^2 carbon of the scissile carbonyl to form the TI. The 105° angle in addition reactions to carbonyl is called Burgi-Dunitz angle. EWG stands for 'electron withdrawing group'.

Figure 60: The Michaelis complex showing the protonation via hydrogen bonds (curved arrows) and catalytic water attacking the peptide polarized carbonyl bond (straight arrow).

Figure 61: Superposed structure of unliganded HIV-1 protease (yellow sticks) at pH 6.2 onto **a)** pH 2.0 (green sticks), 1KJG (blue sticks) and TI (light blue sticks) structures and, **b)** pH 7.5 (salmon sticks), 1KJG (blue sticks) and TI (light blue sticks) structures. The pair of water molecules at pHs 2.0 and 7.5 are shown as green spheres and the *gem*-diol oxygen derived from catalytic water is labeled O2. Only carbonyl carbon tetrahedron is shown for TI model.

Figure 62: Crystals of **a)** V82F/ritonavir and **b)** M36I/ritonavir complexes obtained by co-crystallization.

Figure 63: The chemical structure of HIV-1 protease inhibitor, ritonavir.

Figure 64: The F82/F1082 residues in V82F/ritonavir and I36/I1036 residues in M36I/ritonavir were not included in map calculations. **a)** Simulated annealed *mFo-DFc* map (blue) contoured at 2.5 σ showing F82 and F1082 residues fits well. The refined F82 (purple sticks) and F1082 residue (orange sticks) in V82F mutant and the V82 residue (green sticks) in WT HIV-1 protease models are superposed for comparison. **b)** Simulated annealed *mFo-DFc* map (green) contoured at 2.5 σ showing I36 and I1036 residues fitting well. The refined I36 and I1036 residues (yellow sticks) in M36I mutant and the M36 residue (magenta sticks) in WT HIV-1 protease models are superposed for comparison. The red *mFc-DFo* map indicates that M1036 residue is incorrect.

Figure 65: Simulated annealed *mFo-DFc* map, contoured at 2.2 σ, covering the refined coordinates of the inhibitor, ritonavir, bound in the active site cavity. Ritonavir atoms were not included in map calculations. **a)** Grey coloured map corresponds to ritonavir (black sticks) soaked in WT protease crystal. **b)** Red coloured map corresponds to ritonavir (cyan sticks) co-crystallized with WT protease. **c)** Purple coloured map corresponds to ritonavir (yellow sticks) co-crystallized with V82F protease. **d)** Blue coloured map corresponds to the two orientations of ritonavir (yellow sticks) co-crystallized with M36I protease.

Figure 66: The inhibitor, ritonavir, is shown within the active site of WT (blue sticks) and V82F (yellow sticks) complexes after superposition. Note the difference in orientations of the benzyl residues of ritonavir at the S1/S1' pockets. There is a slight shift perpendicular to the plane of the isopropyl thiazolyl ring in the S3 pocket and slight rotation in thiazolyl ring in the S2' pocket. Note that the catalytic D25 and D1025 residues of the protein are perfectly superposed.

Figure 67: Hydrogen bonding involving ritonavir in the active site of WT protease. Ritonavir central hydroxyl (O41) is placed asymmetrically with respect to the D25 and D1025.

Figure 68: Hydrogen bonding involving ritonavir in the active site of V82F mutant protease. Ritonavir central hydroxyl (O41) is placed symmetrically with respect to the D25 and D1025.

Figure 69: Structural superposition of V82F/ritonavir (magenta) and the WT/ritonavir (green) complexes. Cartoon diagram of the HIV-1 protease showing the F82 side chain (yellow stick) is pointing away from the active site cavity. The V82 (magenta stick) of the WT complex is pointing into the active site cavity. P81 and P1081 also have different conformations in the two complexes.

Figure 70: The inhibitor, ritonavir, is shown within the active site of WT (green sticks) and M36I (yellow sticks) complexes after superposition. Note there is very little difference in orientations of the benzyl residues of ritonavir at the S1/S1' pockets. Note that the catalytic D25 and D1025 residues of the protein are perfectly superposed.

Figure 71: Hydrogen bonding involving ritonavir in the active site of M36I mutant protease. Ritonavir central hydroxyl (O41) is placed symmetrically with respect to the D25 and D1025.

Figure 72: Structural superposition of M36I/RTV (yellow ribbon) and the WT/RTV (green ribbon) complexes. The shifts in the backbone near the 36/1036th and 81/1081st residues may be noticed.

Figure 73: Superposed stick models of V82F/ritonavir (yellow), V82F/DMP323 (magenta) and V82F/I84V/DMP323 (green) complexes. It may be noted that V84 of double mutant is superposed well with I84 of V82F single mutants whereas the F82 exhibits more flexibility.

Figure 74: Space filling diagrams of the atoms showing interactions near the 82nd residue in active site cavity of HIV-1 protease. **Top (left and right)**: In the V82F complex, CH-π interactions are disrupted as the F82/F1082 are pointing away from the P1/P1' phenyl rings of ritonavir. **Bottom (left and right)**: In the WT complex, CH-π interactions are present between γ-methyl groups of V82/V1081 residues with the P1/P1' phenyl rings of ritonavir.

Figure 75: Superposition of docked V82F/I84V/ritonavir complex model (green sticks) with the V82F/ritonavir (yellow sticks) complex. The drug molecule in V82F mutant is not shown for clarity.

Figure 76: The interactions of the three hydrophobic residues I84, L23 and L10 with the 82nd residue of HIV-1 protease are shown after superposition of the WT (blue sticks) and V82F (yellow sticks) complexes. Note the shifts in the backbone and different rotameric conformations of L10.

Figure 77: The space filling model of V82F/L10I double mutant showing reduced interactions with F82 and L23 when L10 (green sticks) is mutated to I10 (cyan sticks).

Figure 78: In the M36I (yellow sticks) and V82F (green sticks) complexes the environment around **a)** M36 and **b)** M1036 residues are shown. The hydrogen bonding interactions are shown only for M36I mutant. The water molecules in the V82F complex are shown as green spheres.

List of Tables

CHAPTER 1

HIV-1 Protease-A Target for Drug Design against AIDS

1.1 History of Human Immunodeficiency Virus and AIDS

Human Immunodeficiency Virus (HIV) is the causative agent of Acquired Immune Deficiency Syndrome (AIDS). More than 40 million people all over the world are infected by the virus. It was first isolated in 1983 from a patient with a condition linked to AIDS [1] and was then identified to be a retrovirus. In 1984, three research groups isolated the AIDS virus and named the virus separately as Human T-lymphotrophic virus [2], Lymphadenopathy-associated virus [3] and AIDS associated retrovirus [4]. In 1986, the AIDS virus was rechristened 'Human Immunodeficiency Virus' (HIV). Luc Montagnier and his colleague Françoise Barré-Sinoussi, from France were awarded 2/3rd of the 2008 Nobel Prize in Physiology or Medicine for their "discovery of human immunodeficiency virus". After the discovery of HIV-1 in 1983, a separate strain, HIV-2 was identified in West Africa [5]. HIV-2 is less infectious and the disease develops over a longer period. HIV-1 is the common type of HIV. HIV-1 is thought to be of zoonotic origin and the first transmission from chimpanzee to people is placed at around 80 years ago in Central Africa [6, 7]. Accordingly, these transmitted viruses differentiated in human beings into three groups: M (main), O (outlier), and N (non-M,

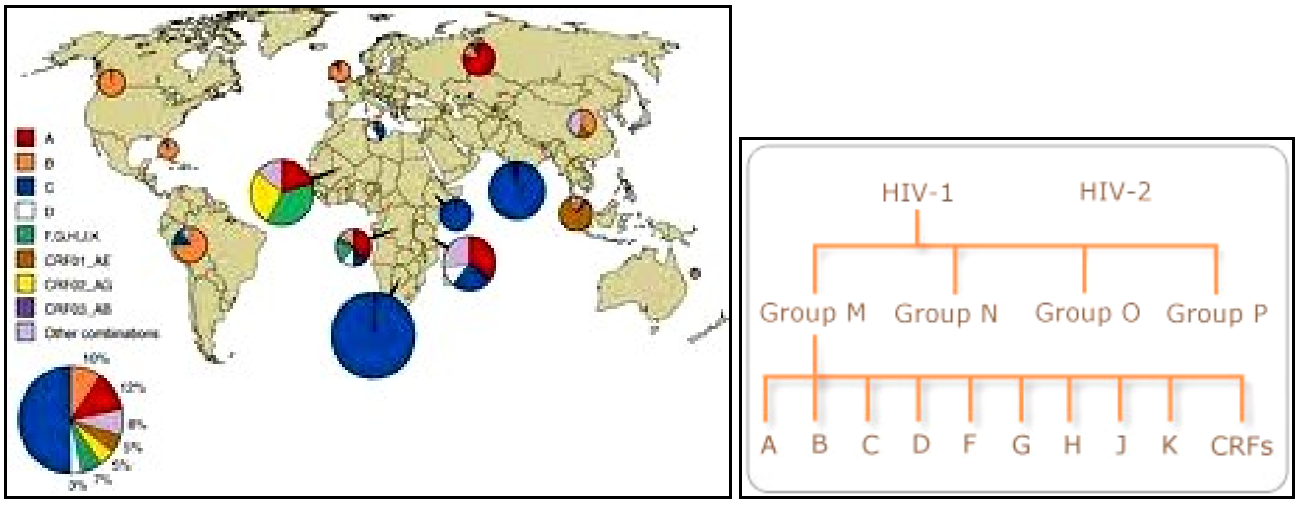


Figure 1: Geographical distribution (left) and clades (right) of HIV-1 genetic forms.

non-O) [8-10]. The last two groups remain essentially restricted to West Africa, whereas the M group is responsible for majority of infections worldwide. In 2009, a new strain closely relating to gorilla simian immunodeficiency virus was discovered in a Cameroonian woman, designated HIV-1 group P [11]. The M-group viruses have nine genetic subtypes identified by the letters A–D, F–H, J and K. Geographical distributions and clades of HIV-1 genetic forms are shown in Figure 1. Within A and F subtypes, separate subclusters are distinguished, designated by sub-subtypes A1 and A2, and F1 and

F2, respectively. Several inter-subtype hybrid strains appear which are referred to as circulating recombinant forms (CRFs). There are 14 CRFs that have been identified till now [12]. However, with the increasing number of gene sequence characterization of HIV-1, this number is likely to increase in future.

1.2 The Human Immunodeficiency Virus

HIV-1 belongs to *lentivirus* genus of retroviruses family. The primary target of HIV-1 is the T-lymphocytes, which play key role in the regulation of immune response [13]. As a result of HIV-1 infection the immune system is weakened resulting in onset of opportunistic bacterial, viral and fungal infections and of certain types of cancers. HIV-1 is spherical in shape having glycoprotein spikes on its surface and two positive sense single stranded RNA molecules as its genetic material in its bullet shaped core [14]. The isolation of HIV-1 and sequencing of its entire genome was done [15, 16] resulting in the recombinant expression and functional characterization of most viral proteins. It is possible to check the viral life cycle which can be exploited in the development of drugs for AIDS [17-19].

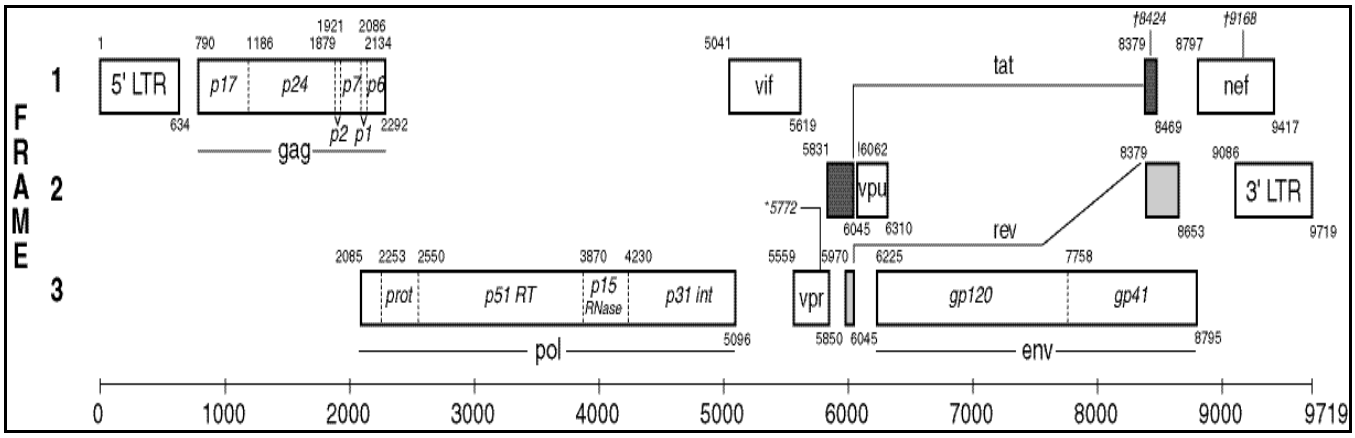


Figure 2: The HIV-1 genome.

HIV-1 has three genes called group specific antigen (*gag*), polymerase (*pol*) and envelope (*env*) (Figure 2) [20]. The *gag* gene encodes a 55 kDa *gag* polyprotein that is proteolytically cleaved by a virally encoded protease (PR) during maturation into three discrete proteins, MA, CA and NC and two spacer peptides, p1 and p6 (Figure 2). The matrix has role in the import of the HIV-1 genetic material into the nucleus of the infected cell in the early stage of infection and in export in the late stage of infection by directing new HIV-1 proteins towards the cell membrane where new viruses are assembled [21]. The nucleocapsid plays a role in the packaging of the viral RNA [22]. The two copies

of the unspliced viral RNA in the capsid are stabilized as ribonucleoprotein complexes. The *pol* gene encodes the viral enzymes necessary for replication, namely the protease (PR), reverse transcriptase (RT) and integrase (IN) and are associated with the ribonucleoprotein complex. The *env* gene encodes a precursor protein gp160, which is cleaved into gp120 (SU) and gp41 (TM) by a cellular enzyme.

Six additional genes encode the regulatory proteins (Tat and Rev) [23-27] and the accessory proteins (Vpu, Vpr, Vif and Nef) of HIV-1 [28-33].

1.2.1 HIV-1 life cycle

Like other viruses, HIV-1 requires a host cell to replicate. HIV-1 infects primarily vital cells in the human immune system such as CD4+ T cells, macrophages and dendritic cells [34]. There are two stages in the life cycle of HIV-1:

1) Early Phase: This involves target cell recognition and entry into the host cell and the processes leading to integration of HIV-1 genome into the host chromosome.

2) Late Phase: This involves the processes from the regulated expression of the integrated proviral genome upon activation of the infected cell to budding and virion maturation.

HIV-1 infection begins with the binding of the trimeric envelope complex gp160 (envelope protein) to the CD4 receptor on the T-cell surface (Figure 3) [19]. It has been found that although CD4 is necessary for HIV-1 infection, it is not by itself sufficient to make a cell susceptible to HIV-1 [35, 36]. On binding to CD4, gp120 undergoes conformational changes that enable it to bind to the chemokine receptors (CXCR4 and CCR5) which serve as viral co-receptors [37, 38]. Once both the cell surface ligands (CD4 and the co-receptors, CXCR4 and CCR5) are bound by gp120, the transmembrane protein gp41 undergoes a conformational change that facilitates the N-terminal fusion peptide of gp41 to penetrate the cell membrane. Repeat sequences in gp41, HR1, and HR2 then interact, causing the collapse of the extracellular portion of gp41 into a hairpin. This loop structure brings the virus and cell membranes close together, allowing fusion of the membranes and subsequent entry of the viral capsid [39]. On entry the core is uncoated and RNA is converted into DNA by its own enzyme reverse transcriptase [40, 41]. This DNA is then transported to the nucleus as part of a pre-integration complex that includes certain viral as well as cellular proteins [42]. The HIV-1 DNA is then integrated into the host genome in a process that requires the viral enzyme integrase [43]. The

late phase of the HIV-1 lifecycle begins with the synthesis of unspliced and spliced mRNA transcripts which are then transported from the nucleus to the cytoplasm [44, 45]. Initially, the multi-spliced RNAs that encode the regulatory proteins are made. The single spliced RNAs encode the accessory proteins and the *env* proteins, whereas the unspliced mRNA code for *gag* and *pol* precursor proteins.

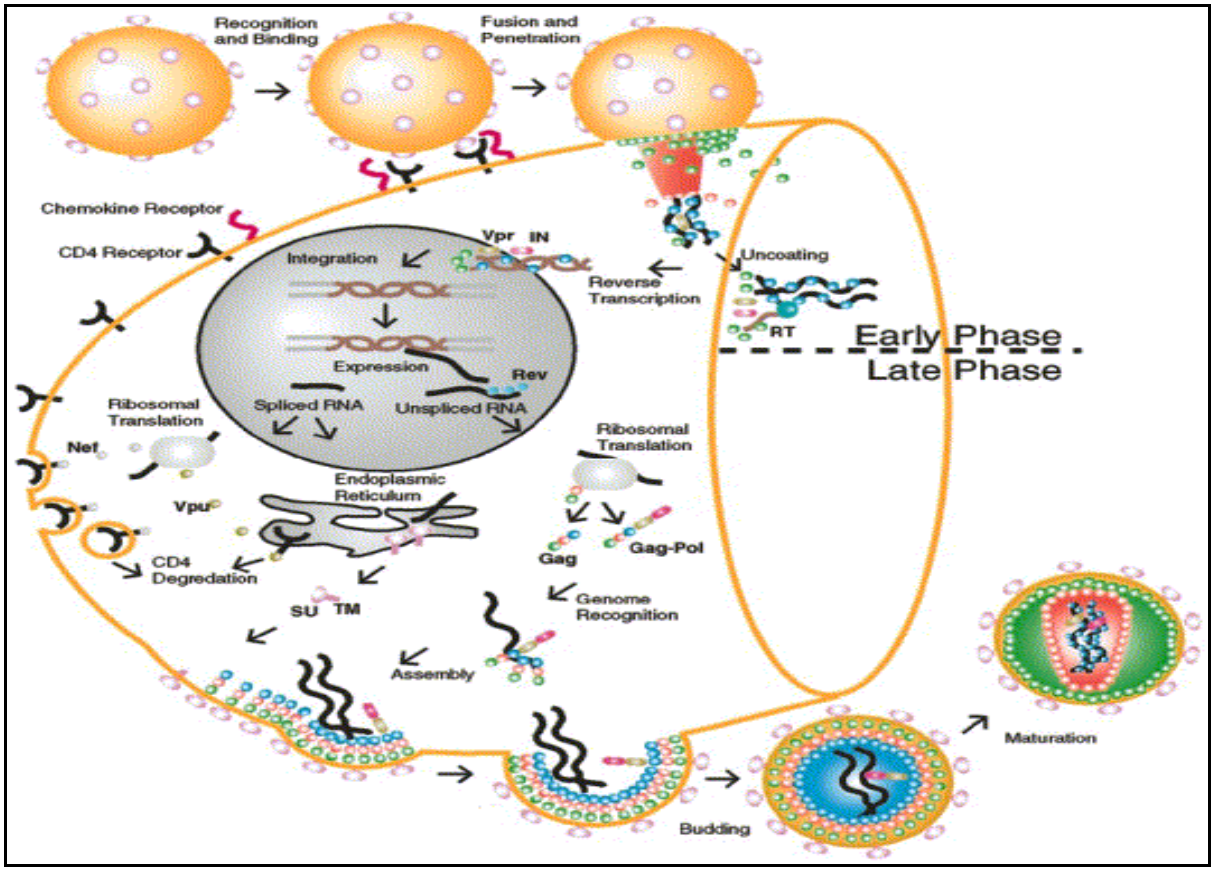


Figure 3: General features of the HIV-1 replication cycle.

The *gag, pol* and *env* gene products are initially expressed as precursors $Pr55^{gag}$, $Pr160^{gag\text{-}pol}$ and gp160 respectively. The $Pr160^{gag\text{-}pol}$ precursor results from a ribosomal frameshift and read through during translation of the *gag* gene [46]. The *gag* and *gag-pol* gene products in assembled immature virions are found in the ratio of 20:1, which represents the frequency of ribosomal frame shifting which is about 5 % [47]. The frame shifting is used as a regulatory mechanism to ensure that large number of variations in the viral proteins. The viral assembly takes place at the cell membrane, where extra-cellular budding of virions results in the acquisition of viral envelope proteins required for receptor recognition and fusion. The virion at this stage is still immature and non-infectious. During maturation, HIV-1 protease cleaves the polyproteins into individual structural proteins and enzymes. The various structural components then assemble to produce a mature HIV-1 virion [48]. The mature

HIV-1 virion is then capable of infecting a new cell starting new round of replication [23] as shown in Figure 3.

1.2.2 Challenges in vaccine development against AIDS

A general consensus is that the development of an effective vaccine is the most economical way to tackle the AIDS epidemic. The are several trial vaccine candidates which include DNA Vaccines, Live Attenuated Virus Vaccines, Virus-like Particle Vaccines, Whole (killed) Inactivated Virus Vaccines, etc. and based on these templates, a large number of studies on HIV-1 and SIV-related vaccines are being generated [49-52].

Researchers are working to optimize the HIV-1 fragments, known as antigens, which can be used to computationally design antigens to deal with the overwhelming genetic diversity of HIV-1. These antigens, called *mosaics*, have only been tested in animal models so far, but there are now plans for three clinical trials evaluating mosaic antigens in Phase I clinical trials in the next couple of years [53]. Another area of recent progress in the HIV-1 vaccine field is the discovery of several antibodies that can neutralize a remarkably high percentage of virus strains in laboratory tests [54]. These broadly neutralizing antibodies (*bNAbs*) were reported along with several new antibodies that were isolated from HIV-infected individuals [55, 56]. More attention has also been directed toward understanding another type of antibody function; instead of neutralizing the virus by binding directly to it, the antibody binds to cells already infected with HIV-1, thus facilitating the killing of these cells by other immune cells [57]. There are reports on new approaches to design vaccine antigens that could coax the immune system to produce such *bNAbs*. One method for designing these antigens involves stitching the precise part of HIV-1 to which the *bNAbs* binds into a computationally designed protein structure. This method, called *scaffolding*, is a promising avenue of work in vaccine development [58]. The rational design and development of a safe, effective, and affordable HIV-1 vaccine still remains a formidable scientific and public health challenge at the dawn of this century [52, 59, 60]. Unfortunately, the effort to develop a good and reliable vaccine against HIV-1 has proven to be difficult.

1.2.3 Possibility of Gene Therapy

Gene therapy offers the promise of preventing progressive HIV-1 infection by sustained interference with viral replication. Gene-targeting strategies are being developed with RNA-based agents, such as

ribozymes, antisense RNA, RNA aptamers and small interfering RNA, and protein-based agents, such as the mutant HIV Rev protein M10, intrakines, intrabodies, fusion inhibitors and zinc-finger nucleases along with recent advances in T-cell based strategies which include gene-modified HIV-1-resistant T-cells, lentiviral gene delivery, CD8+ T cells, T-bodies and engineered T-cell receptors [61-71]. HIV-1-resistant hematopoietic stem cells can be designed which has the potential to protect all cell types susceptible to HIV-1 infection [72]. Till now, treatment regimens primarily target the virus enzymes or virus-cell fusion, but not the integrated provirus. HIV-1 integrates into the host chromosome and persists as a provirus flanked by long terminal repeats (LTRs). A tailored recombinase has been developed that recognizes an asymmetric sequence within an HIV-1 LTR. This evolved recombinase efficiently excised integrated HIV-1 proviral DNA from the genome of infected cells [73]. The emergence of viral resistance can be addressed by therapies that use combinations of genetic agents and that inhibit both viral and host targets.

1.3 Targets in AIDS therapy

The HIV-1 replicates utilizing the host cell machinery. HIV-1 has very high genetic variability eluding host defense as well as drugs. The current strategy for anti-viral therapy is to selectively inhibit the processes that are essential for the replication of the virus, but not for the survival of the cell. The drugs are targeted against the following steps in the life cycle of HIV-1: 1) viral adsorption, through binding to the viral envelope protein gp120 [35, 36], 2) virus-host cell fusion through envelope protein gp41 and co-receptors and is internalized [39], 3) uncoating of the viral core with the help of host TRIM5α protein [74], 4) reverse transcription, 5) integration into host chromosomal DNA by integrase [43], 6) expression of early viral proteins from multiply spliced mRNAs, 7) expression of late mRNAs encoding the structural proteins *env, gag, pol* [75] and 8) packaging of unspliced genomic RNA and release of viral particles with protease activity [23, 75]. The numbers of practical targets for drug are limited due to the fact that agents that block the replication of the virus can be lethal to the host cell. Currently approved drugs for AIDS therapy are shown in Table 1. Although all these drugs delay the progression of the disease, they do not prevent it, as infection readily leads to drug-resistant mutants. Recently developed "drug cocktails", more popularly known as HAART (Highly Active Anti Retroviral Therapy), that contains combinations of PR and RT

inhibitors can reduce viral loads to undetectable levels, and these low levels can be maintained for periods of two years or more [76, 77]. One fusion inhibitor, Enfuvirtide was approved by US FDA which binds to a region of gp41 and prevents the conformational change necessary for fusion of HIV-1 to the CD4+ cell [78]. Of recent, few US FDA approved inhibitors targeted against HIV-1 integrase are currently available in the market. Several inhibitor affecting uncoating, transcription and translation are at different stages of development. The most serious problem is that a pool of virus is maintained in reservoirs within the body that are not accessible to the current drugs [79-81].

Table 1: FDA approved drugs for AIDS therapy.

Fusion Inhibitors (gp41)	Enfuvirtide
Integrase Inhibitor	Raltegravir, Elvitegravir, MK-2048, GSK-572
Nonnucleoside Reverse Transcriptase Inhibitors (NNRTIs)	Efavirenz, Nevirapine, Etravirine, Rilpivirine, Lersivirine
Nucleoside/Nucleotide Reverse Transcriptase Inhibitors (NRTIs)	Zidovudine, Didanosine, Zalcitabine, Stavudine, Lamivudine, Abacavir, Emtricitabine, Apricitabine, Tenofovir
Protease Inhibitors (PIs)	Amprenavir, Atazanavir, Fosamprenavir, Indinavir, Lopinavir, Nelfinavir, Ritonavir, Saquinavir, Tipranavir, Darunavir
CCR5 Inhibitor	Maraviroc

1.4 Role of the HIV-1 protease in the viral life cycle

A protease is an enzyme that cleaves proteins to their component peptides. The HIV-1 protease hydrolyzes viral polyproteins into functional protein products that are essential for viral assembly and subsequent activity [82-84]. This maturation process occurs as the virion buds from the host cell [23]. As already mentioned in the section **1.2.1**, the structural and functional proteins of HIV-1 are initially expressed as $Pr55^{gag}$ and $Pr160^{gag\text{-}pol}$ precursor proteins, which assemble to form the immature virion. The protease catalyzes its own release from the *gag-pol* polyprotein in addition to the highly ordered and sequential cleavage of the *gag* and the *gag-pol* proteins [82]. For the AIDS virus to mature, HIV-1 protease must process these assembled polyproteins specifically at nine non-homologous sites, [84] which are shown in Figure 4. The processed products of HIV-1 protease include the *gag* structural proteins and the *pol* enzymes. All these products are found in the mature virus particles [17] and deletion mutagenesis of HIV-1 protease gene resulted in the production of non-infectious virus particles with an immature morphology [85]. These results were confirmed by mutation of the active

site aspartates and subsequently by chemical inhibition with inhibitors [86, 87]. Hence, HIV-1 protease is essential for the life cycle of HIV-1 and, therefore, makes an important target for the design of antiviral agents against AIDS.

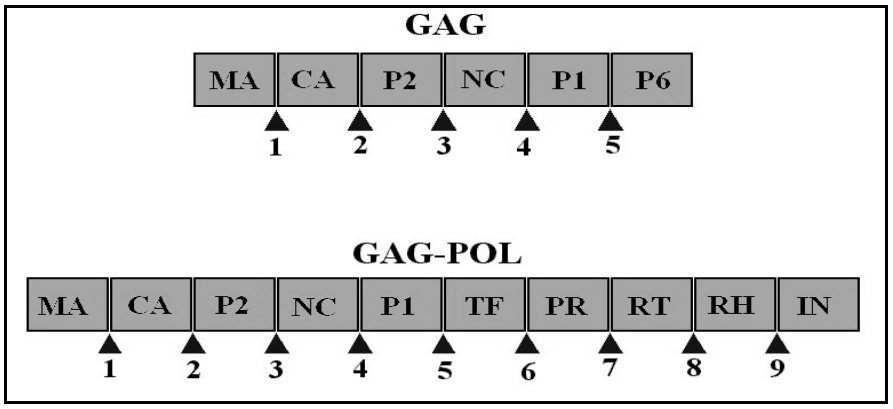


Figure 4: The HIV-1 *gag* and *gag-pol* cleavage domains.

1.5 Structure of HIV-1 protease

HIV-1 protease gene has been cloned in a variety of vectors [88-90] and has also been prepared by total chemical synthesis [91]. Crystal structure of protease reported independently in 1989, by several laboratories [92-94], revealed it to be a homodimer made up of two identical polypeptide chains of 99 amino acids each. Spatially the two chains are related by a twofold axis of symmetry (Figure 5). The secondary structure of each monomer includes one short α-helix and two approximate antiparallel β-sheets. The dimer is stabilized by non-covalent interactions, hydrophobic packing of side chains and hydrogen bonding interactions involving the catalytic aspartates. The major contribution to the dimerization is by hydrogen bonds in the four-stranded antiparallel β-sheet involving N- and C-terminal β-strands form each monomer. The enzyme active site is formed at the interface of the two monomers and comprises of two catalytic aspartates, one from each monomer (Asp25 and Asp1025), at the catalytic centre. The carboxylate groups of aspartic acid are nearly coplanar and form a hydrogen bond. Many aliphatic side-chain residues from each monomer form a hydrophobic pocket of the active site. Each monomer contains two cysteine residues but without any disulfide bridges. Structure of the HIV-1 protease is shown in Figure 5 with the principal elements of secondary structure labeled as ***a-d***, ***a'-d'***, ***h'***and ***q***. For each monomer, the N-terminal β-strand ***a*** (residues 1-4) continues, through a turn, into β-strand ***b*** (residues 9-15) which after a small turn is followed by β-strand ***c*** (residues 18-23) and then the active-site β-hairpin loop consisting of Asp25-Thr26-Gly27.

The active-site loop is followed by β-strand ***d*** (residues 30-35), separated from β-strand ***a'*** by a broad loop (residues 36-42) containing the flap elbow (residues 38-42). The flap region consists of two β-strands, ***a'*** and ***b'*** (residues 43-49 and 52-58 respectively) and participates in the binding of inhibitors and substrates. Residues 49-51 connect these two β-strands is called the flap tip. This flap

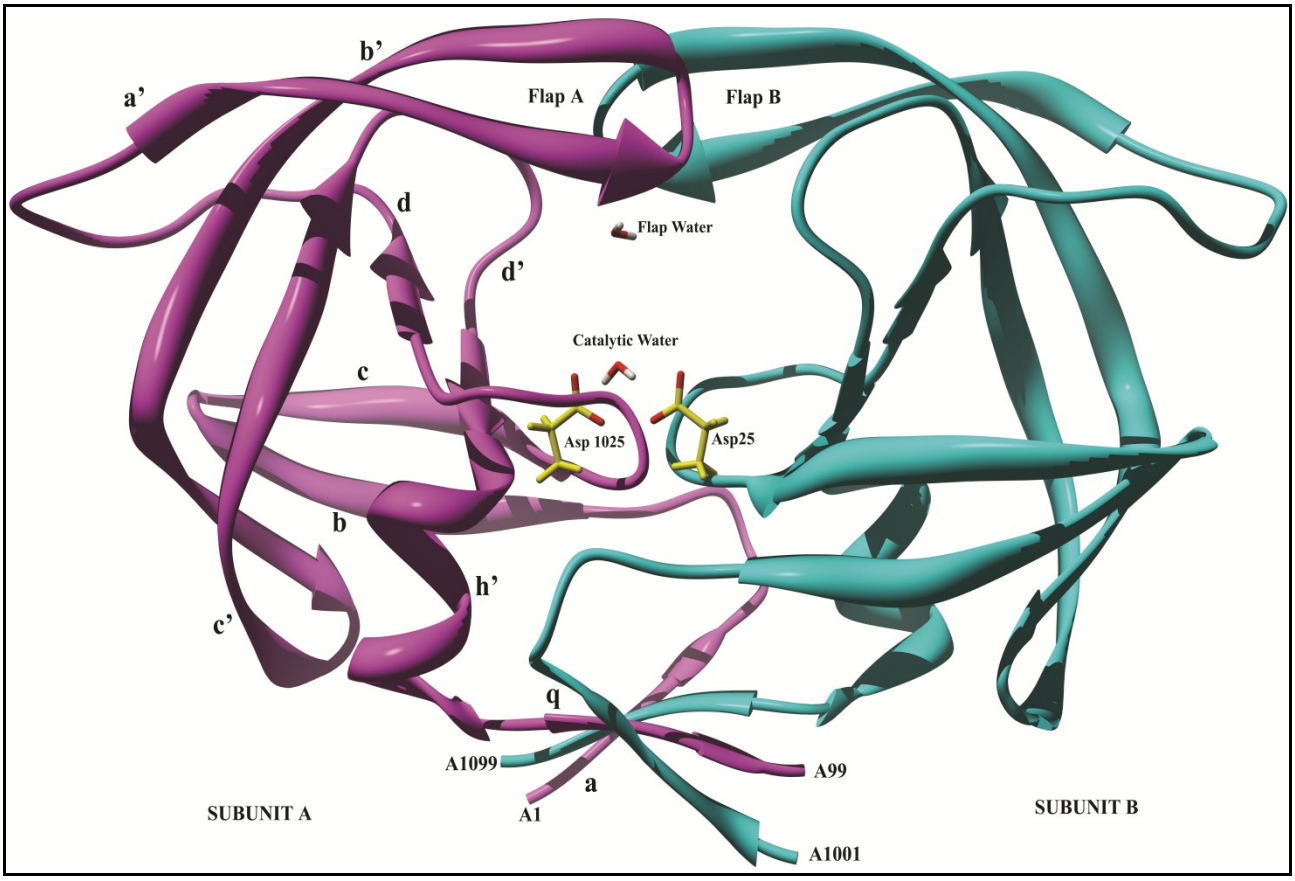


Figure 5: Ribbon diagram showing the secondary structural elements of HIV-1 protease in closed-flap conformation. The two catalytic aspartic acid residues (yellow sticks) are hydrogen bonded to the catalytic water molecule. A structural water molecule anchors the flaps in closed conformation, which otherwise is labile when the flaps are open.

remain flexible and allows for hinge-like mobility. It allows substrate/inhibitor access to the active site by opening and curling the tips into hydrophobic pockets. The flap tips open up by more than 7 Å in apo-enzyme and, when bound to ligands, close in forming a single inter-monomeric hydrogen bond. β-strand ***c'*** (residues 69-78) follows and continues to strand ***d'*** (residues 83-85) after a short loop at residues 79-82. The only helix ***h'*** (residues 86-94) has hydrogen-bonding pattern that is intermediate between an α-helix and a 3_{10} helix. The C-terminus ends with a β-strand ***q*** (residues 95-99). Asp25 or Asp1025 (hereafter will be represented by D25 or D1025 respectively) residues from each monomer bind to a water molecule through hydrogen bonds. In all earlier crystal structures of unliganded HIV-1 protease, these flaps are in open conformation. The closed flap structure of unliganded protease was reported earlier [95].

1.6 Enzymes

Though enzymes catalyse thousands of chemical reactions at any given instant within all living cells, the mechanisms are not completely understood. Enzymes can catalyze thermodynamically favorable reactions by reducing the activation barrier (Figure 6). However, how this reduction is accomplised is not understood at the atomic level. Structural data on complexes between active enzyme and true substrate are needed to fill this void.

All enzymes now are classified into six classes and formally named according to the reaction they catalyze. New classes are observed only recently (Ribozymes and Deoxyribozymes) [96, 97]. Thus HIV-1 protease belongs to the class hydrolases, subclass peptide hydrolase, sub-subclass aspartyl protease and so is designated as ***3.4.23.16***.

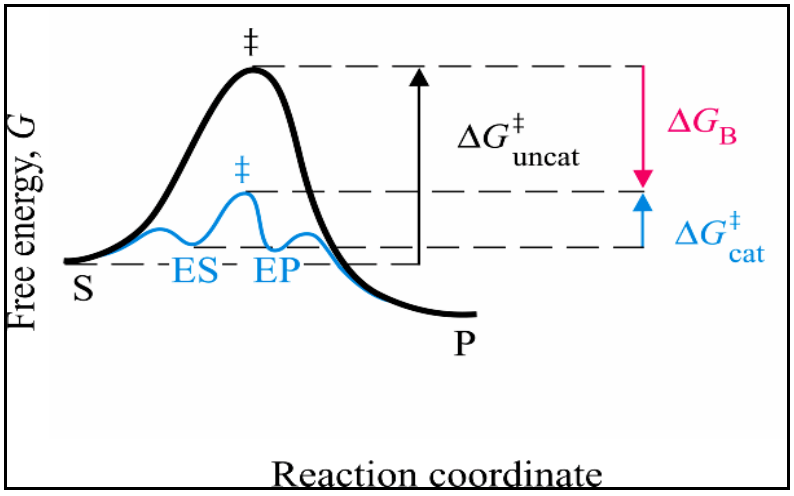


Figure 6: Reaction profile showing large reduction in free energy, ΔG_B ($=\Delta G^{‡}_{uncat} - \Delta G^{‡}_{cat}$) between uncatalyzed and enzyme catalyzed reaction. This free energy change caused by catalysts lowers ΔG_B thereby accelerating rate.

Enzymes are characterized by three distinctive features: catalytic power, specificity and regulation [98]. Figure 7 below provides a chronological order in which different theories evolved, to explain enzyme catalysis [99-106].

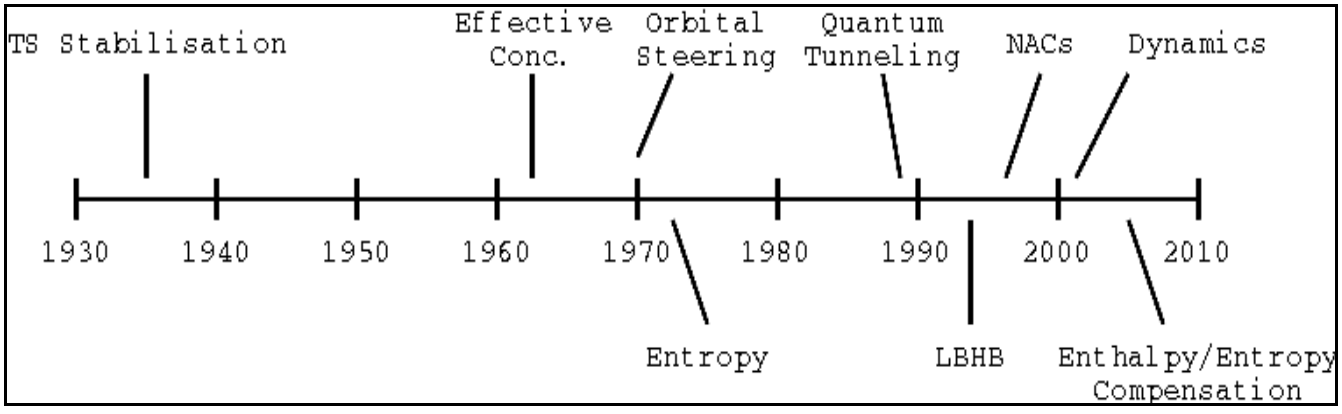


Figure 7: Chronology of the theories proposed to explain enzyme catalysis.

1.7 HIV-1 protease is an aspartyl protease

Aspartic proteases are one of the four major classes of proteolytic enzymes, the other three being the serine proteases, cysteine proteases and metalloproteases. Aspartic proteases are a broad subfamily of proteolytic enzymes [107] which catalyze the specific peptide bond hydrolysis in oligopeptides or proteins. This reaction, also involving eukaryotic proteases, is a key process in a number of serious diseases: hypertension (Renin) [108], AIDS (HIV protease) [17], malaria (Plasmepsins) [109], breast cancer (Cathepsin D) [110], and the Alzheimer disease (β-Secretase and Presenilin) [111]. HIV-1 protease belongs to the family of aspartic proteases which was based on the presence of aspartic protease signature sequence Asp-Thr/Ser-Gly in retroviral proteases [112, 113]. Other features of HIV-1 protease, which are typical of aspartic proteases, are inhibition by pepstatin [86], and mutation of the active site aspartates resulting in loss of enzyme activity [113-115]. Again these enzymes, which have a pair of catalytic aspartic acid residues, are classified into two major groups: retroviral proteases and the cellular aspartic proteases [116]. The cellular aspartic proteases, like pepsin, renin, chymosin and cathepsin D, are single chain molecules in mature form [117]. Crystal structures of cellular aspartic proteases revealed that they are made of two similar domains, named as N- and C-terminal domain. The most significant differences in the structures of retroviral and cellular proteases are attributed to the symmetrical arrangement in the dimeric molecule of the retroviral protease. The two halves of the molecules are exactly or approximately the same, as opposed to the single chain molecule of cellular aspartic protease, two domains of which, although topologically similar, are not identical [117]. Crystal structures of proteases from several other retroviruses have been determined [118-124].

1.8 Inhibitor and substrate complexes of HIV-1 protease

Several structures of various inhibitor complexes of HIV-1 protease have been reported in various crystal forms [125]. In most of these structures the inhibitors are peptidomimetic, although some nonpeptidic inhibitors have also been studied crystallographically [86, 126, 127]. It was initially assumed that binding of an inhibitor introduces substantial conformational changes to the enzyme that could be described as a rotation of the two monomers by up to 2° in opposite directions around a hinge axis located near the β-sheet structure of the dimer interface. This motion, which slightly

tightens the cavity of the active site, is also accompanied by a very large motion of the flap region – as much as 7 Å for the tips of the flap [128]. The interactions between peptidomimetic inhibitors and HIV-1 protease are very similar for all of the complexes. The hydrogen bonds are formed between the main-chain atoms of both the enzyme and the inhibitor, and follow a similar pattern. These interactions are not sequence specific, and most of the time they dominate the total binding energy [129]. The non-hydrolyzable scissile bond analog of each inhibitor was found in close vicinity to the carboxyl groups of the active site aspartates D25/D1025. The hydroxyl group at the non-scissile junction, present in inhibitors other than those containing the reduced peptide bond isosteres, is positioned between the protease aspartate carboxyl groups, within hydrogen bonding distance to at least one carboxylate oxygen of each aspartate. The enzyme substrate complexation has been obtained only when the enzyme is inactive D25N mutant [130]. A feature common to almost all complexes of HIV-1 protease is a water molecule that bridges the carbonyl groups at P2 and P1' of the inhibitor and the I50 (or I1050) NH groups of the flaps. This water is approximately tetrahedrally coordinated and is completely inaccessible to the solvent [131, 132] and is highly specific for retroviral proteases, has no equivalence in cellular aspartic proteases.

A number of distinct subsites that accommodate the side chains of the substrates (or inhibitors) can be identified in HIV-1 protease. These subsites are usually described using the nomenclature of Schechter and Berger [133], in which subsites on the N-terminal side of the scissile bond (or its nonscissile replacement) are unprimed and those on the C-terminal side are primed. The two residues that are immediately adjacent to the scissile bond are therefore labeled P1 and P1'; the next two P2 and P2'; and so on. Their corresponding binding pockets in the enzyme are labeled S1 and S1', S2 and S2', and so on. In all the retroviral proteases, the primed and unprimed sites are formed by identical residues, due to the symmetry of these enzymes. Some of the enzyme residues are also part of more than a single subsite, since the pockets S4, S2, S1' and S3' are located on flap-side of the active site, whereas the pockets S3, S1, S2' and S4' are located in a similar manner on the other side of the active site. Three subsites (S1–S3) are very well defined, whereas the more distant subsites are not as distinct.

The HIV-1 protease side chains comprising pockets S1 and S1', with the exception of the active site aspartates, are mostly hydrophobic. The residues, which contact the P1 or P1' side chains of the

substrates (or inhibitors), include R8, L23, D25, G27, G48, G49, I50, T80, P81 and V82. Almost all of the documented inhibitors have hydrophobic moieties at P1 and P1' with the exception of the statine- and glycine-containing inhibitors, in which no groups occupy the protease subsite S1'. Subsites S2 and S2' are interior pockets and are smaller than the S1/S1' or S3/S3' binding sites. They were also shown to be more specific, restricting the size and the type of the residue at P2/P2' in the substrates or inhibitors relative to the other binding pockets in the HIV-1 proteases [134, 135]. In HIV-1 protease, S2 and S2' are formed by A28, D29, D30, V32, I47, G49, I50, L76 and I84. Although it is clear that these pockets are at least partially hydrophobic, both hydrophilic and hydrophobic residues of substrates/inhibitors have been found to occupy the individual sites [126, 130]. Subsites S3 and S3' are known to have a rather broad specificity, being able to accept residues of different types and sizes [126, 134]. This phenomenon is due to the variation of the residues forming them and to their ability to maintain aliphatic, polar and ionic interactions with different ligands. In addition, these subsites are quite large and exposed to the solvent and are formed by R8, L23, D29, G48, G49, P81 and V82. G48, P81 and V82 also form parts of the S1/S1' pockets mentioned earlier. These subsites can accommodate a variety of side chains, some of which are very large (for example, naphthylalanine in inhibitors such as LP-130 [136], LP-149 [137] and HBY-793 [138]). Subsites S4/S4' and S5/S5' have not been described in comparable detail, due to the availability of only few structures where ligands extend beyond P3/P3' [139]. The structures of HIV-1 protease and its complexes have been catalogued [140].

1.9 HIV-1 protease catalysis – the mechanistic proposals

HIV-1 protease is both structurally and biochemically a well characterized enzyme. Various mechanisms for the reaction catalyzed by aspartic proteases, including HIV-1 protease, have been proposed based on the kinetic, biochemical and structural data [121, 141-155]. There are also reports, where efforts were made to understand how HIV-1 protease cleaves the peptide bond, in the enzyme/substrate complexes, employing various sophisticated theoretical techniques [156-165]. Three distinctly different chemical mechanisms have been proposed for catalysis of peptide bond cleavage by aspartic proteases. In the first one, a direct nucleophilic aspartic acid side chain carboxylate attacks the carbonyl-group of the peptide bond, followed by expulsion of the amine component [145]. However, this mechanism would imply, as an intermediate, a covalently bonded

complex between the enzyme and the substrate molecule. So far, there has been no experimental evidence to support the existence of such covalent intermediate.

In the second mechanism, first proposed by Fruton [166, 167] and then followed by structural and kinetic studies [121, 147-150, 151-155], a water molecule acts as a nucleophile to attack the carbonyl carbon of the scissile peptide bond of the substrate forming a tetrahedral intermediate, which then collapses to form the two products. This general acid general base (GAGB) mechanism for peptide bond hydrolysis [141] is basically a four stage process. In this mechanism, the first stage of catalysis is the substrate bound in the active site cavity with the scissile peptide bond near the catalytic water molecule and the water hydrogen bonded to the inner oxygens of the two catalytic aspartates. In this configuration the catalytic aspartates are monoprotonated and the one (D25) closer to the substrate peptide carbonyl is protonated which acts as a general acid protonating the scissile peptide carbonyl (Figure 8). The other aspartate is anionic and abstracts a proton from the catalytic water and the resulting nucleophilic OH^- attacks the substrate peptide carbonyl to form the *gem*-diol tetrahedral intermediate (stage 2). The D1025 (shown as Asp25' in Figure 8) is now protonated and D25 (shown as Asp25 in Figure 8) is anionic. D25 acts as general base abstracting a proton from the *gem diol* to form anionic tetrahedral and both aspartates protonated (stage 3). The amide nitrogen is protonated by D1025 along with the C-N bond breakage. In the last stage two products are formed and D1025 becomes anionic.

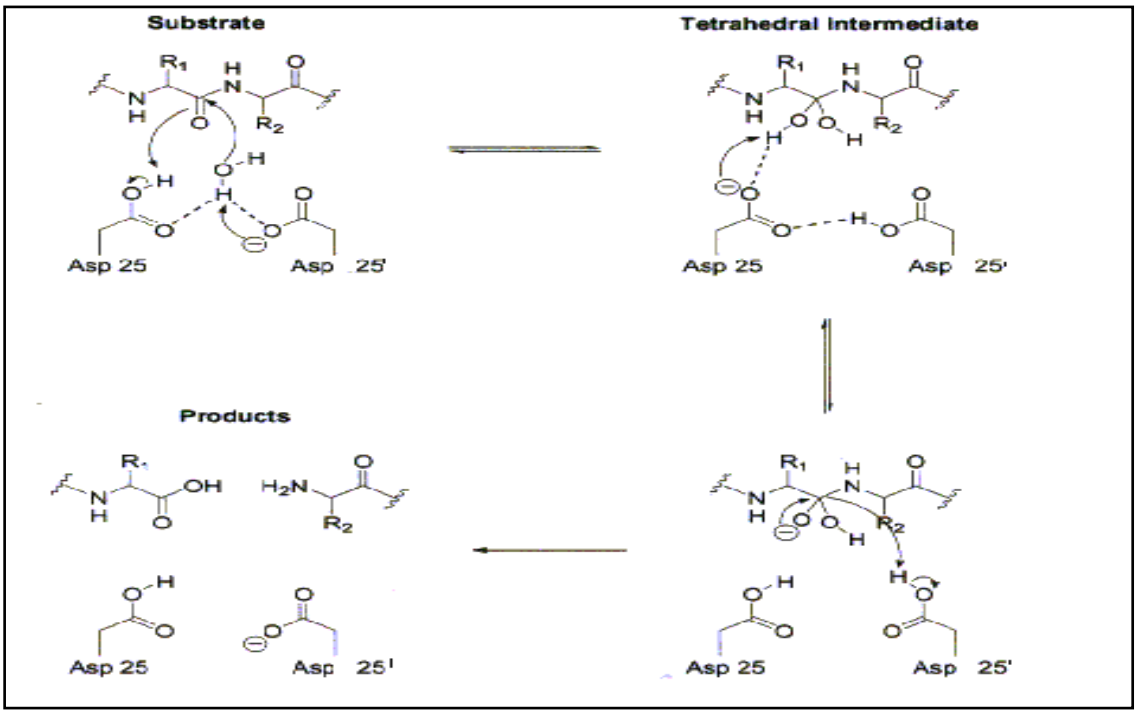


Figure 8: General acid-general base mechanism of peptide bond hydrolysis [141].

The third is a kinetic *iso*-mechanism, proposed by Northrop [150], which is a nine step process (Figure 9), starting with a planar 10-membered hydrogen bonded symmetric ring formed by the catalytic water molecule and the carboxylate/carboxylic acid of monoprotonated catalytic aspartates (state E). In the second step, substrate is bound to HIV-1 protease along with the catalytic water molecule (state ES). The flaps are open in these steps. In third step (E'S), flaps are closed and outer oxygen of catalytic aspartate accepts a proton from the catalytic water and the water attacks the scissile peptide bond with conversion of low barrier hydrogen bond (LBHB) to normal hydrogen

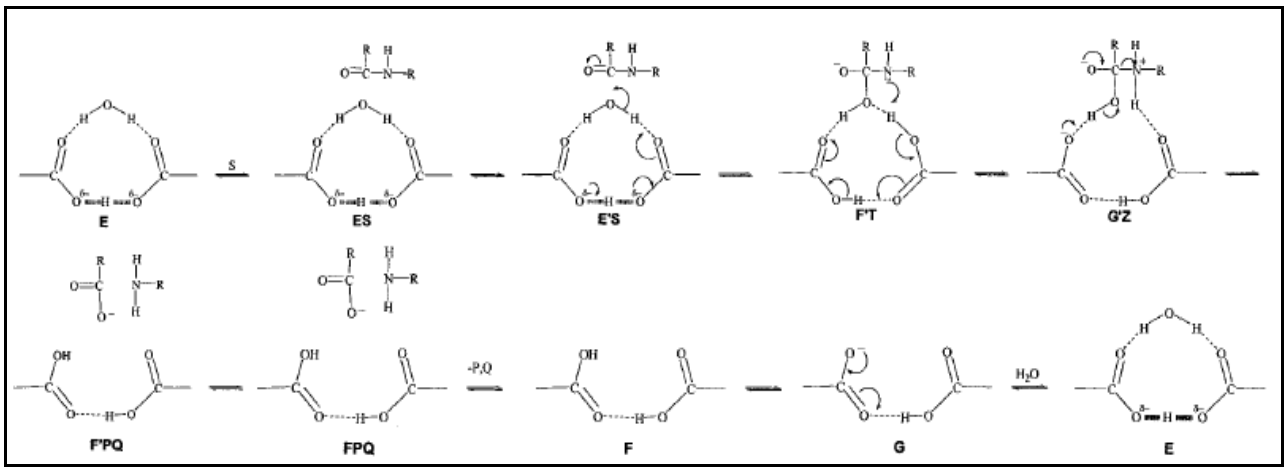


Figure 9: Chemical and kinetic *iso*-mechanism of aspartic protease proposed by Northrop [150].

bond [168]. In X-ray structures, hydrogen bonds shorter than 2.5 Å are classified as "low barrier hydrogen bond" or LBHB [168-170]. The LBHB is maintained throughout stages 1-3. The fourth stage, F'T state is an anionic tetrahedral intermediate having two protonated catalytic aspartates. Two proton transfers occur in this step, the outer oxygen of a catalytic aspartic acid donates a proton to the scissile peptide nitrogen and inner carbonyl oxygen of the same catalytic aspartic acid accepts a proton from the second catalytic aspartic acid. The fifth stage, G'Z state is a zwitterionic tetrahedral intermediate having monoprotonated aspartates with a proton transfer to the outer oxygen of catalytic aspartic acid from the hydroxyl group of tetrahedral intermediate. The distinctive feature of this transition state intermediate is a large shift in the position of N-atom of the scissile peptide bond, to make a hydrogen bonding interactions with the outer oxygen atom of one of the catalytic aspartates and concomitant peptide bond cleavage. This gives the F'PQ state where two products are formed and both the catalytic aspartates are protonated. Next state is FPQ, where the flaps are open but the products are still present in the active site. The F state represents the flap open conformation of HIV-1 protease *sans* the bi-product. Next state, G is the monoprotonated open flap conformation of

the enzyme but without the catalytic water molecule. Finally the E state is achieved with the catalytic water bound to the enzyme and LBHB between the inner oxygens of catalytic aspartates is restored. Between the states F'T to G the LBHB is absent. Northrop proposed that the diprotonated F form of the enzyme catalyses the transpeptidation reactions. The form F of the enzyme goes to the form E through deprotonation and rehydration to make a 10-atom cyclic structure. Thus the concept of isomerisation was invoked to explain the transpeptidation reaction as well as the kinetic isotope effects. The last two mechanisms invoke a low-barrier hydrogen bond (LBHB) - in the GAGB mechanism the LBHB would stabilize the transition state [165], while in the kinetic *iso*-mechanism it allows for hydrogen tunneling [150].

1.10 Rates of cleavage of substrates by HIV-1 protease

The kinetic parameters of oligopeptides corresponding to the HIV-1 protease cleavage sites have been determined [171] and are shown in Table 2 below.

Table 2: HIV-1 protease cleavage sequences with the kinetic parameters.

No.	Peptide sequences	Cleavage domain	K_m (mM)	k_{cat} (sec^{-1})	k_{cat}/K_m ($mM^{-1}. sec^{-1}$)
Cleavage sites in *gag*:					
1.	VSQNY*PIVQ	MA-CA	0.150	6.80	45.3
2.	KARVL*AEAMS	CA-p2	0.010	0.09	90.0
3.	TATIM*MQRGN	p2-NC	0.050	3.70	74.0
4.	TERQAN*FLGKI	NC-p1	not detected	not detected	not detected
5.	RPQNF*LQSRP	p1-p6	0.530	0.30	0.6
Cleavage sites in *pol*:					
6.	VSFNF*PQITL	$TFRp6^{pol}$-PR	< 0.010	0.06	6.9
7.	CTLNF*PISP	PR-RT	0.070	1.50	24.1
8.	AETF*YVDGAA	RT-RH	0.040	0.40	10.0
9.	IRKIL*FLDG	RH-IN	0.006	1.20	202.0

HIV-1 protease has a broad bell-shaped pH dependent enzymatic activity curve, between pH 3.0 to pH 7.0, with optimum activity between pH 4.5 - pH 5.5 (Figure 10) [143].

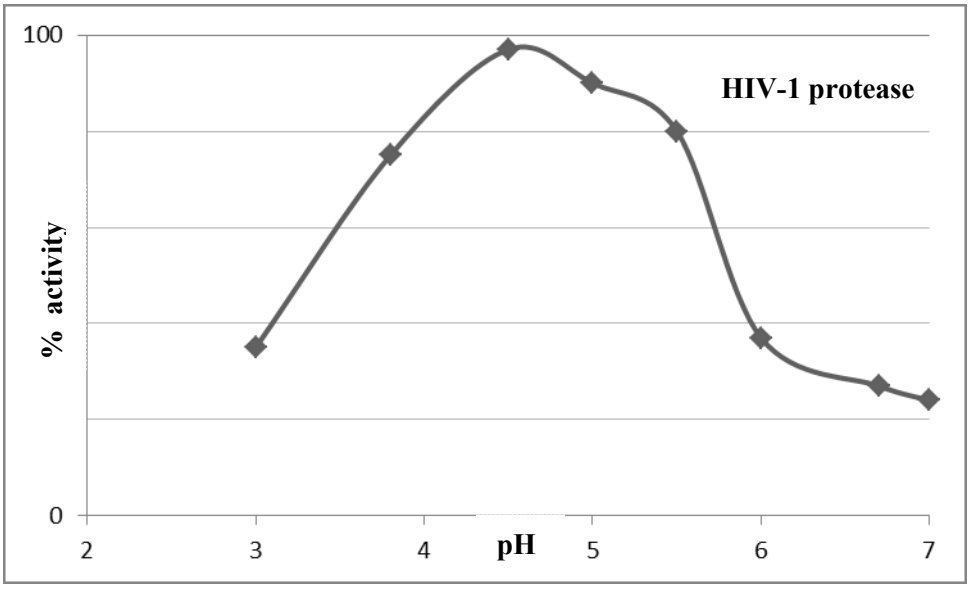


Figure 10: The pH activity profile of HIV-1 protease is bell-shaped arising from two separate active-site ionization constants [143].

1.11 Drug-resistance in HIV-1 protease

Mutant viruses emerge in the presence of antiviral agents whenever the replication of mutant virus is favorable under the selection pressure of administered drugs. Pharmaco-kinetic studies using potent HIV-1 protease inhibitors have revealed that the combined half-life of plasma virus and virus-producing cells in the body is on the order of 2 days or less, with new virus being produced at a rate of 10^{10} virions per day [172, 173]. These conditions, coupled with the high error rate of HIV-1 reverse transcriptase (aproximately 3 x 10^{-5} per nucleotide base per cycle of replication) favor rapid mutation and selection of drug-resistant virus [174]. Although many promising new anti-HIV-1 drugs have been developed, their effectiveness has been hampered by the emergence of drug-resistant variants. Clinical resistance to every newly introduced anti-HIV drugs has become a major hurdle in AIDS therapy [175-179]. Reports of the selection and transmission of multi-drug resistant HIV-1 strains that contain multiple PR and RT mutations in the *pol* gene and also in cleavage site sequences in the *gag* gene are of concern to the AIDS research community [180, 181]. In HIV-1 protease over 70 mutations at 38 positions of the 99-residue polypeptide have been observed in response to drug selection pressure [182, 183] (Figure 11). Many of these mutations were first observed *in vitro,* and presaged the emergence of mutants in the clinical setting *in vivo,* although the actual number and order of appearance of multiple mutations varied [184-186]. The mutations can be classified as active site *vs* nonactive site mutations according to whether they occur inside or outside the inhibitor binding subsites. A variety of resistance mechanisms have been proposed based on our understanding

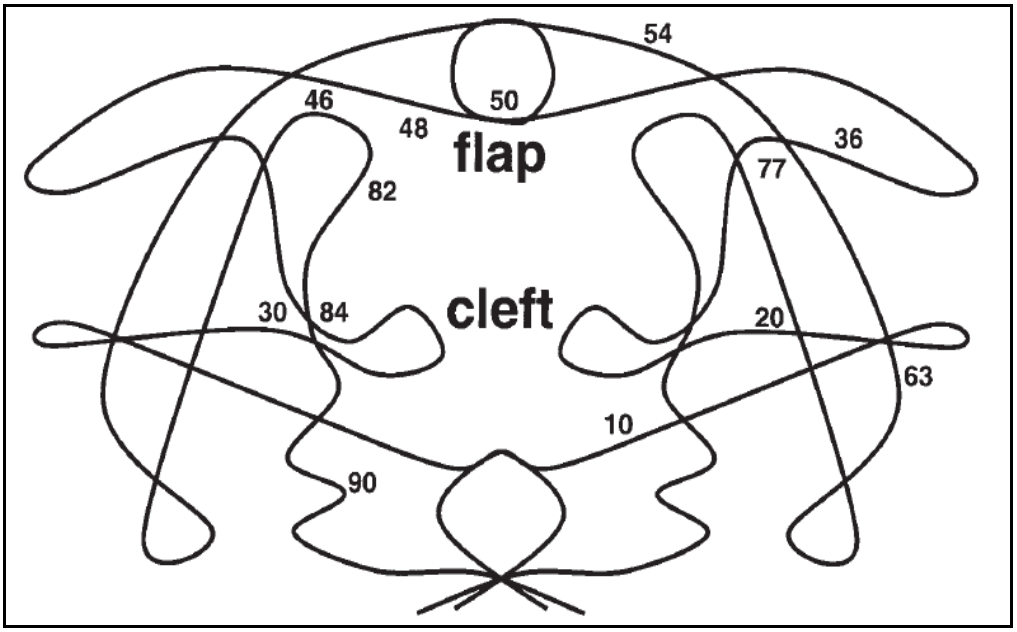


Figure 11: Some of the common mutation sites are shown on the backbone cartoon diagram of HIV-1 protease. The numbers indicate the amino acid residue number.

of the structural biochemistry of the protease and on the nature of inhibitor binding to the enzyme [175]. Mutations of specificity-determining residues that would directly interfere with inhibitor binding and lead to loss of potency constitute an obvious mechanism for resistance to HIV-1 protease inhibitors. Active site mutations are necessary but often are not sufficient for high-level resistance in the clinical setting [184]. The characteristics of the subsites are also responsible for the lower tolerance in the drug-resistant mutants of the residues forming these binding sites [184]. An explanation for this observation comes from biochemical studies that reveal a negative impact of many resistance-conferring active site mutations on enzyme activity [184, 187, 188], suggesting that such mutations result in suboptimal virus.

A second mechanism of resistance involves non-active site mutations that indirectly alter the active site architecture *via* long range structural perturbations. While the precise structural mechanism of drug resistance can often be pinpointed for active site mutations that directly affect inhibitor binding, the evaluation of non-active site mutants is more challenging. Some mutations might act in concert with active site mutations by compensating for a functional deficit caused by the latter.

Mutations that enhance enzyme catalysis in the presence of inhibitors could constitute a third mechanism. Any mutation that influences the binding of a specific inhibitor can be expected to have an effect on substrate cleavage kinetics, as well as perhaps, on substrate recognition. For this reason, combinations of two or more mutations in HIV-1 protease may lead to a variety of additive,

synergistic, or compensatory effects, depending on which property is being measured. Active site mutations can strongly affect catalytic efficiency of HIV-1 protease [188], but the magnitude of the effect also depends on the substrate sequence [189]. The addition of one or more non-active site mutations may compensate for a catalytically defective active site mutation [188, 190]. These conclusions are based on enzymology studies with recombinant HIV-1 protease mutants. Studies of individual drug treated patients demonstrate an ordered accumulation of mutations in which one or two active site mutations usually occur early and are followed by numerous non-active site mutations [185, 191]. Thus, the clinical evolution of drug resistance to protein inhibitors seems to qualitatively mirror expectations based on the enzymology studies. Since active site mutations may be expected to alter the rate of one or more cleavages that must occur during viral maturation, one may imagine that compensating mutations in the cleavage sites on the *gag* or *gag-pol* polyproteins might result in better substrates for particular mutant enzymes. Studies identified a L449F mutation in the p1/p6 Gag polyprotein cleavage site that can synergize with the I84V HIV-1 protease mutation to produce a virus with 350- to 1500-fold decreased sensitivity to substrate-based inhibitors, Bila-1906 and Bila-2185 [192]. This mutation altered the p1/p6 cleavage site from Phe-Leu to Phe-Phe. A synthetic peptide containing the mutant Phe-Phe cleavage site was cleaved at higher catalytic efficiency by the I84V HIV-1 protease mutant than the corresponding peptide with the wild-type sequence [193]. Salzman *et. al.* subsequently identified mutations at the NC/p1 Gag cleavage site in breakthrough resistant virus isolated from patients on indinavir therapy [192]. This is an important finding since it confirms the possibility for this drug resistance mechanism to be operative in the clinical setting. Several groups have followed the lead of these investigators and have confirmed the presence of cleavage site mutations in clinical isolates from multidrug-resistant HIV-1 protease in preliminary reports [194-196].

Finally, mutations that affect dimer stability, cleavage site mutations that lead to altered processing kinetics by mutant enzymes, and ‘‘regulatory’’ mutations elsewhere in the genome that lead to improved viral growth in the presence of HIV-1 protease inhibitors comprise additional resistance pathways [197,198].

1.12 Aims of present study on HIV-1 protease

In last three decades, since the first structures of HIV-1 protease was solved, this important enzyme has been well characterized, both structurally and biochemically and has been hot target against AIDS in structure-based drug design efforts. However, the drug resistant mutations in HIV-1 protease present a never ending challenge. Many of the drug resistant mutants have been well characterized and are evolving with new drugs. There is a need to understand the functioning of the active protease based on structure using true substrate as its bait. We need to solve active enzyme-true substrate 3D structures in order to understand the substrate specificity and recognition. However the basic difficulty is that the true substrate would be processed before crystals could be grown. Thus structures of six peptide substrate complexes with the inactive D25N mutant enzyme are currently available [130]. In my thesis I have selected few of the substrates and inhibitors in order to get the crystal structures of their complexes with active HIV-1 protease. **Chapter 3** and **Chapter 4** describe the binding of two different natural oligopeptide substrates to the active enzyme. These structures provide important insights to the mechanism of cleavage process and substrate binding at the molecular level. **Chapter 5** describes crystal structures of HIV-1 protease at different pHs and preparation of deuterated crystals for neutron diffraction studies. **Chapter 6** describes crystal structures of active site and non-active site drug resistant mutants complexed to Ritonavir in order to understand the mechanism behind drug-resistance.

CHAPTER 2

Macromolecular Crystallographic Method and Protein Structure Determination

2.1 Principles of X-ray crystallography

Single crystal diffraction is a very powerful method of determining accurately three dimensional structures of molecules of any size. A single crystal is a periodic array of identical unit cells repeating translationally on a three-dimensional lattice defined by the basis vectors of the unit cell. If there are atoms in the unit cell, the crystal can be regarded as made up of N interpenetrating lattices of identical basis vectors, but with origins displaced by positional vectors of each of the N-atoms.

Incident monochromatic X-rays are resonantly scattered by the oscillating free electrons of the atoms in the crystal [199]. The long range periodicity of the lattice in three dimensions leads to constructive interference resulting in non-zero intensities of the diffracted secondary X-rays only along specific directions. This is mathematically expressed as the three Laue diffraction conditions [200], which have to be satisfied simultaneously,

$$\begin{aligned} \vec{S} \bullet \vec{a} &= h \\ \vec{S} \bullet \vec{b} &= k \\ \vec{S} \bullet \vec{c} &= l \end{aligned} \tag{1}$$

These Laue conditions lead to Bragg's law in reflection geometry [201],

$$2d(hkl)\sin\theta = \lambda\,, \tag{2}$$

where $\vec{a}$, $\vec{b}$ and $\vec{c}$ are the crystal translation vectors, $\vec{S}$ is the scattering vector with magnitude $2\sin\theta/\lambda$, h, k, l are the Miller indices of the reflecting lattice planes, d is the distance between the lattice planes (hkl), θ is the angle of reflection and λ is the incident X-ray wavelength. Considering the crystal to be made up of interpenetrating lattices, formed by each atom of the unit cell occupying the origin of the respective lattice, the structure factor, $\vec{F}(hkl)$, can be written as,

$$\vec{F}(hkl) = \sum_{j=1}^{N} f_j e^{2\pi i\left(\vec{h}.\vec{r_j}\right)} \tag{3}$$

where f_j is the atomic form factor of the jth atom positioned at r_j. Treating electron density as a continuous function $\rho(xyz)$ with peaks at atomic positions r_j , the the structure factor is written as

$$\vec{F}(hkl) = V \int_{cell} \rho(xyz)\, e^{2\pi i(hx+ky+lz)} dxdydz \tag{4}$$

where V is the volume of the unit cell and ρ is the electron density distribution. Inverse Fourier transformation of equation 4 gives the electron density distribution in the unit cell,

$$\rho(xyz) = \frac{1}{V}\sum_{h}\sum_{k}\sum_{l}\vec{F}(hkl)\,e^{-2\pi i(hx+ky+lz)} \tag{5}$$

The structure factor of equation (4) is a complex quantity, and is therefore characterized by an amplititude, $\left|\vec{F}(hkl)\right|$ and a phase *φ(hkl)*.

$$\vec{F}(hkl) = \left|\vec{F}(hkl)\right| e^{i\varphi(hkl)} \tag{6}$$

The amplitudes $\left|\vec{F}(hkl)\right|$ can be derived from the experimentally measured intensities of the diffracted beam *I*(*hkl*), but the phase angles *φ*(*hkl*) cannot be obtained directly from the diffraction pattern. Without phase angle, equation 5 cannot be solved, and this is known as **'phase problem'** in crystallography [202].

2.2 Tools for X-ray diffraction

2.2.1 X-ray sources

X-radiation or X-rays is a form of electromagnetic radiation having wavelength in the range of 0.01 to 10 nanometers and energies in the range 120 eV to 120 keV. There are two types of X-ray sources used commonly by the crystallographer today, the conventional laboratory X-ray generators and the synchrotrons. A conventional laboratory source consists of an X-ray tube in which the electrons released from a cathode (tungsten filament) are accelerated in high vacuum by high applied voltage. The high velocity electrons collide with a metal target, the anode, creating the X-rays characteristic of the anode [203]. In synchrotrons (Figure 12a) charged particles (electrons or positrons) are accelerated in a storage ring at nearly the speed of light producing electromagnetic radiation by *bremsstrahlung*. Since freely traveling electrons (positrons) are not quantized, the emitted radiation ranges over a wide wavelength region, depending on the energy of the charged particles and on the strength of the magnetic field inside the storage ring [204]. The tangentially emitted radiations are used as X-ray sources at synchrotron beamlines (Figure 12b). A useful characteristic is the median of the distribution of power over the spectral region, called the critical photon energy E_c, which divides the emitted power spectrum in two equal parts,

$$E_c = 0.665E^2B \tag{7}$$

and

$$\lambda c = \frac{18.64}{E^2B} \tag{8}$$

where E_c is in *keV* and E is the circulating power in *GeV* given by,

$$E = \frac{Particle\ \ Energy \times Current}{\text{Re}volution\ \ frequency}, \tag{9}$$

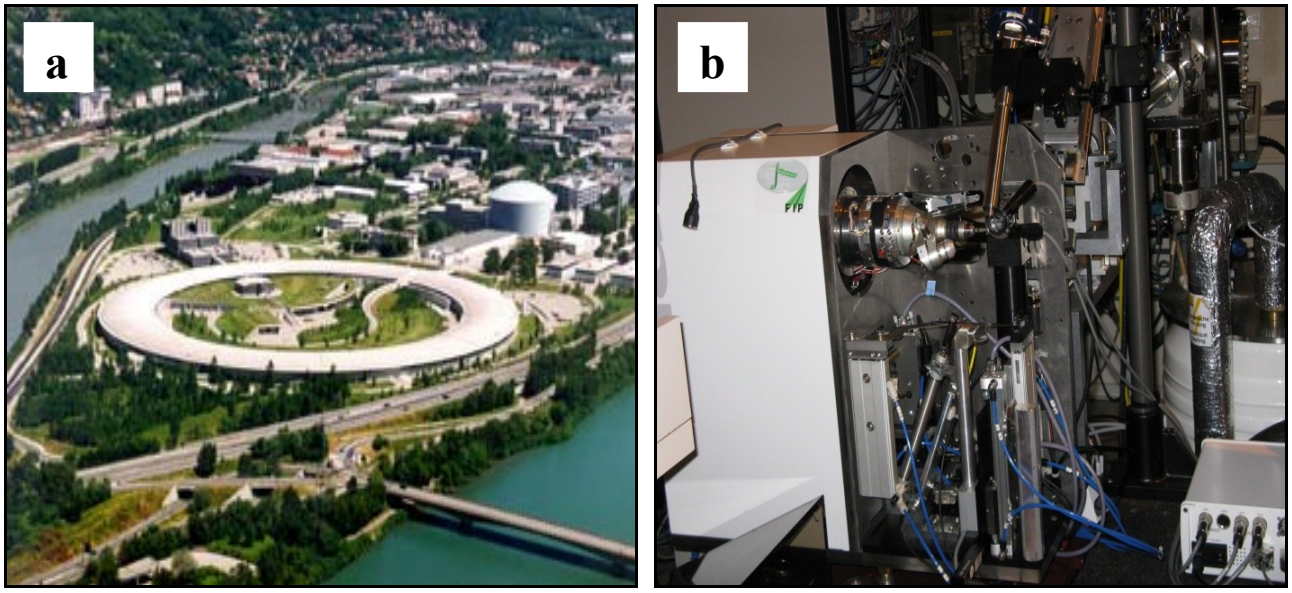


Figure 12: a) The European Radiation Synchrotron Facility (ESRF) at Grenoble, France. **b)** A view of the BM30A synchrotron beamline where diffraction data was collected.

B is the magnetic field strength in Tesla and λ_c is in *Å*. The main photon flux is close to E_c, but above E_c it drops exponentially as a function of the photon energy. The magnetic field causes the charged particles to change their direction. There are four types of magnetic devices in storage rings.

a) *Bending magnets* guide the electrons in the storage ring to stay in a circular closed orbit. At each bend the beam is guided into a straight section.

b) *Wavelength shifter* increases the magnetic field strength, B locally resulting in a decrease of λ_c (Equation 8) and produces higher intensity X-rays at shorter wavelength.

c) *Multipole wiggler* is a series of wavelength shifters where the radiation from consecutive magnets being independent, adds up incoherently in the general direction of propagation of the electron beam. The total flux is N times the flux generated by a single period (i.e. bending magnet), where N is the number of periods. It is easily tunable to desired wavelength.

d) *Undulators* are similar to multipole wigglers but with moderate magnetic fields and a large number of poles. The effect of this difference is that in the undulator strong interference occurs between the radiation from the consecutive magnets, which results in a spectral profile with a peak at a specific wavelength and a few harmonics. Therefore this radiation is not polychromatic, but the advantage for

some experiments is that a monochromator is not required. Its emitted intensity can be N^2 times that of a single period, with N the number of poles.

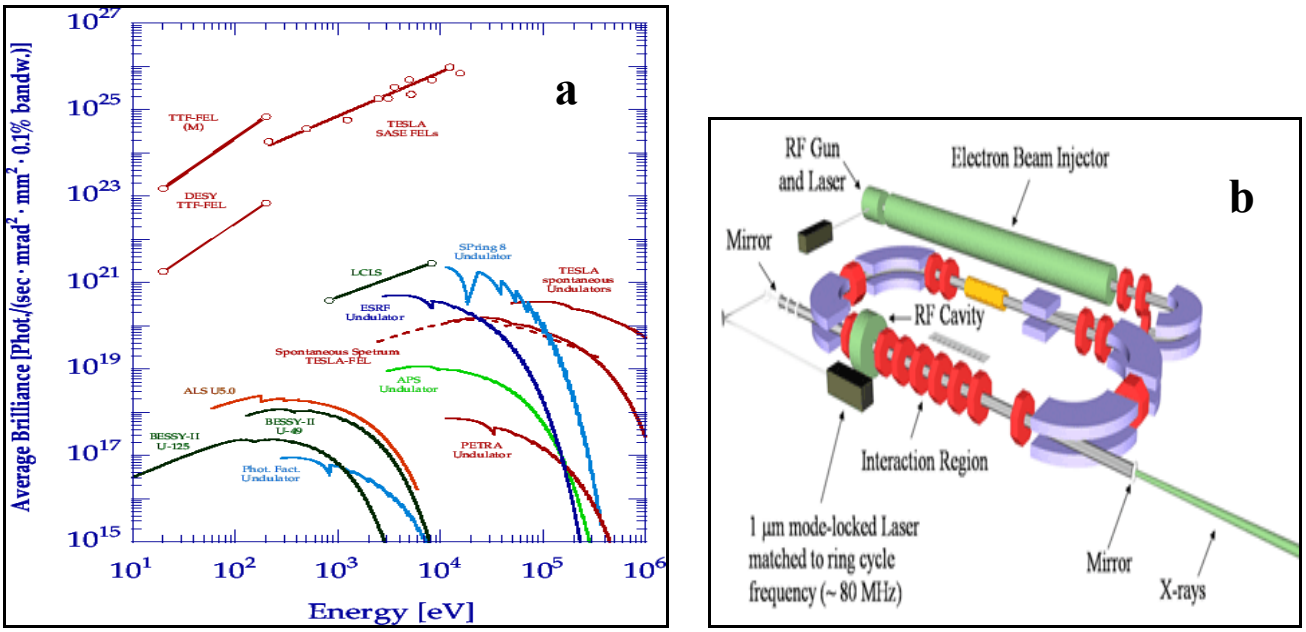


Figure 13: a) Comparison of the brilliance between the 3rd generation synchrotron sources and XFELs. **b)** A prototype of Compact Light Source [214].

Synchrotron sources have evolved through three generations and currently there are about 63 synchrotrons in the world. The advantages of synchrotron radiation for X-ray diffraction are: 1) its high intensity that is about five to six orders of magnitude higher than that from the laboratory-based rotating anode X-ray generators, 2) low divergence of the beam, resulting in sharper diffraction spots, 3) very good energy resolutions (~10^{-4}) of the X-ray photon energies or wavelengths produced and 4) tunability of X-rays to obtain desired wavelengths to solve the "phase problem" by multiple or single wavelength anomalous dispersion methods [205]. Synchrotron radiation is obtained in picosecond flashes produced by circulating bunches of charged particles. This property is being utilized for time resolved measurements in microsecond range towards understanding enzyme catalysis [206]. Together with the increasing availability of high intensity X-rays from synchrotrons and new crystal flash freezing techniques, data collection on very small crystals and on molecular assemblies has become feasible [207]. Synchrotron beams can be circularly polarized and can be used for circular dichroism studies to study protein-protein complexes [208].

The most promising directions for fourth generation sources in the wavelength range from the VUV to hard X-rays, are storage rings with even lower emittance than third generation rings, and short wavelength free-electron lasers (FELs) which offer femtosecond pulses with full transverse coherence

[209, 210]. The brilliance of these beams can be a billion times more than that of a third generation undulator source (Figure 13a). However, technological hurdles are still there in actual feasibility of operating the FELs in 1.0 Å wavelength regions. There are about nine FEL facilities around the world today. In a recent study, using nanocrystals of photosystem I, one of the largest membrane protein complexes, more than 3,000,000 diffraction patterns were collected, and a three-dimensional data set was assembled from individual photosystem I nanocrystals [211-213]. The problem of radiation damage in crystallography was taken care of by using pulses briefer than the timescale of most damage processes. This offers a new approach to structure determination of macromolecules that do not yield crystals of sufficient size for studies using conventional radiation sources or are particularly sensitive to radiation damage. Future beamlines worldwide are aiming to achieve up to a 10^5-fold increase in pulse irradiance by tighter focusing, allowing data collection with low-fluence, 10 fs pulses or pulses of even shorter duration. This provides a route to further reducing radiation damage and may allow measurements on even smaller nanocrystals, down to a single unit cell or a single molecule [211]. In future it has opened up the possibility of elucidation of structure from single molecules without the need of crystals!

Another new development is to produce a new generation of compact X-ray sources - miniature synchrotron light sources (Figure 13b). At Lyncean Technologies Inc., USA, the Compact Light Source, delivers a monochromatic beam of tunable hard X-rays, from a few keV up to 35 keV, comparable in quality to some of the most productive synchrotron beamlines in use today [214]. A laser beam colliding with an opposing electron beam has the same effect as an electron beam passing through an undulator magnet. The electric and magnetic fields of the laser beam cause the electron to wiggle and induces a radiation spectrum similar to that from a long undulator magnet. This radiation is typically referred to as Compton Scattering or Inverse Compton Scattering. If a laser beam with a wavelength of one micron is used, the electron beam energy necessary for 1 Å radiation is only about 25 MeV. The X-ray spectrum of the laser pulse/electron bunch collision is equivalent to that of a 20000 period undulator magnet. The X-rays can be focused using conventional X-ray optics down to a size of about 60 microns. Fine-tuning of the X-ray energy (for scans near absorption edges) is achieved with a monochromator adjustment just as with a synchrotron beamline. The instrument "Beta CLS" has been installed and commissioned at Accelerated Technology Center for Gene to 3D

Structure (ATCG3D). In June 2009, a 2 Å dataset on *Mycobacterium tuberculosis* glycine cleavage system protein crystal was collected (PDBID: 3IFT) [215].

2.2.2 X-ray detectors

X-ray detection for the protein crystallographic applications are carried out by integrating or single photon counting type detection system. Integrating detectors like image plates and CCDs [216, 217] are mostly used. Due to their large active surface, good spatial resolution and high dynamic range, IP detectors are popular in protein crystallography laboratories across the world [218]. However, the main drawback of IPs is its long readout time (1 min - 8 min), which becomes a problem at synchrotron sources, where exposure times are of the order of seconds. The main advantage of CCDs over IPs is their extremely small readout times (~ a few seconds) well suited at the synchrotron sources [219, 220]. CCDs achieve sensitivity or a quantum efficiency of about 80% as compared to only a few percent in case of IPs. The typical dynamic range is about 10^5, which is similar to the IPs. The main disadvantage of CCDs is that they are physically much smaller than IPs and consequently an array of such devices are required for practical applications. However, fabricating a large array of such devices is technologically challenging as the CCDs need to be cooled and joined properly.

A completely new type of detector for protein crystallography applications is a solid-state pixel array detector that offers a huge performance advantage over IPs and CCDs. The detector consists of an array of reverse-biased Si diode array bump-bonded to an Application Specific Integrated Circuit (ASIC) [221-223]. The hybrid technique of bump-bonding gives many advantages. The individual channels have low leakage current and low input capacitance allowing low readout noise. It can be operated in true single photon counting mode without dead-time and provide imaging at high speed. It offers frameless readout for high-speed and time-resolved crystallography, especially when the direct conversion type is used. Such detectors, called PILATUS, are being used at synchrotrons.

2.2.3 Other methods to study 3-D structure of macromolecules

The necessity to determine the three dimensional structures of proteins has led to the development of various sophisticated tools based on sound scientific principles. Following four approaches are currently being used to study the structures at a resolution range where individual macromolecules and their internal structure could be visualized.

1. **Neutron Crystallography:** Hydrogen atoms are ubiquitous in a functioning protein and play vital role in molecular recognition via hydrogen bonds, and hence their location is very important to understanding biomolecular function [224-226]. Exchangeable hydrogens can be precisely located by the neutron diffraction technique, which is similar to X-ray diffraction in principle [227]. Neutron sources are specialized and costly and are of two types: nuclear reactor based (fission neutrons) and accelerator based (spallation neutrons). The neutron data can be coupled to X-ray data in order to refine the atomic positions more accurately and obtain the charge separation in polarized bonds [228].

2. **Cryo-electron microscopy:** In this technique, the biological sample is flash frozen to liquid nitrogen temperatures and investigated in high vacuum by high-energy electrons (typically 200-300 keV) using the principles of transmission electron microscopy (TEM) [229]. In this direct imaging technique 3-D structure is reconstructed from several 2-D projections of the molecule. Cryo-electron microscopy becomes most useful for large assemblies showing them in their native environment in comparison to X-ray crystallography [230]. However there are problems with harsh environment due to high vacuum and sample damage, which means that very low electron dose must be used to avoid destroying the sample, so that the images have extremely poor signal-to-noise ratio and must be averaged out.

3. **Nuclear Magnetic Resonance:** NMR detects chemical shifts of atomic nuclei with non-zero spin (e.g. ^{1}H, ^{13}C and ^{15}N) [231]. The shifts depend on the electronic environments of the nuclei, namely, the identities and distances of nearby atoms. From these constraints, three-dimensional structure of proteins is derived as a set of 10-50 probable models. About 16% of the structures deposited in the PDB have been determined by NMR. NMR allows the study of the proteins in solution. The protein must however be soluble to a high concentration (0.2-1.0 mM) and be stable for days without aggregation under the experimental conditions. These limitations make the structure elucidation of typical membrane proteins and filamentous proteins unlikely. But intrinsically unstructured proteins, not amenable to crystallization, can be studied by NMR [232]. In principle the size limitation can be overcome by solid state NMR [233].

4. **Small angle scattering:** The scattering peak near the direct beam, below 10°, of a macromolecular solution contains the size and shape information about the biomolecule in solution [234]. The probe can be X-rays or thermal neutrons. This technique is useful to identify the

homogeneity, multimeric states, complexes and foldability of the sample [235, 236]. Low resolution structures can be determined, especially for proteins difficult to crystallize. Contrast matching and upcoming data analysis and refinement programs shall enable low resolution structure determination of multimeric proteins, protein-DNA/RNA and protein-protein complexes in solution [237].

5. Powder diffraction: Powder diffraction is an emerging tool to elucidate the structures of proteins from polycrystalline powder samples [238, 239]. This technique is useful when the protein crystals show poor diffraction, are radiation sensitive and form weak complexes [240].

2.3 Single Crystal Diffraction Data Collection

Different strategies are used to record within reasonable time and accuracy intensities of all Bragg reflections produced by the crystal. The most common method is to change the orientation of the crystal relative to the incoming monochromatic X-ray beam. This causes a corresponding rotation of the reciprocal lattice and in this manner all reflections can be passed through the Ewald sphere. It is however also possible to change the Ewald sphere radius by changing the wavelength and this is the principle of Laue diffraction. The methods of data collection are given below.

- Oscillation method (single axis rotating crystal, stationery 2D detector, overlap problem) [241]
- Weissenberg method (single axis rotating crystal, moving 2D detector, gives distorted picture of reciprocal lattice) [242]
- Precession method (single axis rotating crystal, moving 2D detector synchronized with the crystal movement, gives undistorted picture of reciprocal lattice) [243]
- Multi-circle diffraction method (multiple axis rotating crystal, moving 1D detector, one reflection at a time, more accurate but relatively slow) [203]
- Laue method (all wavelengths and many reflections at a time, fast, overlap problem) [244]

2.3.1 Oscillation Method

The oscillation method, which has been used for data collection in the present thesis work, will be briefly discussed. With the advent of 2D detectors and their fast response, currently, oscillation method is the most popular method. The oscillation method is a straightforward technique where the crystal is rotated around an axis (the oscillation axis) that is perpendicular to the incident X-ray beam (Figure 14). To avoid overlapping of reflections on the stationary 2D detector, the crystal can be

rotated over only a small angle (0.25°-2°) per diffraction image. Crystal is oscillated up and down this angular range several times (called passes) during each exposure, ensuring changes in beam intensity are averaged out. Depending upon the spot overlaps the oscillation width needs to be reduced. The total rotation should be 180° in order to collect all diffracted rays. However, there is redundancy in the reciprocal lattice due to point group symmetry, and hence rotation range is determined by 180°/n (where 'n' is the symmetry number).

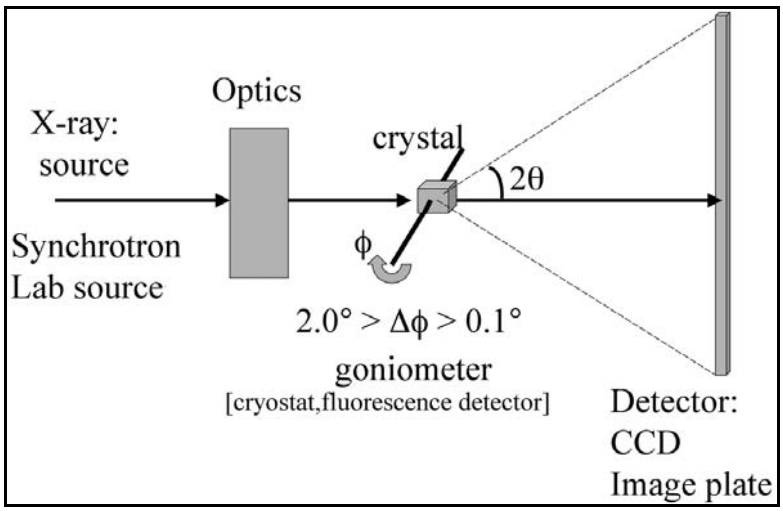


Figure 14: A schematic view of an X-ray diffraction experiment. The X-ray beam is produced by the source on the left and is conditioned by the optics before interacting with the crystal mounted on the goniometer. Scattered X-rays are detected on the CCD or imaging-plate detectors.

2.4 Data Processing

The observed intensity of a reflection is given by Darwin's equation,

$$I\,(hkl) = I_0 .\, r_e^2 .\, (\lambda^3 / \omega)\, \{Vx / V^2\} \,.\, L \,.P.\, A.\, |F_{hkl}|^2 \qquad (10)$$

where $I\,(hkl)$ is the measured intensity on the detector, I_0 is the incident beam intensity, V_x is the volume of the crystal illuminated in the X-rays, V is the volume of the crystal unit cell, r_e is the classical electron radius (2.818×10^{-15} m) of an electron, λ is the X-ray wavelength, ω is the angular velocity of the crystal, L is Lorentz correction, P is polarization correction and A is the absorption correction [245-248]. The analysis and reduction of a single crystal raw diffraction data consists of seven major steps.

1) Visualization and preliminary analysis of the original, unprocessed detector data.

2) Indexing of the diffraction pattern, i.e., to determine the indices h, k, l of each reflection, as well as the parameters that define the size and shape of the crystal unit cell, and its orientation on the diffractometer.

3) Refinement of the crystal and detector parameters.

4) Integration of the diffraction maxima.

5) Apply geometric corrections and finding the relative scale factors between measurements.

6) Precise refinement of crystal parameters using whole data set.

7) Merging and statistical analyses of the measurements related by space group symmetry, putting all the reflections from all of the images on the same relative numeric scale and then merge and average them to produce a unique reflection list with associated intensities and realistic error estimation.

Several computer programs are available to perform all these steps. Widely used data reduction program packages are *XDS* [249-254], *MOSFLM* [255-257], *DPS* [258-261], d*TREK [262, 263] and *HKL* [241].

For the work presented in this thesis, the *XDS* software suite was used for data reduction/processing. The XDS software suite consists of the following programs: *XDS* for indexing [241, 251, 258, 264-273] and integration and *XSCALE* for scaling and merging data [253, 274-281].

A reduced cell is extracted from the observed diffraction pattern and processing of the data images continues to completion as if the crystal were triclinic. The reflection indices then refer to the reduced cell and reindexed once the space group is known. For all space groups, the required reindexing transformation is linear and involves only whole numbers [282]. Automatic space-group assignment is carried out in two steps once integrated intensities of all reflections are available [253]. Firstly, the Bravais lattices are identified that are compatible with the reduced cell derived from the observed diffraction pattern. In the second step, all enantiomorphous space groups compatible with the observed lattice symmetry are rated by a redundancy-independent *R* factor [283, 284]. The group is selected that explains all integrated intensities in the data set at an acceptable *R* factor requiring a minimum number of unique reflections.

The quality of the data is judged by the parameters, *R*-merged, I/σ (I) and completeness, at automatically determined resolution shells.

$$R\text{-merge} = \Sigma_{hkl} \Sigma_i \mid I_i(hkl) - \langle I(hkl) \rangle \mid / \Sigma_{hkl} \Sigma_i I_i(hkl) \qquad (11)$$

The summation is over all observations, and the values of *R* are typically 0.05 - 0.12 for reasonably well-measured data. The criteria for the inclusion of the highest resolution shell should roughly

satisfy all or at least a pair of the following conditions: I (hkl) / σ {I (hkl)} ≥ 2.0, completeness ≥ 75% and R-merged ≤ 50 %.

2.5 Phase solution

Once the diffraction data has been obtained, the next step is to solve for the phases of the reflections. The protein structure solution scheme is given in Figure 15 and methods used are given in Table 3.

Table 3: Methods used in phase solution.

Method	Prior knowledge
Direct methods	ρ > 0, discrete atoms
Molecular replacement	Structurally similar model
Isomorphous replacement	Heavy-atom substructure
Anomalous scattering	Anomalous-atom substructure
Density modification (phase improvement)	Solvent flattening Histogram matching Noncrystallographic symmetry averaging Automatic partial structure detection Phase extension

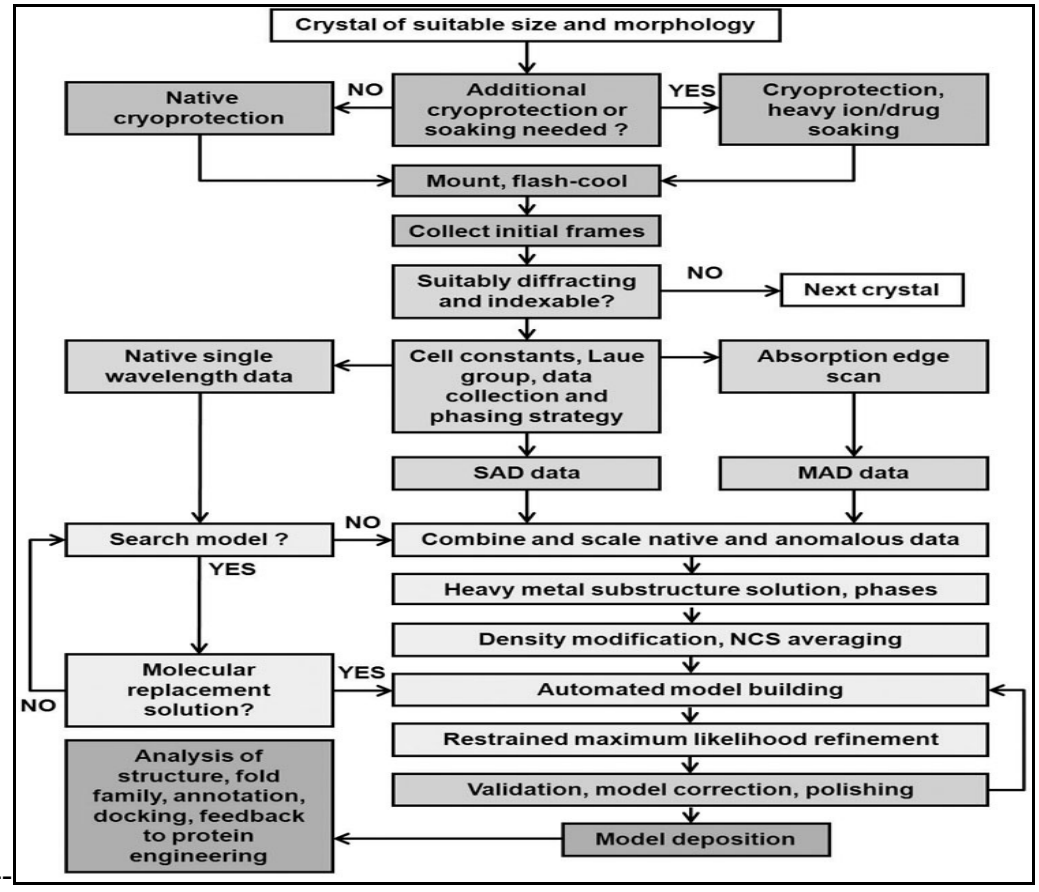


Figure 15: The structure solution protocol.

The methods such as Molecular Replacement (MR) [285-297, 300-305], Multiple Isomorphous Replacement (MIR) [306-309], Multiple Wavelength Anomalous Diffraction (MAD) and Single

Wavelength Anomalous Diffraction (SAD) [310-319] are used for this purpose. In this thesis Molecular Replacement method is used and will be described next, followed by the other methods.

2.5.1 Molecular replacement (MR) method

Molecular replacement is fundamentally a search method of solving crystal structures when a suitable related model is available [285-297]. The model structure may correspond to a crystal structure or a solution structure of a homologue of the unknown structure. The required sequence identity should be > 25 %. MR enables the solution of the crystallographic phase problem by providing initial estimates of the phases of the new structure from a previously known structure. MR is a good phasing method because it involves the collection of a single data set from a single crystal and has the advantage of minimizing the effects of radiation damage. With the success of structural genomics and proteomics projects, the number of available structures will increase rapidly, leading more harness of MR method. The pre-requisites for MR method are: a) an observed diffraction pattern (intensities) of the unknown target structure, and b) the atomic coordinates of a homologous probe protein structure.

The MR task involves positioning the probe within the unit cell of the target crystal in such a way that the theoretical diffraction pattern that would result from this model closely matches the experimental one. If only one molecule is present in the asymmetric unit, then six parameters (three rotational and three translational), which fully describe how the probe is placed in the unit cell, would have to be determined. In principle, one could do a search on these six parameters to determine the position of the probe that gives the best agreement between observed and calculated structure factor. However, such a search would be computationally too demanding. From the theoretical analysis of the properties of the Patterson function [298, 299] it became obvious that such a six-parameter search could be reduced to two three-dimensional problems. The first is the determination of the correct orientation of the probe, and the second is the determination of the position of the correctly oriented molecule within the unit cell. Therefore, the main aim of the MR method is to find these two operators to solve the rotation and the translation functions. The searches can be by using either match in the Patterson function or through likelihood functions. There have been several advances in the MR technique which extend its scope [290-297, 320, 321].

2.5.1.1 The Rotation Function

Patterson function is used to obtain the rotation function [286]. Patterson function $P(\vec{x})$ is the Fourier transform of $\left|\vec{F}(hkl)\right|^2$ and can be calculated using the observed structure factor amplitudes without knowing the phases. If the Patterson function is calculated from the structure factor amplitudes of the crystal, $P_1(\vec{x})$ and that calculated from the model, rotated by the rotation operator $\vec{R}$, $P_2(\vec{R}x)$, then the rotation function is given by,

$$RF(\vec{R}) = \int_U P_1(\vec{x}) P_2(\vec{R}x) dV \tag{12}$$

where the integral is over a volume U, and large values are obtained when there is a maximum overlap between $P_1(\vec{x})$ and $P_2(\vec{R}x)$. Such a function describing the overlap, calculated in reciprocal space, has been first presented by Rossmann and Blow [286, 289]. A maximum in this rotation function $RF(\vec{R})$ indicates a potentially correct orientation of the search probe in the target cell. However, the problem with the Rossman-Blow function was large computation time required and lack of precision owing to truncation errors in the interference function obtained from the intergrals in equation 12. This is modified to a new form and was described as the fast rotation function by Crowther [300] which is based on spherical coordinate system and is computationally faster.

2.5.1.2 The Translation Function

Once the orientation of a probe molecule is known from the rotation function, the actual position of the molecule in the unit cell has to be determined. In the translation function the correlation between the observed intensities and the Patterson cross-vectors of the symmetry-related molecules of the probe is calculated as the molecule is moved within the cell [286, 322]. When the molecule is correctly positioned, the translation function should have peaks at values corresponding to the translation vectors between the symmetry related molecules. The combination of a three-dimensional translation function with a packing analysis [301] gives more realistic and accurate results. The translation is also determined from the correlation coefficient between the observed and calculated amplitudes squared, as a function of model translation. The Patterson correlation (PC) score is,

$$C(t) = \frac{\sum \left(|F_O|^2 - \overline{|F_O|^2} \right) \left(|F_C(t)|^2 - \overline{|F_C(t)|^2} \right)}{\sqrt{\sum \left(|F_O|^2 - \overline{|F_O|^2} \right)^2 \sum \left(|F_C(t)|^2 - \overline{|F_C(t)|^2} \right)^2}} \quad (13)$$

The goal of PC refinement is to improve the overall orientation of the search model and is carried out for rigid bodies such as domains, subdomains or secondary-structure elements. The target function of PC refinement is typically defined as the standard linear correlation between observed and calculated squared normalized structure-factor amplitudes ($|E^2|$). By improving the accuracy of the search model for the correct angular orientation, PC refinement improves the discrimination between correct and incorrect orientations and therefore enables the location of the correct peak in a noisy rotation function.

2.5.1.3 Maximum Likelihood based MR method

The likelihood of the model given the data is defined as the probability of the data given the model. Maximum likelihood (ML) model is that model which explains the measured data with highest probability [323]. The likelihood of the model given the data is defined as the probability of the data given the model. Where the data have independent probability distributions, the joint probability of the data given the model is the product of the individual distributions. In crystallography, the data are the individual reflection intensities which are not strictly independent. Systematic variation in intensities and correlations present due to non-crystallographic symmetry hinders estimation of joint probability as a simple product. Nevertheless, the assumption of independence and anisotropy corrections are necessary to make the problem tractable. The log of the likelihood function (LL) is used to give a reasonable numeric value compared to the product of small probabilities. In the present thesis *Phaser* is used for MR solutions. In the ML functions in *Phaser*, the structure-factor amplitudes (Fs), or normalized structure-factor amplitudes (Es) are used [324]. Integrating out the unknown phases gives the *Rice* distribution or *Sim* [325] distribution from a two-dimensional Gaussian distribution. The structure factors have four probability distributions: 1) for an acentric reflection, the probability distribution is (*Rice* distribution) dependent on F_O with its expectation value, $\langle F_O \rangle$ centered on $\langle F_O \rangle$, 2) for a centric reflection, the probability distribution is a *Woolfson* distribution, 3) for an acentric reflection when the atoms are totally random, i.e. $\langle F_O \rangle = 0$, the *Rice* becomes *Wilson* distribution and, 4) for a centric reflection and random atoms *Woolfson* becomes centric

Wilson distribution. In MR, the expection values $<F_O>$ of the structure factors, using the set of free reflections, are calculated from the model, F_C multiplied by the fraction of correct structure, D ($0 > D > 1$). These probability distributions are the basis for all the maximum likelihood functions used in *Phaser*. In a typical run of the program, the fast rotation and translation functions are used to identify potential solutions, which are then rescored using the full likelihood functions to determine the best solution. Packing criteria are a very powerful constraint on the translation function. It has been shown that judicious editing can make a significant difference in the quality of a distant model. Any available phase information, be it from experimental phasing (i.e. multi-wavelength anomalous diffraction (MAD), single-wavelength anomalous diffraction (SAD), Multiple Isomorphous Replacement (MIR), etc.) or a partial molecular replacement solution, can be used in solving molecular replacement problems through calculations of phased-rotation and phased-trnslation functions. In models having poor sequence identity, an ensemble (e.g. ensemble of NMR models or homology models) may be provided in *Phaser*. The best ensembles are generally derived using alignment protocols, such as *FFAS* protocol [326]. Automated Molecular Replacement in *Phaser* combines the anisotropy correction, likelihood enhanced fast rotation function, likelihood enhanced fast translation function, packing and refinement modes for multiple search models and a set of possible space groups to automatically solve a structure by molecular replacement.

For molecular replacement, software suites *AMORE* [302, 303] and *MOLREP* [305] use fast rotation and translation functions, but both lack likelihood-based scoring functions. In *MOLREP* the search for multiple copies of a molecule in the unit cell and incorporation of the packing function provides better contrast indentifying correct solutions. Another approach in MR that is implemented in the program *CNS* [290, 292] is based on two concepts- *i*) the direct rotation search and *ii*) Patterson correlation (PC) refinement. In the traditional rotation search, the computationally expensive structure-factor calculation is carried out only once to obtain a model Patterson map, which is then rotated and superimposed with the observed Patterson map. In the direct rotation search, however, the model is rotated directly and a structure-factor calculation is carried out for each sampled angular orientation. This has the advantage of avoiding approximations such as interpolations, but has the disadvantage of being computationally expensive. The intervening step between the rotation search and the translation search in *XPLOR/CNS* is the PC refinement [290].

Programs to perform six dimensional searches similar to conventional MR are implemented in programs such as *'Queens of Spades', 'EPMR'* and '*SOMoRe'* [294, 295, 327].

2.5.2 Multiple isomorphous replacement (MIR)

This method of phasing involves the isomorphous attachment of heavy atoms to the protein molecules in the crystal [306, 307]. Since different atoms contribute to the scattered intensity in proportion to the square of the number of electrons they contain, a heavy atom like U, Hg, I, As, W, etc. which contain many electrons than atoms typically present in proteins, will contribute much more to the diffracted intensities. As a result, the change in intensity from the addition of one heavy atom to a typical protein is easily measured. If there are two crystals, one containing just the protein (native crystal) and other containing in addition bound heavy atoms (derivative crystal); one can measure diffraction data from both. The differences in scattered intensities will largely reflect the scattering contribution of the heavy atoms, and these differences can be used to compute a Patterson map [308]. Because there are only a few heavy atoms, such a Patterson map will be relatively simple and easy to deconvolute (Alternatively, direct methods can be applied to the intensity differences). Once the location of the heavy atoms in the crystal is known, their contribution to the structure factors can be calculated. Here the assumption is made that the heavy atom doesn't change the rest of the protein structure and scattering from the protein atoms is unchanged by the addition of heavy atoms. Thus, the structure factor for the derivative crystal $\left(\vec{F}_{PH}\right)$ is equal to the sum of the protein structure factor $\left(\vec{F}_{P}\right)$ and the heavy atom structure factor $\left(\vec{F}_{H}\right)$,

$$\vec{F}_{PH} = \vec{F}_{P} + \vec{F}_{H} \tag{14}$$

In this equation only the lengths of $\vec{F}_{PH}$ and $\vec{F}_{P}$ are known, but $\vec{F}_{H}$ is known both in length and direction. Through Harker construction [309], it turns out that there are two ways to draw vector triangle that will satisfy above vector equation leading to two possible phases of $\vec{F}_{P}$. This twofold phase ambiguity can be removed by preparing a second derivative crystal with heavy atoms that bind at other sites. Only one phase choice of $\vec{F}_{P}$ will be consistent with all three observations. Hence, in principle, at least two heavy atom derivatives are required to get the phase information.

2.5.3 Multiple wavelength anomalous diffraction (MAD)

With the advent of powerful synchrotron sources, this method exploits the fact that, at certain wavelengths, tightly bound electrons in an atom do not behave like free electrons in the scattering process [205, 310]. This causes the anomalous scattering, an effect that depends on the wavelength and is generally stronger for heavy atoms [205, 311, 312]. The consequence of anomalous scattering is that the intensities of certain otherwise equivalent reflections called Friedel and Bijvoet mates [311-318], are no longer equal, and this difference in intensities can be used for phase angle determination. This method facilitates phase determination from one single crystal and the earliest experiments showing the practical feasibility of this approach were carried out by Ramaseshan *et al.* in 1957 [205]. Most often Se-atom is introduced in the protein by growing the microorganisms on Se-methionine substrate instead of the normal methionine containing substrate [319, 328]. The single-wavelength anomalous dispersion (SAD) can be used as a phasing method [329-333]. The advantage is single dataset is required but the statistical accuracy has to be increased by collecting a highly redundant intensity dataset. Another advantage is that in-house laboratory sources can be used to collect a sulphur-SAD dataset utilizing the intrinsic sulphurs of proteins [332, 333].

2.5.4 Direct methods

Direct methods are based on the positivity and atomicity of electron density that leads to phase relationships between the (normalized) structure factors. For developing direct methods for phasing Hauptmann and Karle shared the 1985 Nobel Prize in Chemistry [334-339]. The direct method attempts to calculate the phases directly from the magnitudes of normalized structure factors $\left|\vec{E}_h\right|$. The method relies on the fact that certain linear combinations of phases are structure invariants, which do not depend on the choice of origin. The most useful structure invariants are triplet invariants $\Phi_{hk} = \phi_h + \phi_{-k} + \phi_{-h+k}$ where h and k are two arbitrary indices of the reciprocal lattice and ϕ_h is the phase of a structure factor $\vec{F}_h$. The triplet relation shows how the phases of three reflections are related. Therefore, knowing the phases of two reflections and the structure invariant Φ_{hk} allows one to derive the phase of a third. Assuming a random distribution of identical atoms, the probability distribution of the invariant, $P(\Phi_{hk})$, is given by,

$$P(\Phi_{hk}) = (1/L)\exp(G_{hk}\cos\Phi_{hk}) \tag{15}$$

where L is a normalization constant and $G_{hk} = (2/\sqrt{N})|E_h E_k E_{h-k}|$; N being the number of atoms in the unit cell. This distribution is maximal when $\Phi = 0$ and decreases as Φ deviates further from 0. Also, $P(\Phi_{hk} = 0)$ increases as $|E_h E_k E_{h-k}|$ increases and N decreases. Each reflection is involved in multiple invariants having multiple probability distributions. The tangent formula [335],

$$\tan\phi_h = \frac{\sum_{j=1}^{r} G_{hk_j} \sin(\phi_{k_j} + \phi_{h-k_j})}{\sum_{j=1}^{r} G_{hk_j} \cos(\phi_{k_k} + \phi_{h-k_j})} \tag{16}$$

combines these distributions and assigns the most probable phase of h given other phase angles. Most classical direct methods use the probability distributions or the tangent formula to generate phases with high overall probability. As the number of atoms (N) increases, the probability distribution flattens. One of the reasons the classical direct method fails for molecules with more than ~200 atoms in the asymmetric unit, is that it becomes much harder to predict the phases when $P(\Phi_{hk})$ is almost the same for all values of Φ_{hk}.

The direct methods are also being utilized increasingly now in combination with anomalous-dispersion methods and isomorphous replacement methods to locate the position of heavy or anomalous atoms from a very small signal using direct methods programs, *SnB* [340-344] and *SHELXD* [345, 346]. Several programs, mostly based on maximum-likelihood principles, are available for carrying out heavy-atom refinement and phasing. These include: *CNS* [347], *MLPHARE* [348], *SHARP* [349], *SOLVE* [350], and *HySS* [351].

2.5.5 Other method of phasing

A new phasing method [352-354], called *VLD* (*Vive la Difference*), is based on difference and hybrid Fourier syntheses [355-357]. It is claimed that this new method can obtain the correct difference structure even when the model is completely uncorrelated or is completely random from the true structure. *VLD* has been implemented in the program *IL MILIONE* [358]. It can also use a starting seed model, found by other phasing methods (e.g., Patterson techniques, MR, SAD, MAD, SIR-MIR), and extend and refine the phases. This phasing algorithm is suitable for structural solution in Structural Proteomics projects.

2.6 Density modification

Once the phase information is available and a preliminary model of the protein is known, the next step is the calculation of the electron density map. The quality of the phases determines the quality of the resulting map. There are methods to improve the initial electron density maps, and these are referred to as density modification methods [359-361]. Density modification works by incorporating features that are expected to appear in a correct map, such as the similarity of regions related by non-crystallographic symmetry [362], flatness or disorder of the solvent region [329, 363, 364] and the similarity of the density-map histogram to histograms of macromolecules whose structures have been determined earlier [365]. Histogram matching is a robust statistical density modification technique which is done along with solvent flattening. The average features of electron density of known proteins are similar, as they are composed of twenty amino acids, and this is utilized in obtaining the frequency distribution of the electron density values. These distributions are a function of electron density, and are independent of protein at a given resolution. The plots of frequency distributions of the electron density levels calculated at grid points as a function of electron density are scaled to resemble a standard plot. Solvent flattening and averaging improve the phases of low resolution reflections and when combined with histogram matching, the phases of higher resolution reflections improve considerably [360].

2.7 Electron density map interpretation

Once the initial electron density maps are improved, the next step is the interpretation of these maps, which is nowadays achieved with the help of a broad spectrum of programs and powerful computers. Interactive computer graphics programs *O* [366] and *Coot* [367] were used to build the model into the calculated *2mFo-DFc* electron density maps [368]. The model is usually built by interpreting simultaneously two types of electron density maps, the *mFo-DFc* [369] and the *2mFo-DFc* map contoured at 2 σ and 1 σ levels respectively. The *mFo-DFc* map contains positive peaks where density is not accounted for by the model used to calculate the *Fc*, and hence is useful to locate missing or wrongly placed atoms. The *2mFo-DFc* map is the sum of a *Fo* map and *mFo-DFc* map, and contains information from both the *Fo* map and the difference *mFo-DFc* map. In the simulated annealed (SA) omit electron density map calculation, portion/portions of the model in the region of

interest is/are left out and the rest of the model is used in phase calculations reducing model bias [370, 371].

2.8 Crystallographic refinement

The aim of refinement is optimization of the atomic model to simultaneously fit both observed diffraction data and *a priori* chemical information. The refinement of a macromolecular structure is a difficult optimization problem because of poor parameters to observations ratio and because of correlations between the coordinate of each atom to all the other (nearby/bonded) atoms [372]. The target functions are functions of many parameters and their errors [373, 374]. The large number of adjustable parameters results in the multiple minima problem — the target function contains many local minima in addition to the desired global minimum. The challenges of crystallographic refinement arise not only from this high dimensionality of the parameter space, but also from the poorer quality and/or lower resolution of experimental phases which guide correct shifts to atomic parameters.

The refinement programs differ in the nature of the target function(s) and the method(s) by which the target function is minimized. In energy-based methods, such as *CNS* [370] and *PHENIX* [375] the target function is expressed as energy and the weights of the restraints are estimates of the energy penalty associated with deviations from the target values. Other refinement programs are implemented in *REFMAC* [376], *TNT* [377], and *SHELX* [378]. The force constants and stereochemical parameters are derived from parameter sets of Engh and Huber [379] and Parkinson et. al. [380], derived from the high-resolution small molecule crystal structures. All the structures reported in the present thesis were refined using maximum likelihood target function and simulated annealing optimization to overcome multiple minima problem [372].

2.8.1 Target functions

Crystallographic refinement is a search for the global minimum of the target,

$$E = E_{chem} + w_{xray} E_{xray} \tag{17}$$

where E is a function of the parameters of an atomic model. E_{chem} comprises empirical information about chemical interactions; it is a function of all atomic positions, describing covalent (bond lengths, bond angles, torsion angles, chiral centers and planarity of aromatic rings) and non-bonded

(intramolecular as well as intermolecular and symmetry related) interactions [381, 382]. E_{xray} is related to the difference between the observed (*Fo*) and calculated data (*Fc*) and w_{xray} is a weight appropriately chosen to balance the gradients (with respect to atomic parameters) arising from the two terms.

2.8.2 Maximum Likelihood refinement targets

The most commonly used target function (E^{LSQ}) for macromolecular refinement employs the least-squares residual for the diffraction data [383],

$$E^{LSQ} = E_{restraints} + w_a \sum_{hkl} \left(|F_o| - k|F_c| \right)^2 \tag{18}$$

where $|F_o|$ and $|F_c|$ are the observed and calculated structure-factor amplitudes, k is a relative scale factor, w_a is a weight, and $E_{\text{restraints}}$ are geometric (bond length, bond angle, and atomic repulsion) restraints. A decrease of this function can sometimes be due to accumulation of systematic errors in the model without improvement or even a worsening of the model. The underlying reason can be found in the fact that the least-squares residual does not account for the effects of phase errors in the calculated structure factors, so it is poorly justified when the model is far away from the correct answer or incomplete. A more appropriate target for macromolecular refinement can be obtained through a maximum likelihood formulation [323, 384-389]. The goal of the maximum likelihood method is to determine the probability of making a set of measurements, given the model, and estimates of its errors and of errors in the measured intensities [372]. The effects of model errors (misplaced atoms and missing atoms) on the calculated structure factors are first quantified with σ_A values, which correspond roughly to the fraction of each structure factor that is expected to be correct [371]. To achieve an improvement over the least-squares residual (Equation 19), cross-validation [390-393] was used for the computation of σ_A, necessitating its calculation with a randomly selected test set of diffraction data that was never included in the refinement process. The cross-validated σ_A values (σ_A^{cv}) are then used to compute the expected value of $\langle |F_o| \rangle^{cv}$. $\langle |F_o| \rangle^{cv}$ and $\sigma^2_{ML^{cv}}$ can be readily incorporated into a maximum likelihood target function,

$$E^{ML} = E_{restraints} + w_a \sum_{hkl \in working set} \left(\frac{1}{\sigma^2_{ML^{cv}}} \right) \left(|F_o| - \langle |F_o| \rangle^{cv} \right)^2 \tag{19}$$

The least-squares residual is a limiting case of the maximum likelihood theory and is only justified if the model is nearly complete e.g. rigid-body refinement [394]. For many structures, some initial experimental phase information is available from either MIR or MAD methods. These phases represent additional observations that can be incorporated in the refinement target. The addition of experimental phase information in the target function greatly improves the results of refinement [388].

2.8.3 Minimization method - Simulated Annealing

The suitable target function has to be minimized. In protein structure refinement, two minimization methods are used- a) gradient descent and b) simulated annealing. Crystallographic target function has multiple minima and it is difficult to cross the barrier between the minima by the gradient descent method. Simulated annealing method [395-399] can overcome the multiple minima having greater radius of convergence, thereby reducing the manual intervention required during refinement.

By defining the target *E* (equations 18 and 19) to be the equivalent of the potential energy of the system, one can simulate annealing process. The likelihood of uphill motion is determined by a control parameter referred to as temperature. The higher the temperature, the more likely it is that simulated annealing will overcome barriers. The two most commonly used protocols are linear slow cooling or constant temperature but better model is obtained with slow cooling [397]. Many examples have shown that simulated annealing refinement, starting from initial models obtained by standard crystallographic techniques, produces significantly better final models compared to those produced by conjugate-gradient minimization method [400]. The parameter space can be reduced drastically for proteins by employing simulated annealed torsion angle molecular dynamics [401, 402].

In the present thesis, *CNS, CCP4* and *Phenix* were used for refinement. The structures were refined in *CNS* using standard simulated annealing protocols and the amplitude based maximum likelihood target function [400, 402]. A total of 5 % of randomly selected reflections were set aside for cross validation [391]. All reflections in the respective resolution ranges were included in the refinement. In the initial stages of SA refinement and during calculation of SA omit maps, the model was heated to a temperature of 2000 ^{o}C, and then annealed at a cooling rate of 25 ^{o}C per iteration. Initial anisotropic B-factor and bulk solvent corrections were applied [403]. The relative weighting between geometric and X-ray terms in the target function was determined automatically in *CNS*. The

parameters and the minimization methods used in various commom refinement programs are documented [404].

2.8.4 Multi-start refinement

Multi-start refinement is multiple SA refinements starting from a single model [405]. Some of the models resulting from multi-start refinement may give lower free *R*-value. Each model coming from a multi-start refinement fits the data slightly differently and more variance indicates intrinsic flexibility within the molecule. Also the regions in the starting model that contain significant errors or poor electron density show increased variability after multi-start refinement and a visual inspection of the ensemble of models produced can be helpful in identifying these incorrectly modeled regions.

In order to better identify the correct conformation, structure factors from each of the models can be averaged [405]. This averaging tends to reduce the effect of local errors (noise), which are presumably different for each member of the family. The average structure factors produce phases that contain less model bias than phases computed from a single model. It also produces better estimates of errors in the model for maximum likelihood targets and σ_A-weighted electron density maps because *Fc* is used in the computation of these parameters. Multi-start refinement followed by structure-factor averaging is useful in situations in which there is significant noise, namely when the data to parameter ratio is very low. The program *wARP* [406] makes efficient use of structure-factor averaging in the context of phase improvement and automated model completion.

2.8.5 TLS Refinement

TLS is rigid body collective displacements of whole molecules, domains and secondary structure elements. TLS stands for translation, liberation and screw motions comprising of 20 TLS parameters per group [407-409]. In this refinement, motions of groups of atoms are utilized, instead of individual atoms, which reduce the number of parameters drastically. A protein chain is divided into few groups depending upon domains, flexible and rigid groups [408]. After fitting the model, *PHENIX* [375] and *CCP4* [376] were used for TLS refinement on the protein model.

2.8.6 R-factor and cross-validation

A widely used indicator to represent the correctness of the model structure is the R-factor or the disagreement factor. It is defined as,

$$R = \frac{\sum_{hkl} \left| |F_{obs}| - k|F_{calc}| \right|}{\sum_{hkl} |F_{obs}|} \times 100\% \qquad (20)$$

For a well-refined protein structure it is generally below 20 %. A low R-factor means that the model agrees with the diffraction data more closely. However there is always a danger of overfitting the diffraction data. To reduce this danger the concept of cross-validation in the form of the free *R*-value has been introduced [393]. Cross-validation also produces more realistic coordinate error estimates based on the *Luzzati* or σ_A methods [410, 371]. For cross-validation, the diffraction data are divided into two sets—a large working set (typically comprising 95% of the data) and a complementary test set (comprising the remaining 5%). The diffraction data in the working set are used in the normal crystallographic refinement process, whereas the test set data are not. The cross-validated *R*-value (or R_{free}) computed by equation 20 using the reflections in the test set is a better indicator of model quality. It provides a more objective guide during the model building and refinement process than the conventional *R*-value that is computed by above equation using reflections in the working set. R_{free} also indicates whether the introduction of additional parameters (e.g. water molecules, the relaxation of non-crystallographic symmetry restraints or multi-conformer models) improves the quality of the model or, instead, increases overfitting.

Water molecules are added manually by examining environment around electron densities that were present in both *mFo-DFc* and *2mFo-DFc* maps. A composite omit map, as implemented in *CNS*, is calculated to enable unbiased interpretation of the electron density map. Composite omit maps are calculated by leaving out 3% of the amino acid residues at a time. All superpositions of the structures are carried out using the softwares *O* [366] and *Coot* [367].

2.9 Validation of the model

The last step in protein structure determination is the validation of the model. The crystallographic R-value is an indicator of the model quality, but it has been shown that R_{free} is more reliable. The computer program *PROCHECK* [411] has been used to check the stereochemical and geometrical parameters of the model by comparing them with ideal values obtained from a database of well refined high-resolution protein structures in the Protein Data Bank (PDB) [412]. The checks performed are on covalent geometry, planarity, dihedral angles, chirality, non-bonded interactions,

main chain hydrogen bond, disulfide bonds, stereochemical parameters and residue-by-residue comparisons. It checks the main-chain bond lengths and angles against the ideal values given by Engh and Huber analysis [379] of small molecule structures in Cambridge Structure Database [413]. The computer program *Coot* [367] uses interactive graphics outputs of various plots and a residue-by-residue listing, providing an assessment of the overall quality of the structure and solvents highlighting regions which may need further investigation. The Ramachandran plot [414] of the structure is given by both the programs. Most of the residues in a protein model should lie within the allowed regions of the Ramachandran plot. *CNS* is also used to analyze geometrical parameters and list rms deviations in bond lengths, bond angles, dihedral angles, short contacts between atoms etc. All the structures were validated by using ADIT prior to PDB submission [412]. The figures were made using softwares *Pymol* [415] and *Chimera* [416]. Real-space correlation coefficients of the final refined models provide the accuracy of the positions of the atoms within the positive electron density maxima [417]. More precisely the SA omit electron density map is correlated to the positions of the refined atoms and are used in real-space correlation coefficients calculations. Automatic fitting into the electron density, building the atomic model and refinement are incorporated in current programs viz. *Coot*, *Phenix-Autobuild*, *ARP/wARP*, etc. [367, 375, 417-420].

The atomic coordinate errors can be found by inversion of least-squares full matrices provided an atomic or ultra-high resolution protein structure is present [378]. But when normal resolution structures are available, the coordinate errors can be estimated by using Cruickshank's empirical formula (Equation 21) for an atom with $B = B_i$ [421].

$$\sigma_i (r, B_{avg}) = k (N_i / p)^{1/2} . [g (B_i) / g (B_{avg})] . C^{-1/3} . R . d_{min} \quad (21)$$

where k is 1.0, $N_i = \sum Z_j^2 / Z_i^2$, p is difference between number of observations and parameters, B_{avg} is the average B for fully occupied sites, C is the fractional completeness of the data to d_{min}, and R is refinement R-factor. $g(B) = 1 + a_1 B + a_2 B^2$ is an empirical function to allow for the dependence of $\sigma_i(r)$ on B. The parameters a_1 and a_2 depend upon the structure [422]. The $\sigma(r)$ is called diffraction-component precision index (DPI). This formula may be used to estimate the errors in the hydrogen bonding distances in a refined protein structure. In all the figures in subsequent chapters, 3.4 Å was used as the hydrogen bond length cutoff. The next chapter shall provide specific experimental details of HIV-1 protease preparation, data collection and structure of complex with a type-2 substrate.

CHAPTER 3

X-ray snapshots of HIV-1 protease in action through crystal structures of HIV-1 protease/type-2 substrate complexes

3.1 Introduction

The HIV-1 protease is an important target for drugs against AIDS [16, 17, 83, 94] and knowing, at the atomic level, how the active enzyme interacts with true substrates and with the reaction intermediate resembling the transition state, is critical to structure-based drug design. Many crystal structures of complexes of an active HIV-1 protease with inhibitors and substrate-analogs have been reported [126, 131-138, 159-161]. Several attempts have been made earlier to determine structures of active HIV-1 and SIV proteases complexed with substrate oligopeptides [138, 155]. However, HIV-1 protease was found to be complexed with only the P-product peptide [138], while SIV protease was found to be complexed with only the Q-product peptide [155]. No structures with both products bound simultaneously were obtained. The discovery of closed flap HIV-1 protease, even when unliganded, enabled the possibility of soaking substrates [99, 423-425] to trap both products into crystals of active enzyme.

Since the catalytic activity of HIV-1 protease is pH dependant, exhibiting a bell-shaped pH-activity curve (Chapter 1, Figure 10) [143], HIV-1 protease/substrate complexes were prepared at different pH values with a view to structurally map different stages of the cleavage process. In this chapter, I will present structure analysis of a type-2 cleavage site decapeptide substrate complexed to HIV-1 protease at low, normal and high pH values.

3.2 Methods

3.2.1 The HIV-1 protease gene

Escherichia coli strain BL21 (DE3) cells, with plasmid carrying HIV-1 protease tethered dimer gene (HP) in pET11a vector (hereafter plasmid will be called pET11a-HP), were used to isolate the plasmid. The plasmid has β-lactamase gene conferring ampicillin resistance and multi-cloning site (MCS) is under the control of T7 RNA polymerase. The pET11a-HP consists of 6.2 kb of DNA. The HP gene translates into a single polypeptide chain pro-HIV-1 protease consisting of two HIV-1 protease subunits, each having 99 residues (residues in subunit A and B are numbered 1-99 and 1001-1099 respectively) and the subunits are tethered by the pentapeptide, Gly-Gly-Ser-Ser-Gly. These linker residues are numbered 100-104. The tethered sequence linking the C-terminal residue of subunit A to N-terminal residue of subunit B is known to form β-hairpin loop [88]. The N-terminus of

subunit A is preceded by 57 amino acids of the transframe region of *gag-pol* region of the polypeptide. The peptide bond between the 57th phenylalanine and N-terminal proline of subunit A corresponds to *TFR-PR* natural cleavage site of HIV-1 protease. Thus the initially over-expressed 260 residue (29 kDa) pro-enzyme yields 203 residue (22 kDa) HIV-1 protease tethered dimer (hereafter called HIV-1 protease) after proteolysis during refolding.

3.2.2 Plasmid isolation

The plasmid was isolated by using QIAprep kit (QIAGEN, Germany). It was based on alkaline lysis of bacterial cells followed by selective adsorption of plasmid DNA onto silica membrane of a column equilibrated in high salt buffer. The plasmid was subsequently eluted in low salt buffer. The purity of the plasmid was determined by agarose gel electrophoresis.

3.2.3 Site-directed mutagenesis

The pET11a-HP plasmid was used for site-directed mutagenesis of the HIV-1 protease gene. The Quick Change multi site-directed mutagenesis kit (Stratagene Ltd., Cambridge, U.K.) was used to make the desired mutations in HIV-1 protease gene. The primer oligonucleotides (consisting of ~40 bases) containing the desired mutations were designed using the Stratagene web service.

The three step method uses a thermal cycling procedure to achieve multiple rounds of mutant strands synthesis. Components of the thermal cycling reaction include a super coiled double-stranded DNA template, two or more synthetic phosphorylated oligonucleotide primers containing the desired mutation and the kit–provided enzyme blend featuring *Pfu Turbo* DNA polymerase and DNA ligase. First the mutagenic primers are annealed to denatured template DNA. *Pfu Turbo* DNA polymerase then extends the mutagenic primer with high fidelity and without primer displacement, generating ds-DNA molecules with one strand bearing multiple mutations and containing nicks. The nicks are sealed by components in the enzyme blend. Parental plasmid is digested by Dpn1.

The HP insert region was sequenced commercially using the T7 promoter and terminator primers. The results were analyzed and confirmed for the desired mutations. The plasmid harboring the mutations was re-inserted into for protein expression in *E.coli* BL21 (DE3) cells.

3.2.4 Expression, purification and crystallization of HIV-1 protease

The *E. coli* BL21 (DE3) cells, carrying pET11a-HP gene, were grown in LB media at 37 ˚C. The

growth media also contained the antibiotic, ampicillin (1 mM), so that only those bacterial cells could be grown that have the desired plasmid. The protease expression was induced with 1 mM IPTG (isopropyl-β-D-thiogalactopyranoside) when the absorbance at 600 nm, measured using a UV–Vis spectrophotometer, reached 0.6-0.7. Two hours after induction, cells were harvested and lysed using ultrasonication. The 29 kDa gene product was found to accumulate in inclusion bodies. The following steps are followed to obtain the pure protein:

1. The frozen cell-pellet from ½ litre bacterial cell culture is thawed in ice. 15 ml of ice-cold lysis buffer, TE (100 mM Tris base, 5 mM EDTA, 1 % Triton X-100, pH 7.5) was added to the cell paste. It was mixed using a telflon pestle on an ice bath and transferred to a 50 ml tube for ultrasonification.

2. The sonicator was programmed to pulse mode, with 1 second sonication and 2 second pause, for one minute duration. The cell suspension was sonicated under ice-cold condition. After pausing a minute, repeated the sonication 2-3 times till a homogeneous suspension results.

3. The homogenate was centrifuged at 10000 rpm using Sorvall SS-34 rotor for 10 minutes at 4 ºC. Discarded the supernatant. A greyish white pellet was obtained.

4. Washed the greyish white inclusion bodies with ice-cold 15 ml TE buffer and sonicated under ice-cold condition for a 30 seconds to make a homogeneous suspension. Centrifuged the suspension at 10000 rpm for 5 minutes at 4 ºC using Sorvall SS-34 rotor. The washings were repeated thrice. The washed inclusion bodies appeared as off-white pellet.

5. To the washed inclusion bodies 3 ml of 67 % cold acetic acid was added to extract the soluble protein. A brief sonication was done followed by centrifugation at 10000 rpm for 10 minutes at 4 ºC. The centrifugate containing the soluble protein was collected in a sterile measuring cylinder. The pellet was stored at -20 ºC.

6. The acidified inclusion bodies were flash diluted with ice-cold sterile water to obtain a final concentration of 2 % - 5 % acetic acid. This solution was dialyzed in a 10 kDa cutoff dialysis bag against 5 litre cold distilled water at 4 ºC. The distilled water was changed and then dialyzed again against 5 litre distilled water overnight at 4 ºC. At this stage small amount of precipitate was noticed and was removed by centrifugation at 12000 rpm for 10 minutes at 4 ºC. The centrifugate was collected to be put back in the dialysis bag.

7. The protein was refolded by dialyzing against cold 3 litre refolding buffer (20 mM Tris/Pipes/Mes, 100 mM NaCl, 1 mM DTT, 10 % glycerol, pH 6.5) for about 6 hours at 4 ºC. The contents were centrifuged at 6000 rpm for 10 minutes at 4 ºC in a Sorvall GS-3 rotor. The centrifugate containing the refolded protein solution was collected. To prevent the oxidation of cysteines, all the steps were performed under reducing conditions by adding either 1 mM DTT or 1 mM 2-mercaptoethanol to the buffers. The absorbance was measured at 280 nm. Protein solution was stored at 4 ºC.

8. The refolded protein solution was concentrated, using a 15 ml 10 kD cutoff concentrator (Millipore), to 2 - 4 mg/ml. Measured the absorbance at 280 nm of both the filtrate and flow-through during concentration to check for leakage of the protein.

Purity was checked by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) analysis on a 10 % gel. After refolding, expected single 22 kDa band was seen, indicating that during refolding 57 residue N-terminal flanking peptide (~7 kDa) was cleaved from 29 kDa precursor protein to yield an active mutant protein of 22 kDa (Figure 16). Most of the product after refolding was a mature protein of 22 kDa as seen in SDS-PAGE. The refolded protein is exchanged with 50 mM NaOAc buffer, pH 4.5 containing 1 mM dithioerythritol (DTT) / β-mercapto ethanol (BME) using 15 ml 10 kDa cutoff concentrator and again concentrated to a final concentration of 3 mg/ml to 8 mg/ml. This protein was used for crystallization.

The enzyme activity of the tethered dimer construct is comparable to that of the native homodimer having two separate chains [88]. Enzymatic assay of the refolded protein is also done using a HIV-1 protease specific chromogenic substrate at 310 nm (Figure 17). The UV-substrate has (*para*-NO_2-Phe) - Leu cleavage site which has absorbance in the UV-Vis region of the electro-magnetic spectrum [426]. The assay conditions are: 3 µM enzyme in 50 mM NaOAc, pH 4.5 and 1 mM DTT / β-mercaptoethanol is mixed with 120 µM UV chromogenic substrate in the same buffer. Stock solution of 5 mM substrate is prepared in 100 % DMSO. The absorbance at 310 nm is measured for few hours showing a gradual decrease at 310 nm (Figure 17). The peak of the UV spectrum also shows a hypochromic shift.

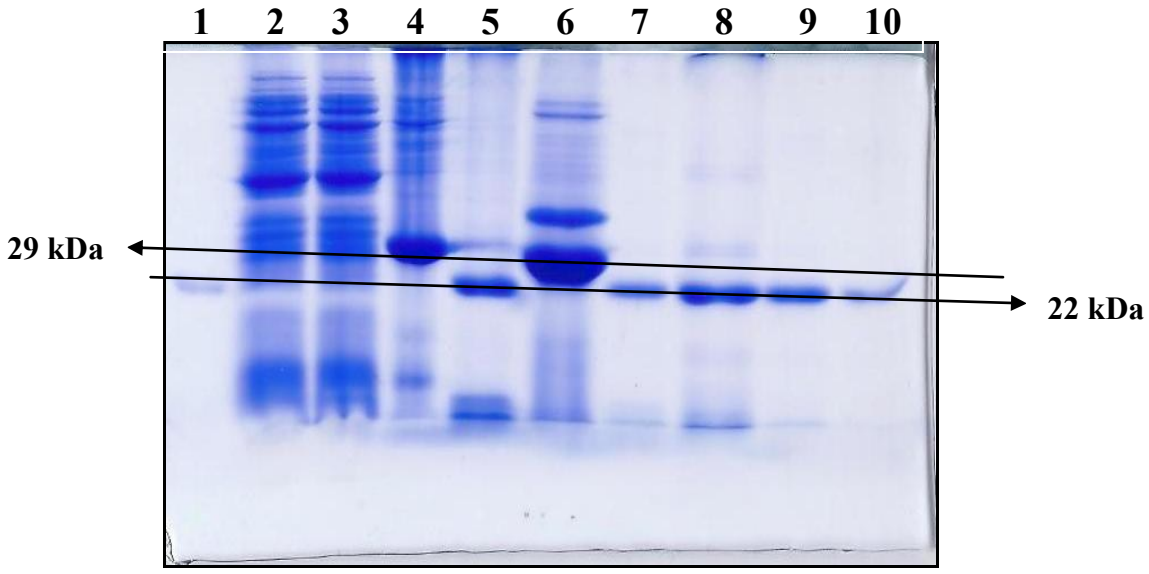


Figure 16: SDS-PAGE bands showing the 29 kDa HIV-1 protease precursor protein and mature 22 kDa HIV-1 protease.

Lane 1: Dilute and completely refolded 22 kDa mature HIV-1 protease.

Lane 2: Un-induced bacterial cell lysate having no IPTG.

Lane 3: Bacterial cell lysate just before induction with 1 mM IPTG.

Lane 4: Bacterial cell lysate induced with 1 mM IPTG showing 29 kDa precursor.

Lane 5: Concentrated and refolded 22 kDa mature HIV-1 protease after 2 hours.

Lane 6: Carbonic Anhydrase, 29 kDa used as protein marker.

Lane 7: Concentrated and completely refolded 22 kDa mature HIV-1 protease.

Lane 8: Concentrated and refolded 22 kDa mature HIV-1 protease after 4 hours.

Lane 9: Concentrated 22 kDa mature HIV-1 protease exchanged with NaOAc buffer, pH 4.5.

Lane 10: Dilute 22 kDa mature HIV-1 protease exchanged with NaOAc buffer, pH 4.5.

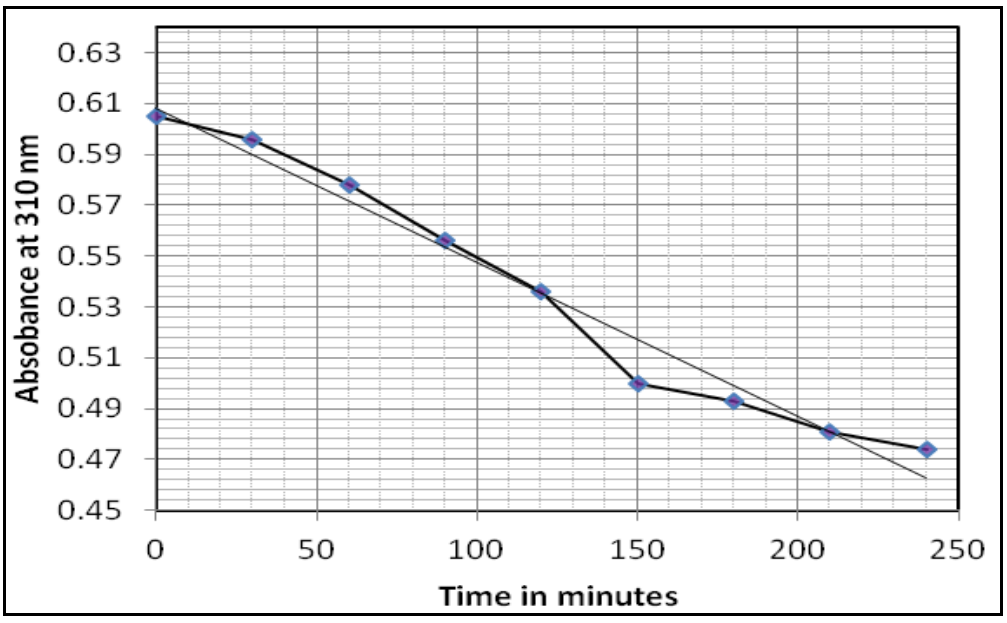


Figure 17: Assay on a chromogenic substrate cleaved by HIV-1 protease monitored by UV-Vis spectrophotometer at 310 nm. The straight line fitted to the data points indicate a linear decrease in the absorbance with time.

3.2.5 Crystallization and soaking experiments

Vapor diffusion method is used to explore and optimize the crystal growth conditions to obtain a diffraction quality crystal [427]. Typically in a hanging drop or a sitting drop vapor diffusion setup, a drop containing the protein solution is mixed with a precipitant solution at concentration insufficient to precipitate the protein. The drop is then equilibrated against a larger reservoir of solution that usually contains higher concentration of precipitant or another dehydrating agent (Figure 18a). As the drop equilibrates with this reservoir solution, the protein starts to reach supersaturating concentrations in the drop, which induces crystallization. High-throughput methods mostly employ sitting drop vapour diffusion method in a 96-well crystallization plate [428-430]. Once a crystallization hit is obtained further optimization to obtain larger crystal size were made using hanging drop vapor diffusion method. A multi-cavity plate was used in manual HIV-1 protease crystallization setups for screening several conditions. Each tray has 24 wells, providing a new crystallization condition to be tested. The protein was concentrated to 5 mg/ml in 50 mM sodium acetate buffer at pH 4.5 containing 1 mM DTT. 2 µl of protein solution and 2 µl of reservoir solution were mixed together and placed on a siliconized coverslip, which was then inverted and sealed with vacuum grease over a well containing 800 µl of the reservoir solution. Glass cover slips (16 mm x 16 mm) were siliconized in-house using 1 % dichlorodimethyl silane in toluene.

Crystals of the HIV-1 protease were obtained (Figure 18b) at 25°C using buffer containing 200 mM sodium dihydrogen phosphate and 100 mM sodium citrate at pH 6.2 and ammonium sulphate precipitant varying between 2 % - 10 %. Crystals appeared from few days to a week and grew in the form of hexagonal rods of size 0.3 x 0.05 x 0.05 mm^3. An additive like glycerol, which stops nucleation, was used and gave fewer, larger crystals [431]. It also has the advantage of doubling as a cryo-protectant [432]. HIV-1 protease cleaves nine specific peptide sequences (Chapter 1, Figure 4). Oligopeptides serve as effective substrates and are amenable to crystallographic analysis. The cleavage sites are classified as type-1 and type-2 [433]. When one of the amino acid flanking the scissile peptide bond is proline then it is designated as a type-1 cleavage site and without proline it is called type-2 cleavage site. The crystals of HIV-1 protease were soaked in the solution of substrate peptides corresponding to all the nine natural cleavage sites. Some of these peptides were initially

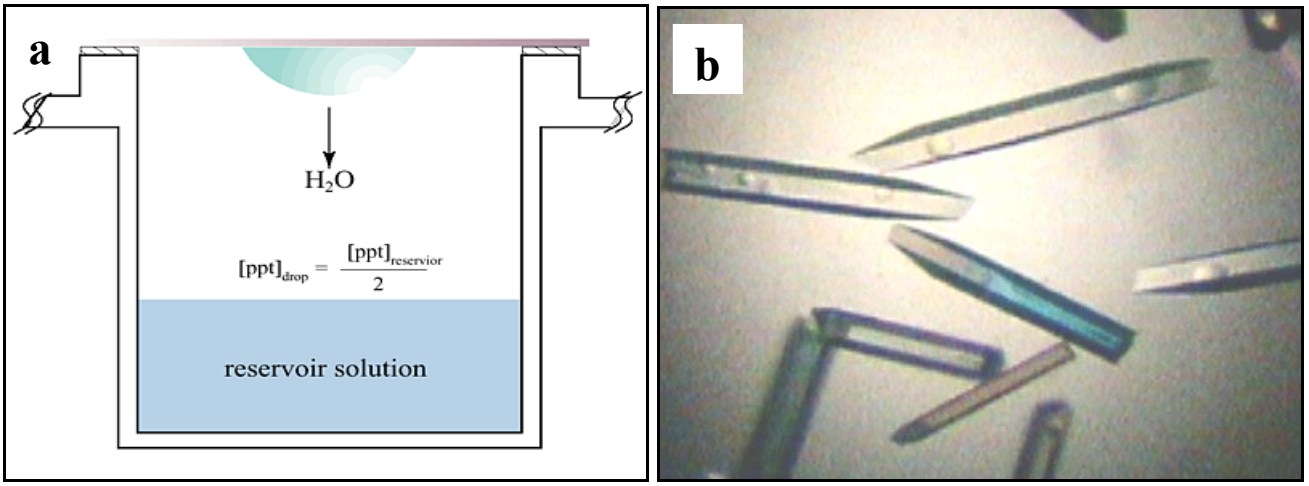


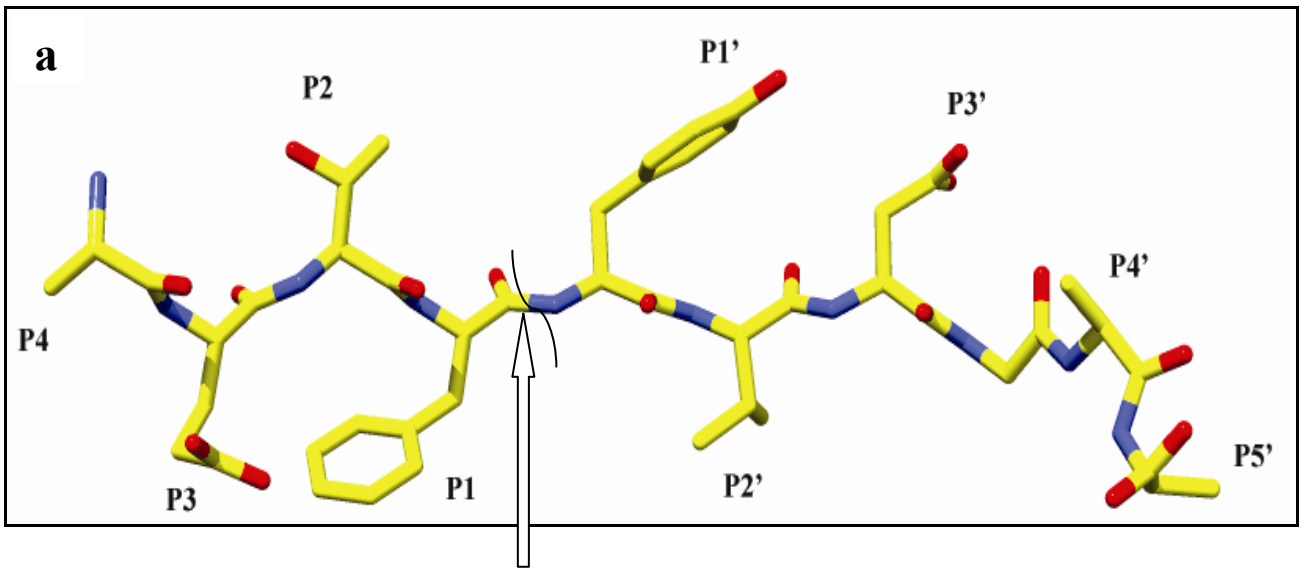
Figure 18: a) Hanging drop vapour diffusion method. **b)** Hexagonal rod shaped crystals of HIV-1 protease grown by hanging drop vapour diffusion method.

synthesized by our collaborator at the National Institute for Research in Reproductive Health, Parel, Mumbai using an Applied BioSystem Peptide Synthesizer and later all the 10-11 residue peptides were commercially purchased. Lyophilized powder of the peptide was dissolved in 10 % acetic acid or in reservoir buffer (0.2 M sodium dihydrogen phosphate and 0.1 M sodium citrate) to prepare a 5 mM solution. For soaking experiment, this stock solution was diluted ten fold into the reservoir solution to prepare 10 µl soaking droplet which was placed on a coverslip. Into this drop, native protease crystals were transferred using a nylon cryo-loop and the coverslip was inverted and sealed over the same reservoir well where crystals had appeared.

To ascertain if, during soaking, any enzyme was released from the crystal into the soaking solution, a UV spectroscopic protease-assay was used to determine the status of the substrate in the soak-droplet. Four to five crystals of size eventually used for data collection were soaked into a 10 µl drop made of the soaking solution mentioned above, but without the substrate. At the end of 72 hours, the crystals were still intact, and only the solution was extracted from the drop and added to a reaction mixture containing *p*-nitrophenyl chromogenic substrate at a concentration of 95 µM. Absorption spectra were recorded daily at regular intervals for over four days using Jenway 6500 UV spectrophotometer. The absorbance at 310 nm only showed minor fluctuations around the initial reading of 0.510, and did not decrease linearly as would happen if the *p*-nitrophenyl containing chromogenic substrate were being cleaved by any enzyme present. Further, the absorption peak also did not shift to lower wavelengths as would happen if there were cleavage reaction taking place inside the cuvette. Therefore it is believed that no enzyme was released from the crystals into the soaking solution, and that the

decapeptide substrate is not cleaved outside the crystals in the droplet. This observation is also consistent with absence of any change in the appearance of the crystals on soaking. Several crystals were soaked in the substrate peptide solutions by employing the following methodology: varying the chemical condition of the soaking solution (pHs ranging from 2.0 to 9.0) and/or duration of the soak (from 6 hours to 7 days). The crystals were left for incubation in the peptide solutions for various soak periods at room temperature, before harvesting them for data collection under cryo-conditions. Crystalline complexes of HIV-1 protease/oligopeptide substrates were successfully prepared where all the nine substrates and their variants were used as ligands. The crystals were monitored every 6 hours for disintegration or dissolution and/or development of cracks upon soaking. Many crystals lost their crystallinity or disintegrated at extremes of pHs and at longer soak times.

Results with a 10-residue peptide, AETF*YVDGAA (where the '*' represents the peptide bond that is cleaved), (Figure 19a) that acts as a substrate for HIV-1 protease are described here. The amino acid sequence of the decapeptide substrate used corresponds to the RT-RH junction of the viral *gag-pol* polyprotein. Figure 19b shows the standard nomenclature used to designate the side chain residues of intact substrate as Pn and the corresponding subsites or pockets on the protease as Sn, where n= 1, 2, … , etc. The numbering starts from the cleavage site and Pn' or Sn' are used for C-terminal side. When the substrate is cleaved, two fragments are generated, the N-terminal fragment is designated as P-product and the C-terminal as Q-product.



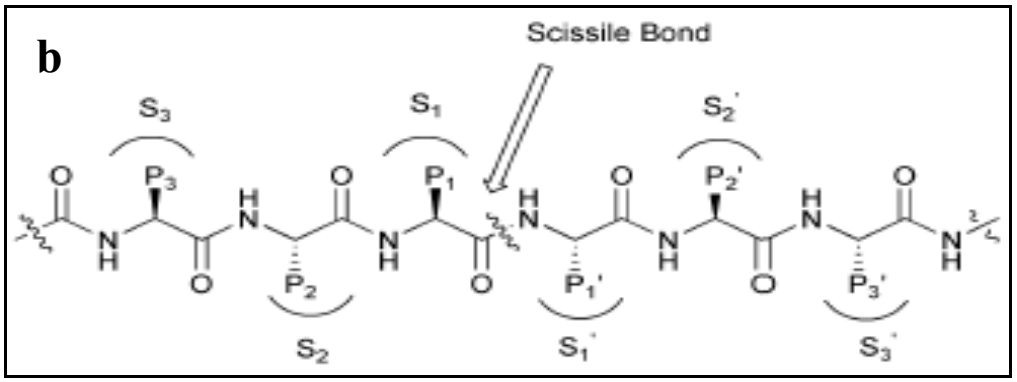


Figure 19: a) 10-residue type-2 peptide substrate (H_2N-AETF*YVDGAA-COOH) of HIV-1 protease that was used for soaking experiments. * denotes the cleavage site. The scissile peptide bond between the Phe (P1) and Tyr (P1') residues is indicated by the wavy line.
b) Standard nomenclature P1, _ _ _ , Pn and P1', _ _ _ _ , Pn' is used to designate amino acid residues of peptide substrates. The corresponding binding sites on the HIV-1 protease are referred to as S1, _ _ _ Sn and S1', _ _ _ _ Sn' subsites.

3.2.6 X-ray Data Collection and processing

X-ray diffraction data on each of the enzyme/substrate complexes were collected, under cryo-condition on various beamlines at European Synchrotron Radiation Facility (ESRF), Grenoble, France. The radiation damage to the protein crystals at high-energy synchrotron sources is reduced considerably by flash freezing the crystals to cryogenic temperatures [434, 435]. This often also improves the microscopic crystal order resulting in stronger high-resolution diffraction. However, in most cases, flash freezing increased the mosaicity, the macroscopic crystal disorder. To prevent formation of ice crystals and subsequent shattering of the crystals, different cryo-protectants are introduced into most macromolecular crystals [436]. The cryo-conditions were optimized in order to minimize radiation damage to the crystals by using many cryoprotectants like glycerol, ethylene glycol, ammonium sulphate, oils, etc. [437, 438]. Glycerol was found to give good diffraction spots (Figure 20). At the end of peptide soaking at room temperature, the crystals were equilibrated for 1-5 minutes in the cryoprotectant solution (25 % glycerol and 75 % reservoir buffer) before flash freezing in liquid nitrogen at 77K. The protein crystals were screened at several pHs ranging from 2.0 to 9.0 and for different soak times, by recording the X-ray diffraction data on each one of the soaked crystals using the BM30A beamline at ESRF [439]. A robotic arm which can mount the crystals on the goniometer from a carousel, holding the cryo-vials, was used for screening 20 crystals at a time. Annealing was tried on many crystals with high mosaicity, and ice rings and it was found that the mosaicity and ice rings significantly lessened. Two images were collected at 0^o and 90^o, each for

1^{o} oscillation and 10 seconds of exposure. It was found that most crystals soaked at extremes of pH and for less than 12 hours or more than 3 days showed poor diffraction quality or low substrate occupancy. Most of the crystals did not survive the whole data collection and showed spot spreading with gradual disappearance of high resolution spots. The crystals which diffracted to better than 2.0 Å resolution were preserved at 77K in liquid nitrogen. High-resolution diffraction dataset on the chosen crystals were then collected at 100K on the undulator beamlines at ESRF.

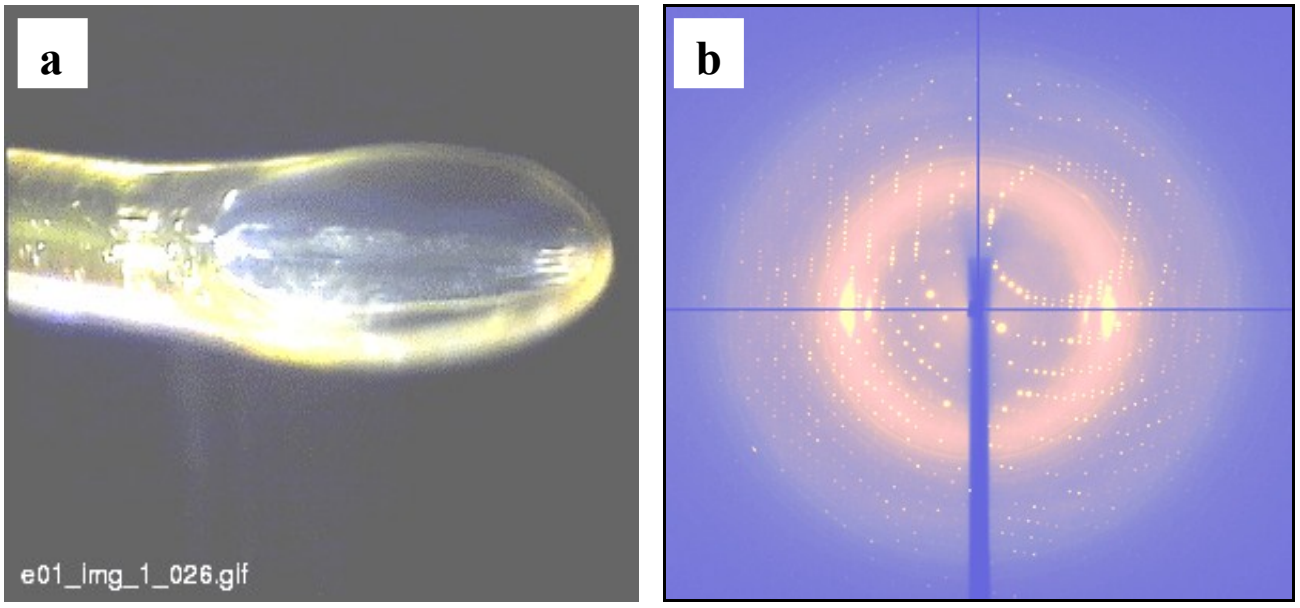


Figure 20: a) A cryoprotected HIV-1 protease crystal on a litholoop during data collection at 100K. **b)** The diffraction frame of a HIV-1 protease crystal.

The exposure times used were 0.8 seconds to 2 minutes depending upon the incident beam intensity of the beamlines used. The detector distance was varied in order to get best possible resolution and the exposure time was selected such that there were few overloads in each frame. Few data collections could be completed by collecting 100-120 oscillation frames, each for an oscillation range of 1^{o}, which could be processed. Each data set was separately processed and analyzed using *XDS,* and scaled using the program *XSCALE* implemented in *XDS* [253]. All the crystals belonged to $P6_1$ space group (no. 169). Crystal and intensity data statistics for three HIV-1 protease/RT-RH junction oligopeptide substrate complex structures at pHs 2.5, 6.2 and 8.0 are given in Table 4.

Table 4: Reflection data statistics of the HIV-1 protease/RT-RH substrate complex at three pHs.

Parameters	pH 6.2	pH 2.5	pH 8.0
Resolution range (Å)	50.0-1.65	50.0-1.76	50.0-1.89
Unit cell parameters (Å)	a=b=62.03, c=81.78	a=b=62.56, c=81.86	a=b=62.89, c=82.80
X-ray wavelength (Å)	0.97945	0.97945	0.87260
Total no. of data frames	120	120	107
Total no. of reflections	112519	128747	78110
No. of unique reflections	20957	17226	14825
Completeness (%)	97.5 92.8 (1.75-1.65)*	95.3 88.5 (1.81-1.76)*	99.0 89.6 (1.99-1.89)*
Mosaicity (°)	0.201	0.213	0.230
R_{merge} # (%)	8.1 60.2 (1.75-1.65)*	7.2 46.5 (1.81-1.76)*	9.9 45.3 (1.99-1.89)*
<I/σ(I)>	10.16 2.90 (1.75-1.65)*	21.76 2.98 (1.81-1.76)*	11.80 3.20 (1.99-1.89)*

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

\# $Rmerge(I) = \left(\sum_i \sum_j |I_j - \langle I \rangle| \right) \Big/ \sum_i \sum_j \langle I \rangle$

where I_j the intensity of the jth observation of reflection I, <I> is the mean of the intensities of all observations of reflection i, the summation Σ_i is taken over all reflections and the summation Σ_j is taken over all observations of each reflection.

3.2.7 Structure solution and refinement

The structures were solved by the Molecular Replacement (MR) method [287]. Unliganded tethered dimer of HIV-1 protease (PDB ID: 1LV1) obtained from Protein Data Bank [412] was used as a search model after removing all the solvent molecules. When one of the amino acid residues, Phe of the substrate was being treated as a tetrahedron, it was modified by conversion of the carbonyl carbon atom into a *gem*-diol carbon atom. *CNS* [347, 440] parameter and topology files for this modified Phe residue, designated as PHD or PHT, were generated either manually or by using the *PRODRG* server [441]. The PHD residue was renamed as HPH (as phenylalanine diol) while submitting in PDB. The force constants in the parameter file for bonds and angles involving the scissile peptide bond were set to comparatively low values so that the actual geometry is not tightly restrained, but is dictated by the diffraction terms. In the final stages of refinement, *PHENIX* [375] and *CCP4* [376] were used for TLS [407, 408] refinement on the protein model. When comparing different complex structures, only Cα atoms of the protein were used in the calculation of the superposition matrices [366, 367, 442].

Figure 21 shows the preliminary electron density map of the HIV-1 protease β–sheet region, calculated after molecular replacement and a cycle of restrained refinement.

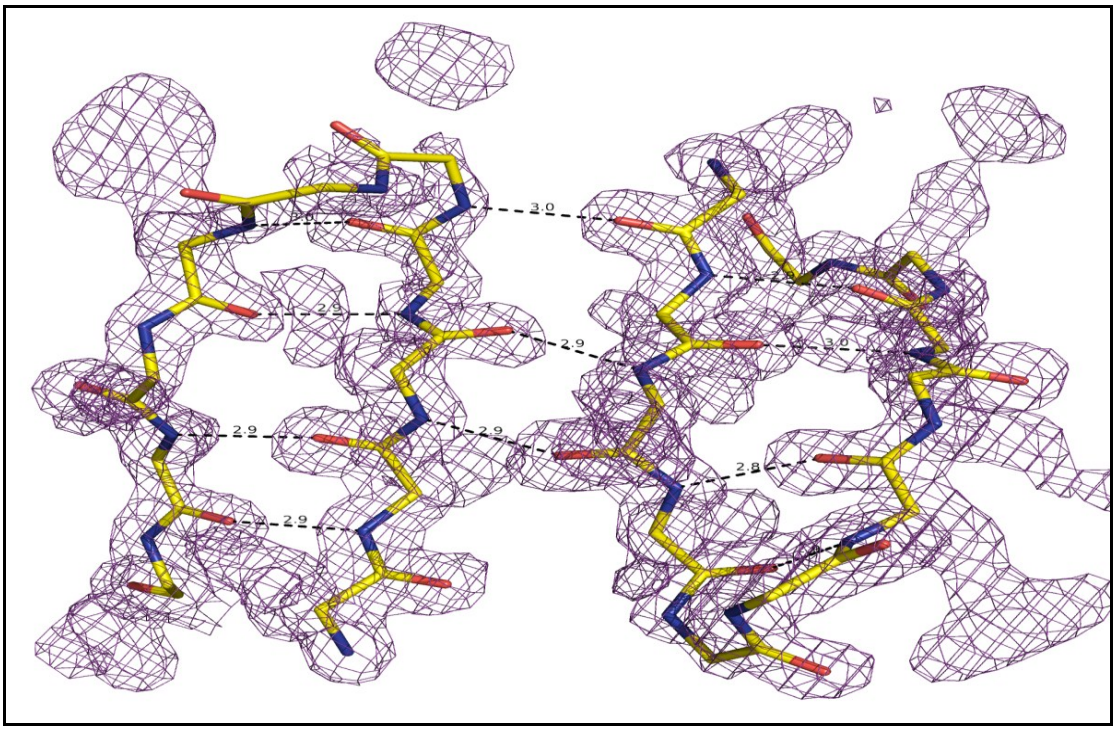


Figure 21: The β–sheet region after structure solution using 1LV1 model along with the *2mFo-DFc* map, after one cycle of restrained refinement.

3.3 Results and discussion

In the next sections, I will discuss the HIV-1 protease/substrate complex structures solved as a function of pH and soaking times.

3.3.1 HIV-1 protease/substrate complex at pH 6.2

Connected positive maxima of the *mFo-DFc* omit map (Figure 22) [369, 397-399] in the active site region indicated presence of the substrate. Electron density for P2, P1, P1' and P2' residues of the peptide AETF*YVDGAA, were clearly visible at a contour level of 2.0 σ. The shapes of omit densities at P1 and P1' were different, and were characteristic of Phe and Tyr residues respectively. Nevertheless, because of the pseudo two fold symmetry of the free enzyme active site, the substrate was modeled in two orientations and treated their occupancies as variables during initial crystallographic refinement. It was found that the occupancy of one orientation increased from the originally assigned value of 0.5 to 1. Absence of any difference electron densities corresponding to the second orientation further confirmed single orientation of the substrate in the active-site. Similar single orientation has been observed in structures of decapeptide substrates complexed with the inactive D25N mutant of HIV-1 protease [130]. The molecular motif refined was a 1:1 complex

between the HIV-1 protease and the substrate in a single orientation. The initial calculated electron density map without the substrate in the active site could not unambiguously indicate whether the substrate is cleaved or not between P1 and P1'. Thus three different types of linkage between P1 and P1' residues were considered: 1) normal *trans*-peptide linkage, 2) hydrated peptide linkage in which the sp^2 hybridized scissile carbonyl carbon atom becomes sp^3, forming tetrahedral carbon bonded to two hydroxyl oxygens, and 3) a cleaved peptide without scissile CO-NH linkage, in which the carbonyl is converted into a carboxylate group.

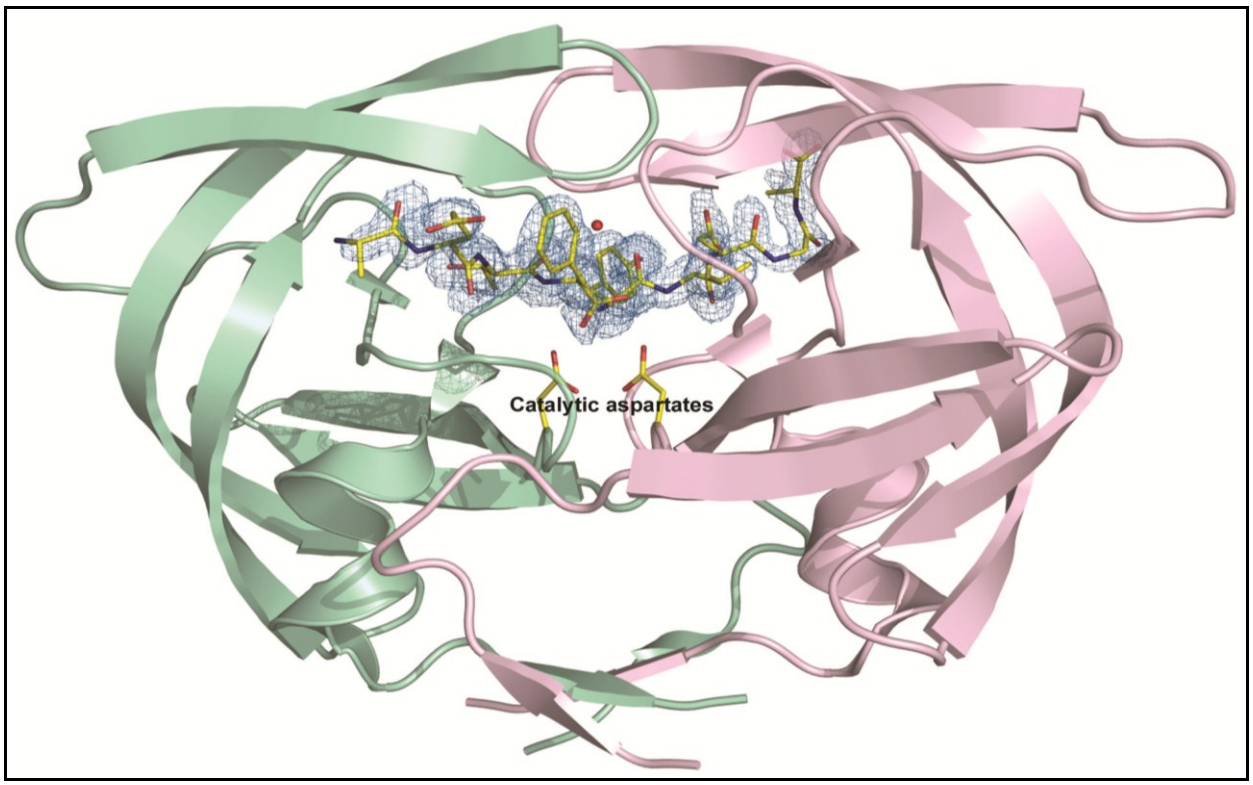


Figure 22: The protein backbone cartoon diagram is shown, where the two subunits are coloured green and pink. Simulated annealed *mFo-DFc* difference electron density in the active-site cavity of HIV-1 protease is shown as blue mesh, where the substrate model is omitted. The refined substrate model is shown along with catalytic aspartates as yellow sticks.

The complex prepared at pH 6.2 is refined to 1.65 Å and the refinement statistics for the three models is given in Table 5. The R_{work} values for the three models are very comparable, while the R_{free} is lowest for the cleaved peptide model. The R_{free} for the tetrahedral model was higher, 26.0 % as compared to that of 25.9 % , for cleaved peptide model.

The stereochemistry of the refined model is good as seen from the r.m.s. deviations of bond lengths, angles and dihedrals. The Ramachandran plot shows that 99 % of the model lies within the allowed region. Figure 23a gives the *2mFo–DFc* and *mFo–DFc* maps for the P1-P1' region, when the substrate AETF*YVDGAA, is modeled, at the scissile bond, as an uncleaved peptide with standard

Table 5: Refinement statistics comparing three states of the substrate peptide model at pH 6.2.

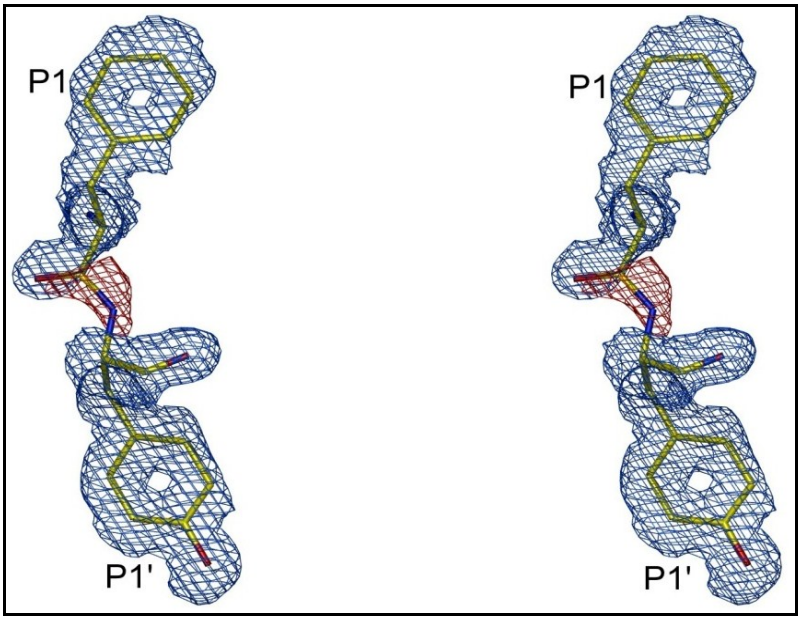
Ligand model refined :	Regular	Cleaved	Tetrahedral intermediate
Resolution range (Å)	50.0 – 1.65	50.0 – 1.65	50.0 – 1.65
R_{work} (%)	21.4 (30.2) #	21.4 (30.2) #	21.5 (30.2) #
R_{free} (%) *	26.3 (31.4) #	25.9 (31.5) #	26.0 (31.6) #
Number of protein atoms	1515	1515	1515
Number of solvent atoms	156	156	156
Number of ligand atoms	69	70	70
r.m.s. deviation of bond lengths (Å)	0.014	0.015	0.014
r.m.s. deviation of bond angles (°)	1.56	1.55	1.54
r.m.s. deviation of dihedral angles (°)	15.0	15.0	15.0
Average B factor (Å^2) for			
Protein atoms	31.2	31.1	31.1
Substrate atoms	52.4	53.3	50.9
Water atoms	47.2	47.2	47.2

** The R_{free} was calculated using 5 % of random reflections which were kept apart from the refinement throughout.*

 # Highest resolution bin is 1.693 – 1.65Å.

peptide geometry. It is seen from *2mFo–DFc* map that, even when included in structure factor calculations, there is no density along the *trans* peptide bond. Further, the scissile peptide bond is in the negative region of the *mFo-DFc* map, suggesting that the bond is either broken or does not have the standard *trans* conformation. Figure 23b gives the SA omit map for the full substrate overlaid with the three models. Real space R-factor and correlation coefficients calculated over the substrate, modeled in the three different ways, are listed in Table 6.

a



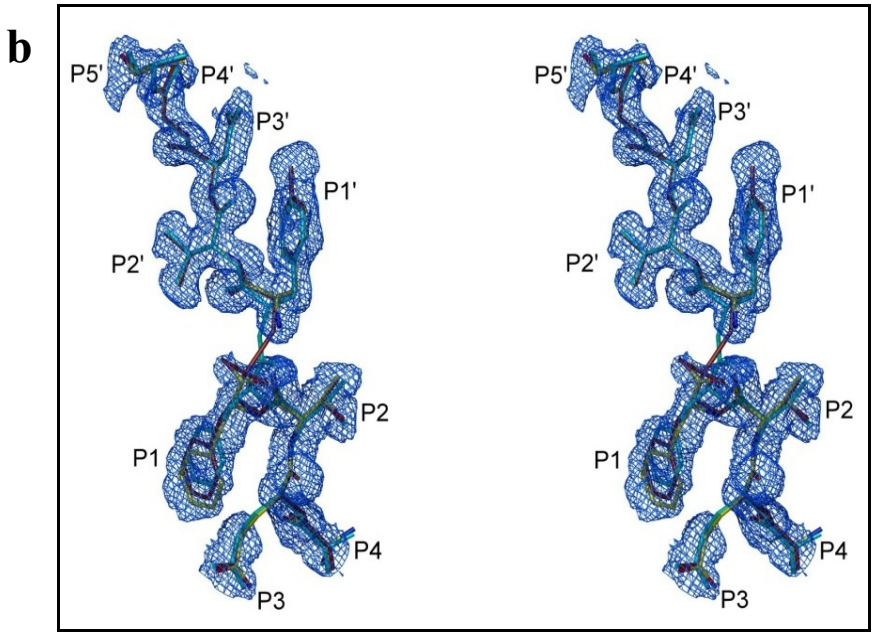


Figure 23: **a)** Stereo diagram of *2mFo-DFc* map in blue and *mFo–DFc* map in red when the substrate model refined is of a regular peptide. The peptide bond is in negative density of *mFo–DFc* map. **b)** Stereo diagram showing SA omit map overlaid with the three models: regular peptide model (cyan), tetrahedral hydrated peptide model (brown) and cleaved peptide model (yellow). Electron density for P1, P3, P1' and P3' residues defines the single orientation of the substrate.

Table 6: Real-space correlation coefficients for nine substrate residues in the refined regular and tetrahedral peptide models. The correlations are between the experimental (ρ_o) and model (ρ_c) electron densities.

Residue no.	Correlation coefficient		
	Cleaved	Regular	Tetrahedral intermediate
1	0.573	0.536	0.483
2	0.599	0.528	0.585
3	0.706	0.646	0.720
4	0.591	0.570	0.605
5	0.727	0.651	0.715
6	0.840	0.840	0.849
7	0.699	0.696	0.680
8	0.607	0.627	0.595
9	0.482	0.529	0.470
<All res>	0.647	0.625	0.634

The correlation coefficient averaged over all the nine residues, <All res> is higher for the cleaved peptide model as indicated in Table 6. Similarly, the R-factor is the lowest for the cleaved peptide model. The structure therefore is of a complex between HIV-1 protease and the two product peptides. The refined coordinates are deposited in Protein Data Bank having ID 2NPH.

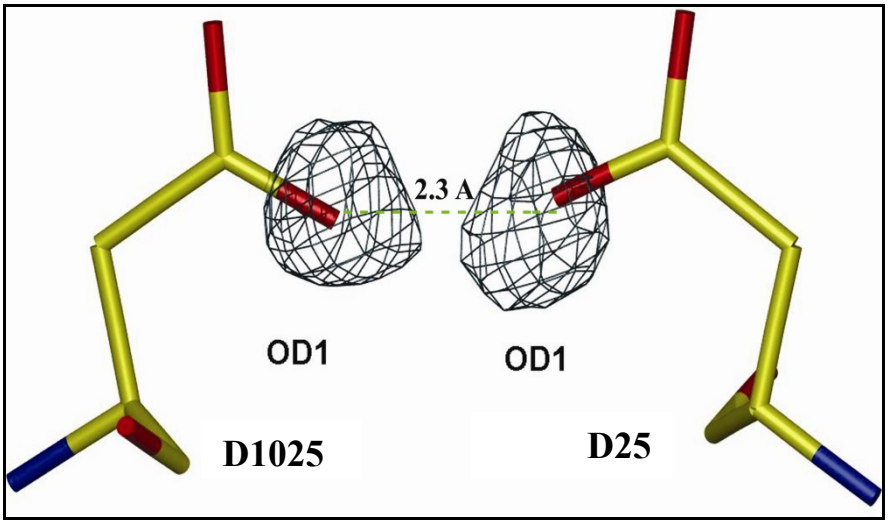


Figure 24: SA omit map to show positions of inner oxygens of catalytic aspartates. The difference density in the *mFo–DFc* map is contoured at 3.0 σ level.

The Figure 24 shows the SA omit map of the two inner oxygens of the catalytic aspartates. The oxygen atoms lie within the omit map indicating the accurate positioning of these oxygens. The two oxygens are separated by 2.30 Å, indicating a low barrier hydrogen bond (LBHB) between them [168, 443]. The O2 oxygen of the newly formed carboxyl group superposes to within 1.6 Å (Figure 25) to the water molecule in the native structure (PDB ID: 1LV1) [374] which is hydrogen bonded to catalytic aspartates. The other oxygen, O1 is 2.5 Å from the water molecule. Therefore O2 could have been derived from catalytic water molecule, while the other carboxyl oxygen, (O1), would be the peptide oxygen.

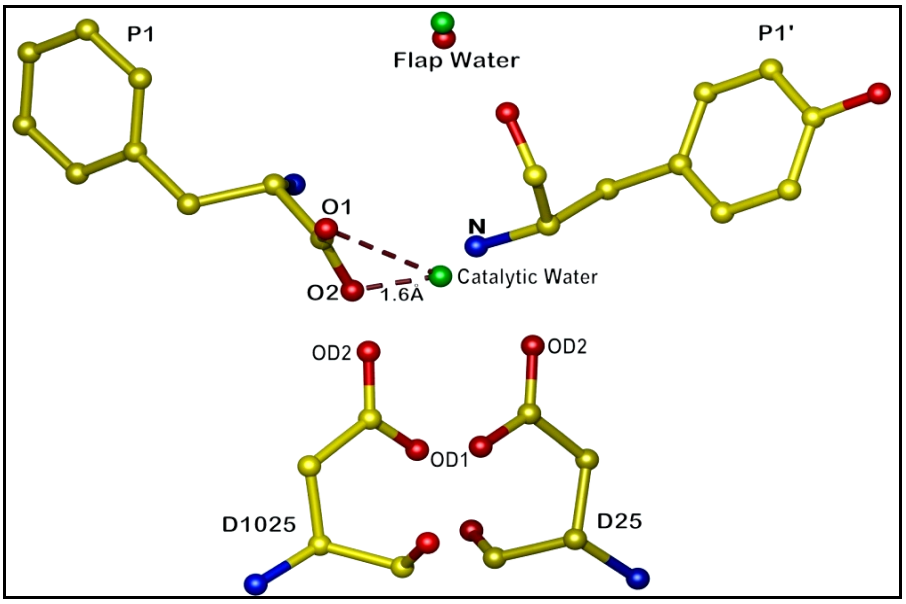


Figure 25: The superposed structures of apo-enzyme with the bi-product complex (yellow stick). The apo-enzyme catalytic water molecule (green sphere) is only shown. Note that the catalytic water is closer to the O2 oxygen of the P-product.

3.3.1.1 Hydrogen bonding in the active site cavity of the cleaved peptide/HIV-1 protease complex

The substrate adopts extended β-sheet conformation in the active site of HIV-1 protease (Figure 26). The flaps of the protein are in closed conformation. The standard errors in the hydrogen bonding distances were estimated using equation 21 (Chapter 2) and ranged from 0.09 Å to 0.127 Å. Substrate is bound by virtue of sixteen hydrogen bonding interactions. Twelve hydrogen bonds involve the peptide backbone of substrate product fragments with the protein and two of them are with flap water (Figure 26). There are two hydrogen bonds involving the OD2 oxygens of catalytic aspartates and the cleaved fragments (O1 of P1 residue and N of P1' residue) as shown in Figure 27. The remaining two hydrogen bonds are between (P3 Glu) OE2 - NH2 (R1008) and (P3' Asp) OD1- NH2 (R8). There are five conserved structural water molecules in the HIV-1 protease active site making extensive hydrogen bonds. Flap water molecule makes four hydrogen bonds having tetrahedral geometry anchoring the two flaps and the substrate P2 and P1' carbonyl oxygens. The (I1050) N - - - O (Flap Water) hydrogen bond (3.4 Å) is very weak or even absent (Figure 26). There are four water molecules, below the P2/P2' residues in each subunit, which make multiple hydrogen bonds with each other and with R8/R1008, T26/T1026, G27/G1027, D29/D1029 and R87/R1087 (Figure 26).

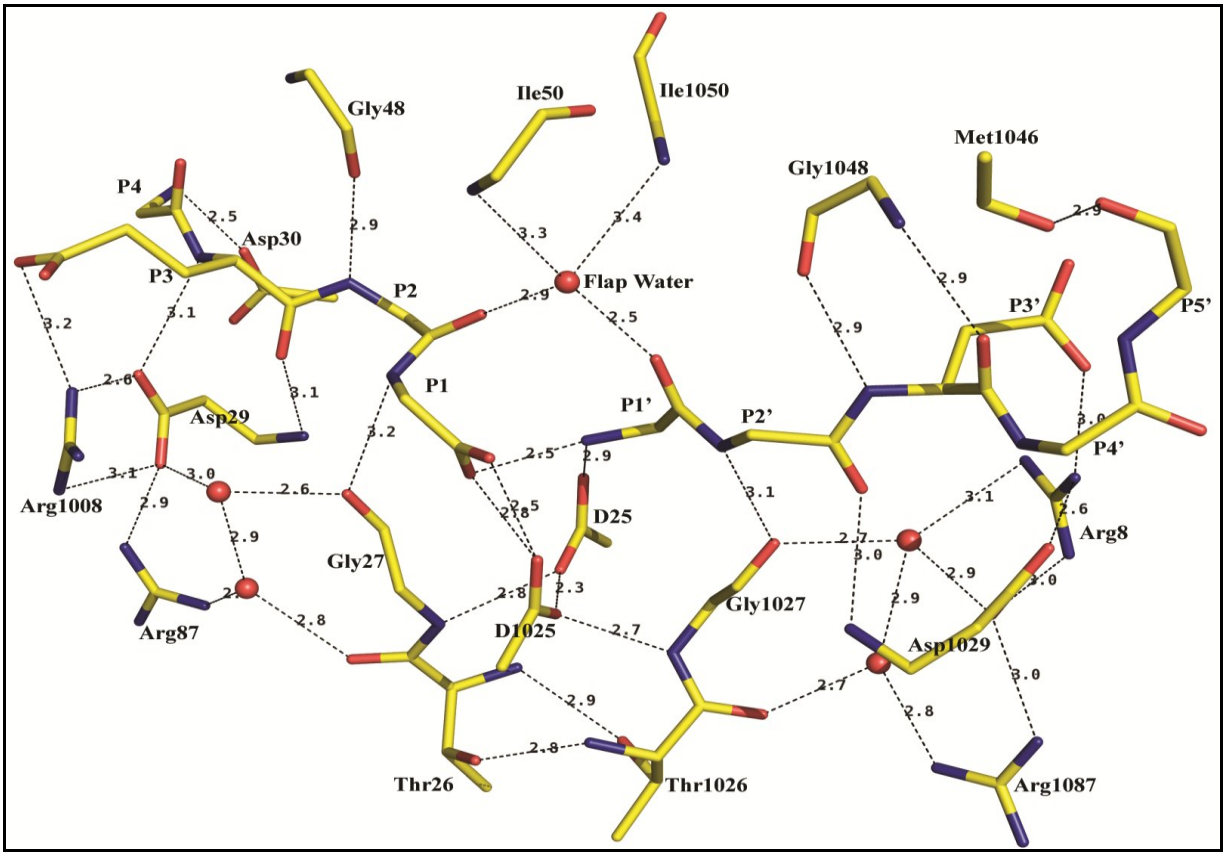


Figure 26: Hydrogen bonding interactions in the active site of the cleaved substrate/HIV-1 protease complex at pH 6.2. Side chain atoms of the substrate, except P3Glu and P3'Asp, are not shown for clarity. Five structural water molecules are also shown as red spheres.

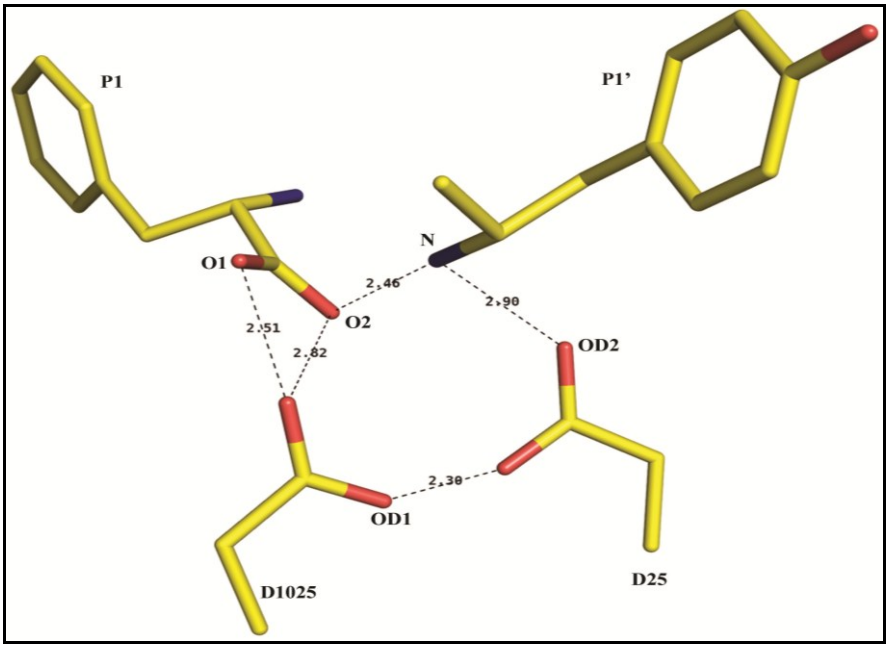


Figure 27: Hydrogen bonding interactions at the catalytic center are shown by dotted lines. O1 and O2 are the carboxylate oxygens of the Phe residue (P1) and the N is the amino nitrogen of the Tyr residue (P1'). Inner oxygens of the catalytic aspartates (D25 and D1025) are labeled as OD1 and the outer oxygens facing the substrates are labeled OD2.

The inner oxygens (OD1) of D25 and D1025 make a low barrier hydrogen bond (2.3 Å) and the carboxylate planes of the two aspartates are almost coplanar (deviation is only 23°). The P- and Q-products interact through (P1) O2 - - - N (P1') hydrogen bond (2.5 Å) (Figure 27).

3.3.1.2 The protonation state of aspartates in the bi-product complex

The identification of the protonation states of the catalytic aspartates in HIV-1 protease is of great interest both in understanding the reaction mechanism and in guiding the design of drugs against HIV/AIDS. At a resolution of 1.65 Å, using X-rays, the hydrogen atom positions cannot be determined and in their absence, a hydrogen bond is inferred from interatomic distance considerations of the heavier atoms. As shown in Figure 27, four hydrogen bonds are formed involving P1, P1', D25 and D1025 residues. The carboxyl oxygen (O1) of P1 forms a hydrogen bond (2.51 Å) to OD2 of D1025. The other carboxyl oxygen (O2) is hydrogen bonded (2.46 Å) with the scissile N on the Q-product. This N-atom forms a second hydrogen bond with the OD2 atom of D25 (2.90 Å) and the inner oxygens (OD1) of the catalytic aspartates are also engaged in a short 2.30 Å hydrogen bond with each other. Such short hydrogen bonds, called LBHB [104, 168, 443], are also seen in other proteins [444]. The hydrogen bonds at the catalytic center will help in determination of the ionization state of the aspartates. There are a total of four hydrogens (one from the monoprotonated catalytic aspartates in the free enzyme, two from the lytic water molecule and one from the scissile peptide NH

group), whose positions need be determined in the present complex. One hydrogen atom has to be shared by the inner oxygens of the catalytic aspartates having LBHB. Since the peptide bond is broken, there has to be two hydrogen atoms bonded to nitrogen of tyrosine as primary amine, NH_2. The fourth hydrogen atom is located between the P-product carboxyl O1 oxygen and OD2 oxygen of D1025 having a strong hydrogen bond (2.51 Å). This hydrogen atom could be attached either to the aspartate oxygen or to the carboxyl oxygen. If it is attached to the aspartate oxygen, the aspartates will be diprotonated and the P-product carboxyl group will be anionic. If the proton is on the carboxyl oxygen, O1, the P-product will be neutral and the aspartates will be monoprotonated. In order to distinguish between these two scenarios an analysis of X-ray structures of HIV-1 protease/inhibitor complexes is done. There is a correlation between protonation state of the aspartates and their hydrogen bonding pattern [145] showing that whenever the aspartates are monoprotonated, they form strong hydrogen bonds to donor groups from the substrate/inhibitor or to water molecules. In contrast, in the diprotonated state, there are no strong hydrogen bonds from the aspartates. Therefore, according to this hypothesis, the catalytic aspartates in the present structure should be monoprotonated. The scissile peptide oxygen, O1 is seen in a strong hydrogen bond with OD2 D25 in the bi-product complex structure. Further, based on maximum orbital overlap criteria, the linearity of hydrogen bonds formed were analyzed, when the hydrogen atom is placed in a standard hybrid orbital of the donor atom. The hybridization of the donor atom is determined by the geometry of its binding to other non-hydrogen heavier atoms, whose positions are known accurately. Such an approach has been shown to give very accurate predictions for positions of polar hydrogen atoms in protein structures [445, 446]. Thus a hydrogen atom is more probable on an atom if the angle subtended by this atom to the hydrogen bonded atom and adjacent atom is close to the tetrahedral angle. The C - O1 - OD2 (D1025) and O1 - OD2(D1025) - C(D1025) angles are 99.6° and 152.4° and their deviations are 9.9° and 42.9° respectively. Thus the hydrogen atom is located on the O1 atom to give P1 carboxylic acid.

3.3.1.3 The post-cleavage reaction intermediate

Since soaking does not result in cleavage of the substrate outside the crystal, the whole substrate would diffuse into the active-site through the solvent channels of the crystal. Earlier work had shown that the active-site tunnel is quite wide, and can be accessed through the solvent channels of the

crystal by molecules as large as acetyl-pepstatin [447], and an undecapeptide [425]. Thus the cleavage reaction has happened inside the crystals, and both the products generated *in situ* are expected to be within the enzyme active-site. Thus the structure presented here is the first report of an aspartyl protease with both cleavage products, P and Q, bound in the active-site. The shorter than van der Waals separation between scissile C and N atoms (2.67 Å) is an indication of the reaction stalled at an intermediate stage. Complete separation and release of products, P and Q, requires loss of hydrogen bonding and other interactions between the products and the enzyme. The residues from the flap and the catalytic aspartates are involved in a majority of these interactions, and opening of the flexible flaps, therefore, could disrupt these hydrogen bonds. However, since flap-opening is not possible in the present crystals due to intermolecular contacts, the species are left in the intermediate stage following the bond cleavage event. The detailed interatomic interactions in this state would give, by extrapolation, an insight into the reaction-mechanism.

Figure 28a shows the comparison with a regular peptide bound to inactive D25N mutant enzyme [130]. This mutant complex structure is assumed to represent the binding of the substrate in the active-site before the onset of the reaction. It may be seen that, just after bond cleavage, the scissile nitrogen and carboxyl group are found closer to and forming hydrogen bonds with the catalytic aspartates. This conformational rearrangement is achieved without making large changes in the positions of the side chain atoms inside subsites, especially on the C-terminal side. The hydrogen bond between the O1 and the aspartate OD2 is also present in the inactive complex and is suggested to be playing a crucial role in proper positioning of the substrate. Figure 28b shows the superposition with the P-product complex prepared by Rose *et. al*., by co-crystallization [149]. There are significant deviations between superposed Cα atom pairs for residues P1, P2 and P3, the maximum shift (1.45 Å̊) being for the P1 residue. The direction of movement is toward the non-existent Q-product, rather than away from it. In the present structure, the enzyme has just cleaved the peptide bond but the product is not yet fully separated for product release. The possible reason for this is the hindrance introduced because of crystal packing of the substrate molecules in the adjacent unit cells. Another reason is the mobility of the flaps of HIV-1 protease is reduced upon substrate binding hindering any large scale movement. The flaps are dynamic which can open and close in the solution. Within the crystals, mobility of the flaps is arrested as reflected from a well-ordered electron density for the flaps. Due to

this rigidity introduced in the enzyme structure, products are not released from the enzyme's active site cavity.

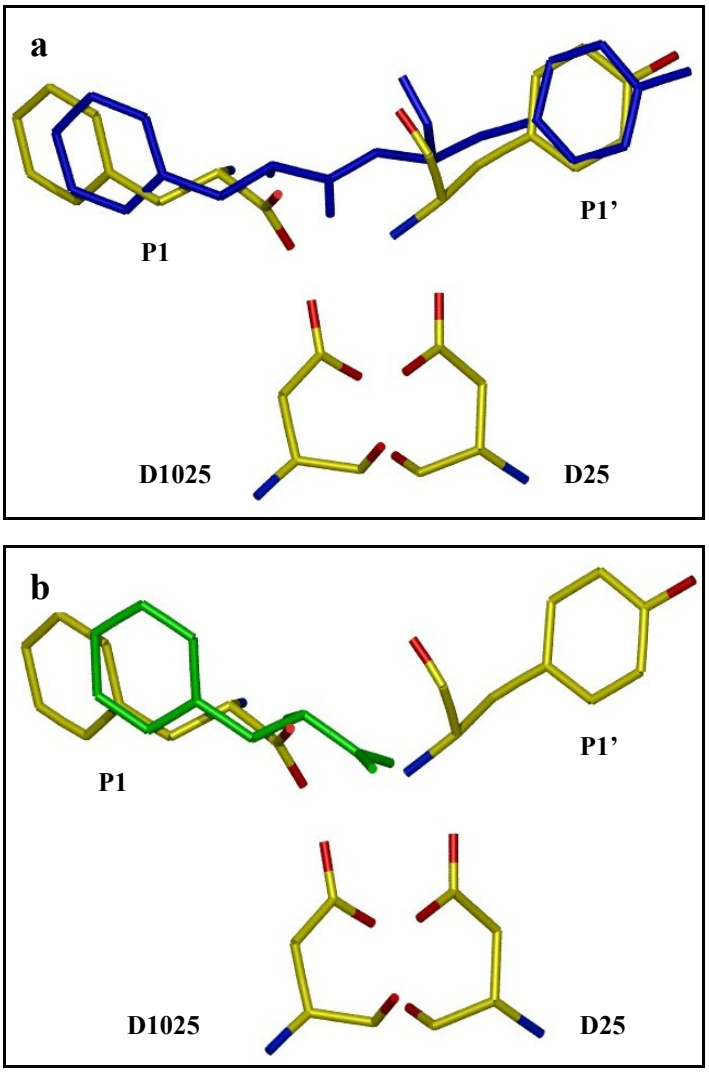


Figure 28: Superposition of the cleaved peptide/active HIV-1 protease structure (yellow) with structures (rmsd 0.75 Å) representing **a)** the regular peptide/inactive HIV-1 protease D25N mutant complex (blue) and **b)** product release stages of the cleavage reaction with the P-product before release (green) where only the P1 residue of the P-product peptide is shown.

3.3.2 HIV-1 protease/substrate complex at pH 2.5

The structure of the HIV-1 protease/substrate complex prepared by soak time of 1 day and at pH 2.5 will be discussed in this section. The same substrate peptide, AETF*YVDGAA was used for soaking. Again three possibilities for the scissile peptide bond were considered during refinement. The difference electron density in the HIV-1 protease active site was continuous in the region of scissile peptide bond, and therefore possibility of a cleaved peptide was ruled out in this case. The pseudo 2-fold symmetry of the active site and symmetrical nature of the difference electron density indicated a possibility of 2-fold related orientations of the substrate. The substrate was modeled in two 2-fold related orientations having the peptide bond as both regular and tetrahedral (TI), each with occupancy

of 0.5. Six substrate residues, ETFYVD could be modeled within the difference electron density. Figure 29 shows the stereo picture of SA omit map for the P1-P1' region overlaid with refined P1 and P1' residues of regular peptide and TI models. In Figure 29, there is no electron density along the C-N bond, linking P1 and P1' residues, in the regular peptide model ruling out its possibility. Further, continuous density near the scissile peptide atoms is unaccounted in the regular peptide model. On the other hand, the C-N peptide bond in the TI model is placed well inside the omit density. The SA

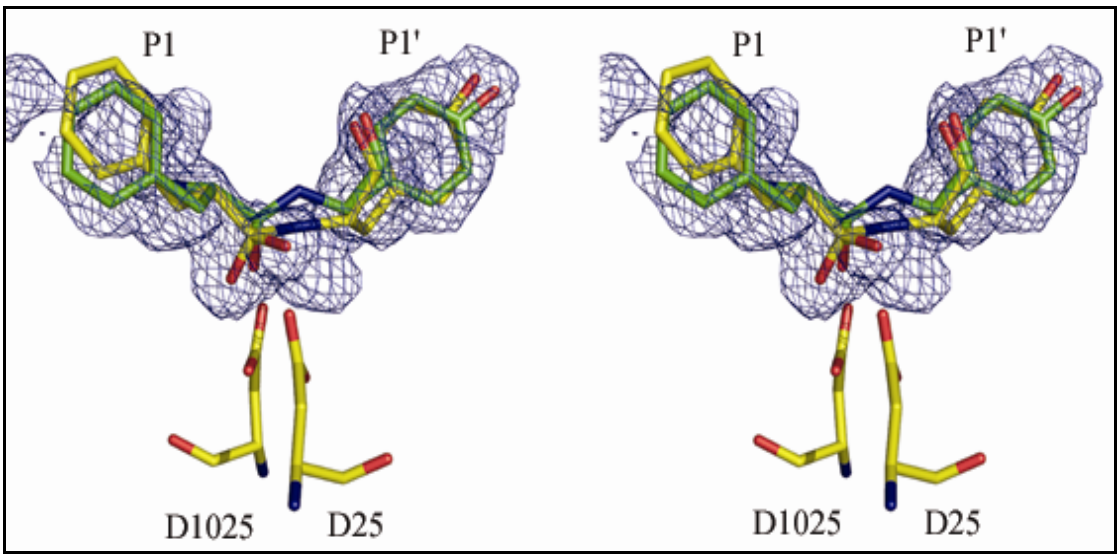


Figure 29: Stereo view of SA omit map contoured at 2.0 σ, when the substrate model is not included in map calculation. The refined tetrahedral intermediate, TI, (yellow sticks) and the regular peptide (green sticks) models are shown for comparison. Note that the scissile nitrogen atom of the regular peptide model is not inside electron density. Only the major orientation is shown for the sake of clarity.

omit electron density map is correlated to the positions of the corresponding refined models, and was used in real-space correlation coefficients calculations. Real-space correlation coefficients are listed in Table 7 for six substrate residues, modeled either as a TI or as a regular peptide.

Table 7: Real-space correlation coefficients for the six residues in the refined regular peptide and TI models of the substrate.

Residue no.	Correlation coefficient	
	Tetrahedral intermediate	Regular
2 (P3)	0.680	0.694
3 (P2)	0.783	0.789
4 (P1)	0.842	0.752
5 (P1')	0.831	0.824
6 (P2')	0.766	0.760
7 (P3')	0.854	0.849
<4 & 5>	0.837	0.788
<All residues>	0.793	0.778

The correlation coefficients averaged over either all the six residues or over only the two residues, P1 and P1', is higher for the TI model. The molecular motif refined further was therefore a complex between the HIV-1 protease and the TI in two orientations.

Occupancy refinement of the substrate models converged to values of 0.51 and 0.32 for the two orientations giving a total occupancy of 0.83. Water molecules have been placed in the remaining 17 % of active sites where the enzyme is in apo form. In all molecular superpositions only the major orientation is shown in stick representation. Table 8 gives details of the refinement statistics. The refined coordinates and the structure factors are deposited at Protein Data Bank, and are assigned ID 3MIM.

Table 8: Refinement statistics of the TI/HIV-1 protease complex.

Ligand model refined:	**Tetrahedral intermediate**
Resolution range (Å)	50.0 – 1.76
R_{work} (%)	22.28 (28.65)#
R_{free} (%)*	25.00 (29.41)#
Number of protein atoms	1515
Number of solvent atoms	194
Number of ligand atoms	56
R.m.s. deviation of bond lengths (Å)	0.012
R.m.s. deviation of bond angles (°)	1.99
Average B factor (Å^2) for	
Protein atoms	31.5
Substrate atoms	56.2
Water atoms	52.5

** The R_{free} was calculated using 5% of random reflections which were kept apart from the refinement throughout.*
 # Highest resolution bin is 1.80 – 1.76 Å.

Figure 30 shows stereo overlays of two SA omit maps, one calculated by omitting the two oxygen atoms bound to the scissile carbon, and the other calculated by omitting the four Oδ atoms of the two catalytic aspartates. These maps unambiguously confirm the refined positions of the respective atoms. Thus the species present in the active-site cavity at pH 2.5 is a TI, consistent with the omit electron density maps and the refinement data. Hence, the interactions deduced from these coordinates are reliable. Figure 31 shows molecular fit in the SA omit map, of the major occupancy TI model for residues P2-P2'. The electron density beyond P2 and P2' is weak suggesting disorder of these substrate residues. To explore which of the two oxygens O1 and O2 in the TI is derived from the

attacking water molecule, I have compared the present structure with two other structures. The structure published by Torbeev *et. al.*, is a complex of a non-cleavable keto-methylene inhibitor with active HIV-1 protease (PDB ID: 3DCR) [154]. The structures superposed with rmsds 0.52 Å. In their complex with active HIV-1 protease, the inhibitor keto group was found to be hydrated into a TI. The additional oxygen atom bound to scissile carbon in the active complex is inferred to be the

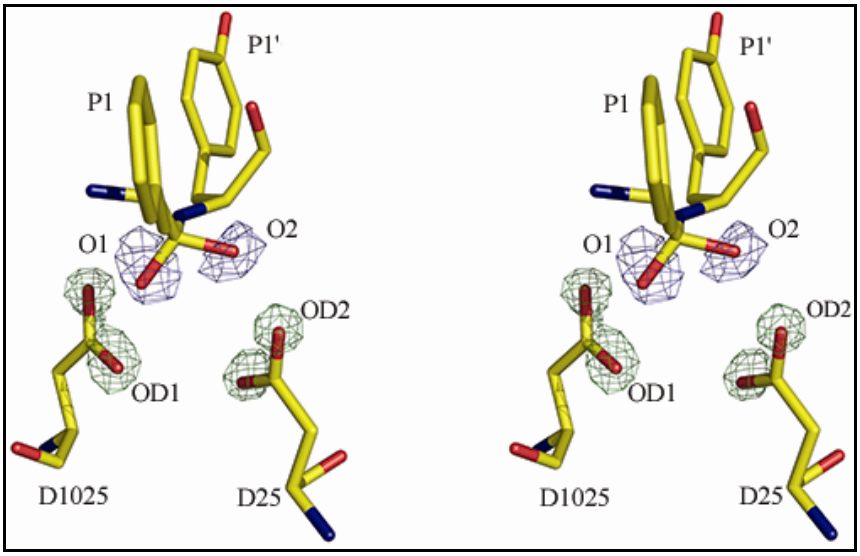


Figure 30: Stereo view of SA omit map (blue) contoured at 2.4 σ when the two oxygen atoms in the major occupancy TI are not included in the calculation of structure factors. Also shown is the stereo view of SA omit map (green) contoured at 5.2 σ when the four oxygen atoms of the catalytic aspartates are omitted from structure factor calculations.

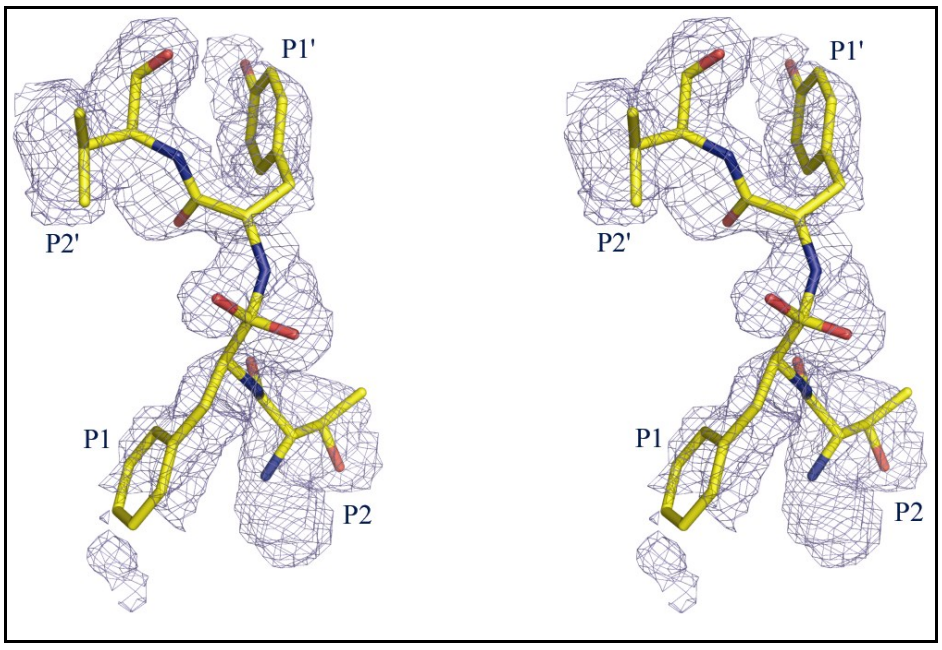


Figure 31: Stereo view of SA omit map contoured at 2.0 σ along with the refined TI model shown in the major orientation.

water-derived hydroxyl oxygen in the TI. This hydroxyl oxygen superposes onto the O2 oxygen atom of the present structure to within 0.7 Å. Further, O2 is closer to the catalytic water molecule (1.2 Å) which hydrogen bonds to active site aspartates in unliganded HIV-1 protease (Figure 32). The second

structure superposed on the present structure is the structure of a complex between inactive HIV-1 protease and the oligopeptide corresponding to the RT-RH cleavage site [PDB ID: 1KJG] [130]. The O1 atom is in close proximity (0.45 Å) to the scissile carbonyl oxygen (Figure 32). Thus the O1 atom in the present structure is likely to be derived from the scissile carbonyl oxygen atom of the substrate. Identification of these oxygen atoms is crucial from the point of view of understanding the molecular mechanism of HIV-1 protease.

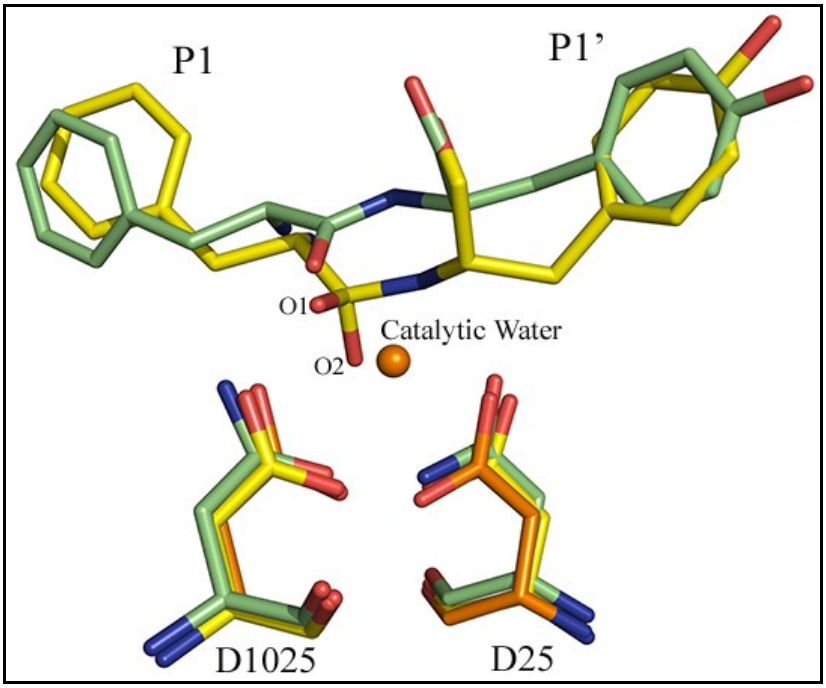


Figure 32: Comparison of catalytic centre from three superposed structures: 1) inactive HIV-1 protease/regular peptide complex, [PDB ID: 1KJG] (light green), 2) present structure (yellow) and 3) unliganded HIV-1 protease [PDB ID: 2G69] (orange). Note the proximity of O2 to the putative catalytic water in unliganded structure, and of O1 to the peptidyl oxygen in the inactive D25N complex.

3.3.2.1 Hydrogen bonds in the active-site of the TI/HIV-1 protease complex

Figures 33 and 34 show the hydrogen bonding interactions observed in the active site cavity of the present structure. The hydrogen bonding distance cutoff used was 3.4 Å. The standard errors in the hydrogen bonding distances ranged from 0.08 Å to 0.12 Å [423]. The TI forms nine hydrogen bonds to protein atoms, either directly or through water molecules (Figure 33). The TI interacts with the carboxyl oxygens of catalytic aspartates in an asymmetric way: O2 forms two hydrogen bonds with both oxygen atoms of D25, while O1 forms only a single hydrogen bond with the outer oxygen atom of D1025 (Figure 34). The hydrogen bond between O1 of TI and OD2 of D1025 is very short (2.2 Å). The scissile nitrogen atom is at a distance of (3.2 Å) from OD2 atom of D25 which is longer than a hydrogen bond. The inner oxygens of the catalytic aspartates are 3.1 Å apart.

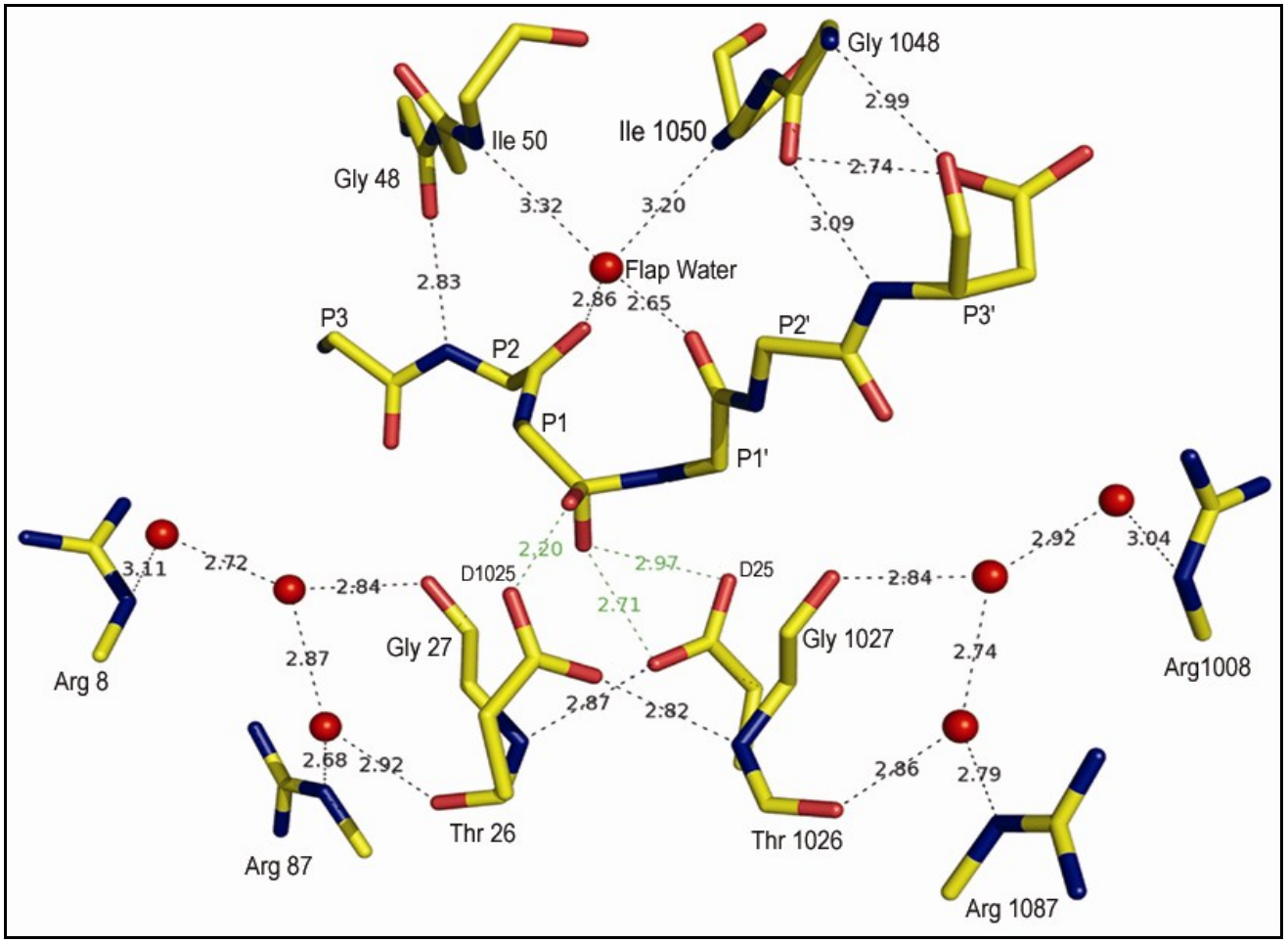


Figure 33: Hydrogen bonding network in the active site cavity of TI/HIV-1 protease complex. The interactions of the *gem*-diol hydroxyls with the catalytic aspartates are shown in green. Only main chain atoms are shown for clarity, except for side-chain atoms of P3', D25/1025, R8/1008 and R87/1087.

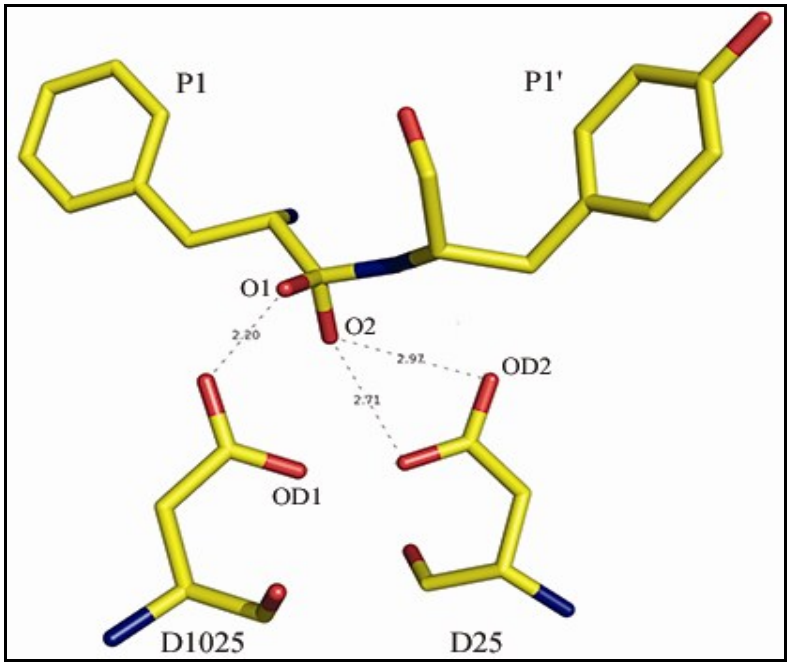


Figure 34: Hydrogen bonds between aspartates and *gem*-diol of the TI at the catalytic center. This hydrogen bonding pattern has not been observed before.

3.3.2.2 The protonation state of the aspartates in TI/HIV-1 protease complex

The protonation state can be different at different steps in the enzyme's catalytic cycle. It has also been recognized that the nature of the ligand bound in the active site is an important determinant of the protonation state of the aspartates. For this reason, protonation states derived from enzyme complexes with true substrates rather than with non-scissile substrate isosteres, are likely to be much more relevant. It is well known that the effective pKa values of titratable groups in proteins are influenced by the microenvironment, electrostatic and hydrogen bonding interactions and solvent accessibility [448]. In a computational study on HIV-1 protease/substrate oligopeptide complex, the pK_a value of one of the aspartates is estimated to be 2.3 while that of the other was much higher, when all solvent molecules and detailed atomic charges are explicitly included in the calculations [449]. Since the pH used in the present study is higher than 2.3, one aspartate is likely to be deprotonated in the present structure. Based on pH rate profile studies for four oligopeptide substrates and two competitive inhibitors Hyland *et al.* concluded that the substrates bind only to a form of HIV-1 protease in which one of the two catalytic aspartyl residues is protonated [143]. Since there are strong hydrogen bonds between TI and catalytic aspartates in the present complex, the aspartic dyad is in a monoprotonated form. A similar correlation between monoprotonation of aspartic dyad and its strong hydrogen bonding interactions with inhibitor molecules has been reported [156]. Support for monoprotonated dyad at pH 2.5 is also provided by experimental studies involving ^{13}C-NMR spectroscopy on HIV-1 protease complexed with inhibitor molecules Pepstatin A [450] and KNI-272 [451]. These authors found that throughout the range pH 2.5 – 6.2, D25 is protonated and D1025 is deprotonated. As before the protonation states were inferred from the hydrogen bond distances and angle criteria. Since D1025 is involved in SIHB, its OD2 is ionized. Thus D25 has to be protonated to give a monoprotonated aspartic dyad. To infer which of the two oxygens of D25 is protonated, the bond angles were analyzed. There are two hydrogen bonds from D25: O2 --- (D25 OD2) and O2 --- (D25 OD1). In the first hydrogen bond O2 is the donor atom as inferred above, and therefore (D25 OD2) is an acceptor. In the second hydrogen bond, the angles were (Cγ D25) - (D25 OD1) - O2 = 101^o and C - O2 - (D25 OD1) = 145^o. The former angle is closer to the average value of 112.2^o for the C – O - H angle in hydrogen-bonded carboxyl groups determined by neutron diffraction [452]. So placing a proton on (D25 OD1) would lead to a more linear OD1 - O2 hydrogen bond. This deduction

is also consistent with the fact that O2 is already a donor in the hydrogen bond with (D25 OD2). Therefore it is concluded that in the monoprotonated catalytic dyad the proton is bonded to (D25 OD1). Thus there is no proton on the outer oxygen atom of D25 at this stage of the reaction, and this fact has important implications toward the reaction mechanism.

3.3.2.3 The tetrahedral intermediate

The bond lengths and angles around the tetrahedral *gem*-diol carbon are very comparable to standard values of a tetrahedron. This indicates no significant strain in the TI. The inter-atomic distance between the scissile N and C atoms is 1.44 Å, while the bond angles around the scissile carbon atom are as follows: Cα – C – N = 122.8^o, O2 – C – O1 = 113.2^o, O2 – C – N = 99.7^o, O1 – C – N = 94.0^o, Cα – C – O2 = 113.3^o and Cα – C – O1 = 112.1^o. The tetrahedron about the scissile carbon is thus slightly distorted as has been observed in other TI/HIV-1 protease structures [153, 154]. The torsion angle around the C-N bond is 106^o as compared to the value of 180^o in the normal *trans*-peptide.

A common feature among short OHO hydrogen bonds between atoms of dissimilar pKa values is localization of negative charge on the acceptor atom [170, 453-458]. The 2.2 Å hydrogen bond between TI O1 and D1025 OD2 is thus a “Short Ionic Hydrogen Bond” (SIHB), with the negative charge located on D1025 OD2, because of the intrinsically higher pKa value of alcoholic OH (O1). This would make the O1 atom of TI to be protonated. To deduce the protonation state of O2 the linearity of hydrogen bonds formed was analyzed, when the hydrogen atom is placed in a standard hybrid orbital of the donor atom. The hybridization of the donor atom is determined by the geometry of its binding to other non-hydrogen atoms, whose positions are known accurately. Such an approach has been shown to give very accurate predictions for positions of polar hydrogen atoms in protein structures [445, 446]. The relevant angles for the hydrogen bond O2 - (D25 OD2) (Figure 34) are: (Cγ D25) - (D25 OD2) - O2 = 89^o, and C - O2 - (D25 OD2) = 117^o. Since the latter is closer to the tetrahedral value and the former is too acute, a hydrogen atom on O2 results in a more linear O2 - (D25 OD2) hydrogen bond leading to greater orbital overlap. This conclusion is also consistent with the fact that O2 is the oxygen derived from the attacking water molecule. Thus the TI is formed just after nucleophilic attack by a water molecule and is a neutral *gem*-diol. This inference is in agreement with quantum mechanical/molecular mechanics (QM/MM) calculations on the nature of the reaction intermediate in HIV-1 protease catalysis [459]. The neutral *gem*-diol in the TI is also

consistent with the fact that there is no oxyanion-binding pocket in any reported HIV-1 protease structures.

3.3.2.4 Comparison of the TI and product complexes of HIV-1 protease

Figures 35 and 36 show the comparison of active site region of the TI/HIV-1 protease complex with the bi-product complex [460] showing the atomic rearrangement that accompanies collapse of TI into the cleavage products. There is a change in the conformation of P1' residue as evidenced by a substantially different position for the Cβ atom in the TI. There is a new hydrogen bond between the carboxyl group of P3' Asp and G1048 carbonyl oxygen atom (d= 2.74 Å) in the low pH structure. This hydrogen bond requires protonation of the P3' Asp residue in the substrate (Figure 33). Shifts in the positions of enzyme residues have been very small, of the order of 0.2–0.5 Å (Figure 35). The most significant change has been in the side-chains of catalytic aspartates. The virtual dihedral angle OD2(D1025) - OD1(D1025) - OD1(D25) - OD2(D25), which is a measure of the co-planarity of the two carboxyl groups of catalytic aspartates, has changed from 67^{o} in the TI/HIV-1 protease complex to the more coplanar value of 23^{o} in the product/HIV-1 protease complex. These conformational changes have led to alterations in hydrogen bond interactions from catalytic aspartates. For example, in the TI/HIV-1 protease complex, the inner oxygen atoms do not form any hydrogen bond whereas they form a LBHB in the product complex. The separation of 2.3 Å observed in the bi-product/HIV-1 protease structure (Figure 27) is the shortest observed so far, and therefore is the first report of a LBHB in the active-site of HIV-1 protease. It is perhaps significant that this LBHB is observed when the ligand is a true substrate rather than a substrate analog. Similarly in the TI/HIV-1 protease structure the 2.2 Å distance between the *gem*-diol, O1 and (D1025 OD2) is the shortest inter-molecular hydrogen bond observed so far and classified as a SIHB. Similarly, both *gem*-diol oxygen atoms form hydrogen bonds to catalytic aspartates in the TI/HIV-1 protease complex, whereas one carboxyl oxygen atom in the product complex does not make any hydrogen bonds with catalytic aspartates. It is also clear that in going from TI stage to product stage, it is the P1 residue that shows variation in its position while the P1' residue has not altered its position (Figure 36), despite there being, in TI/HIV-1 protease complex, no hydrogen bond to anchor in place the scissile N atom of P1'.

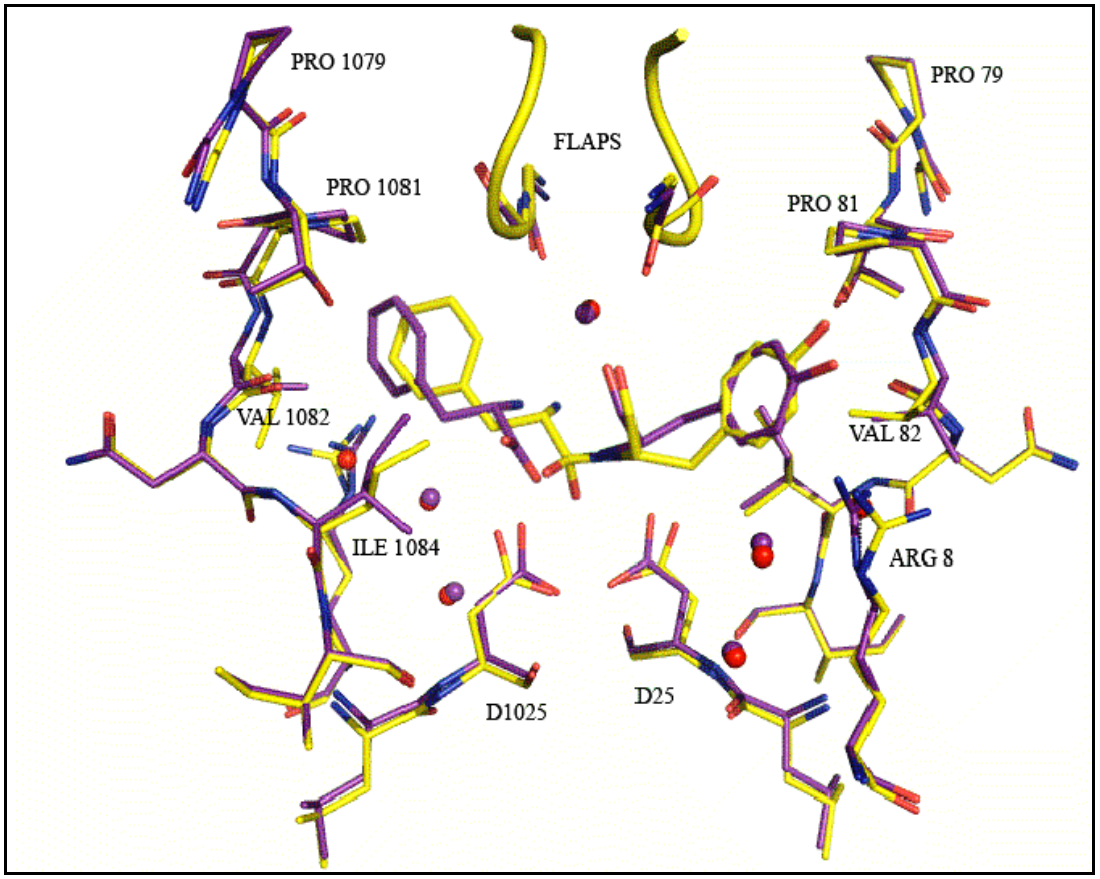


Figure 35: Superposition of the present structure [PDB ID: 3MIM] (yellow) with that of the product complex [PDB ID: 2NPH] (purple). The flaps are shown as cartoon diagram.

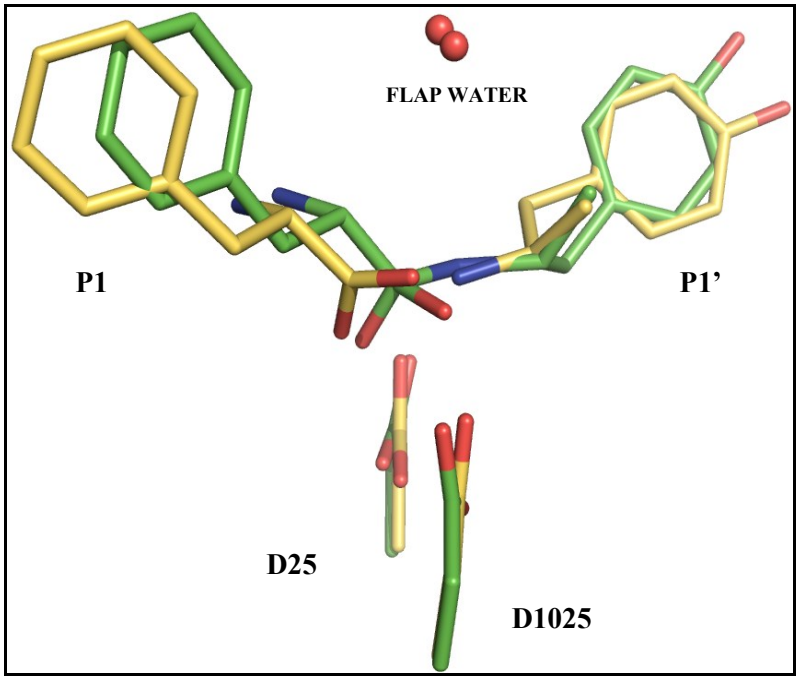
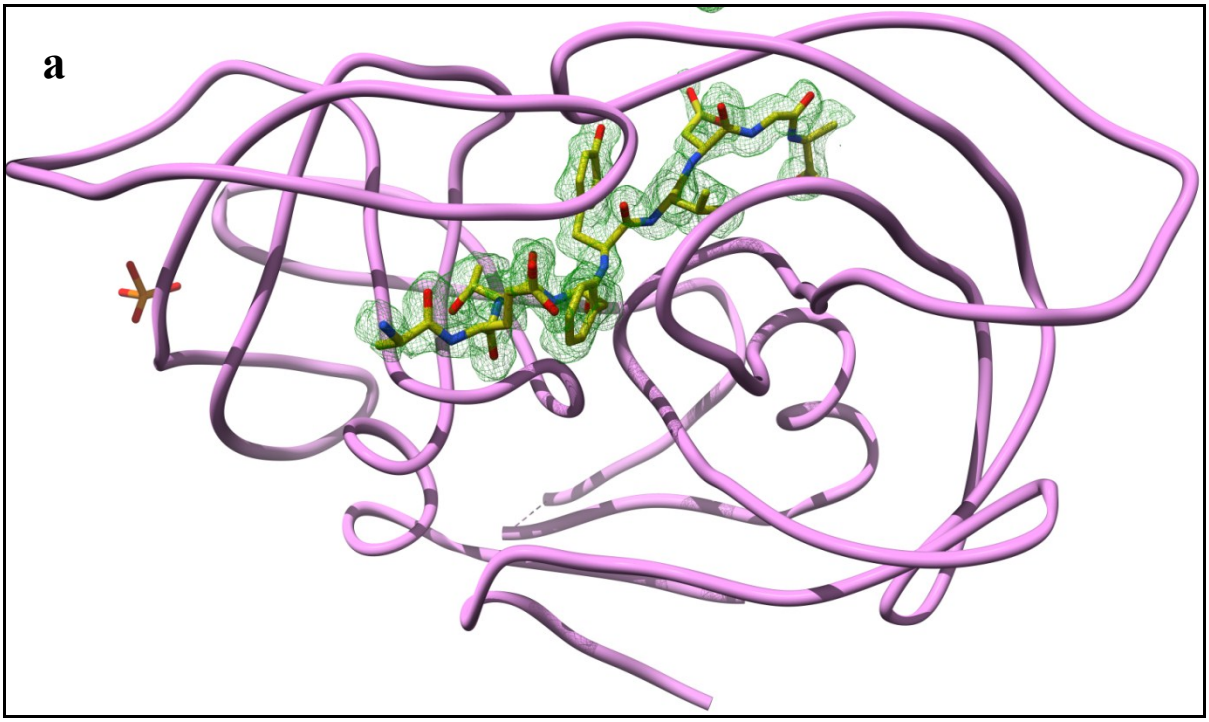


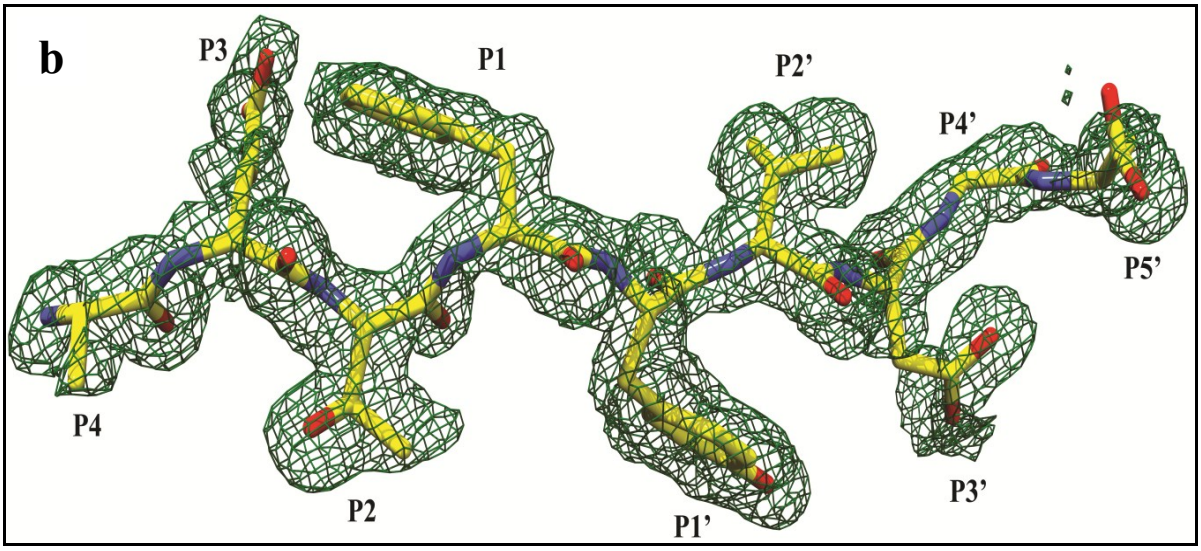
Figure 36: Superposition of the TI complex [PDB ID: 3MIM] (green) with that of the product complex [PDB ID: 2NPH] (yellow) showing the P1-P1' residues and the catalytic aspartates only. Note the change in the planarity of the carboxylate of the catalytic aspartates.

3.3.3 HIV-1 protease/substrate complex at pH 8.0

The crystals of HIV-1 protease were soaked in the substrate at pH 8.0 for durations varying from 6 hours to 3 days. Many datasets were collected and initial SA omit electron density maps were

calculated on each dataset. The quality of the maps in the enzyme active site region were limited by the loss of crystallinity at high pH and radiation damage resulting in low resolutions and completeness; low occupancy and pseudo two-fold orientation of the substrate peptide, AETF*YVDGAA made the interpretation of the maps difficult. Thus the higher resolution datasets showing strong electron density positive peak in the active site of HIV-1 protease were only examined and one such dataset (Table 9) was selected. The native HIV-1 protease crystals were soaked in substrate solution at pH 8.0 for a day. The continuity of *mFo-DFc* map, indicated the presence of the substrate in the enzyme active site. Thus the fitted models at the cleavable peptide bond could be either an uncleaved peptide or a TI. In the Figure 37a the uncleaved substrate bound to active HIV-1 protease is shown. The SA omit map shows that the density near the cleavable peptide bond is thick enough to accommodate only one carbonyl oxygen atom. The uncleaved regular peptide model fits well in one orientation without unaccounted difference density peaks (Figure 37b and 37c). Carefully fitting both the models within the electron density showed that the one of the two hydroxyls of the TI *gem*-diol does not fit into the electron density contour (Figure 38a) whereas the regular peptide carbonyl fits well (Figure 37b). The electron density could be seen beyond the P2/P2' residues and nine residues P1-P5' could be modeled. Real-space correlation coefficients calculated over the substrate residues, modeled either as a regular peptide or as a TI are listed in Table 10. The correlation coefficient averaged over either all the nine residues or over the two residues, P1 and P1', is higher for the regular peptide model. Occupancy refinement of the substrate models converged to values of 0.80 and the rest of the occupancy was filled by four water molecules, which were placed at the positive *mFo-DFc* peaks. Since no *mFo-DFc* difference density is observed in between the two catalytic aspartates the catalytic water molecule was not placed in the vacant unit cells. The catalytic water molecule does not co-exist with the uncleaved substrate. The refinement statistics is given in Table 9. It can be seen that the R- and R-free values for the refined regular peptide model are lower than those for the tetrahedral peptide model. The molecular motif therefore is a complex between the HIV-1 protease and the regular peptide in one orientation. There is one phosphate ion modeled in this structure (Figure 37a) and is included with ligand as shown in Table 9.





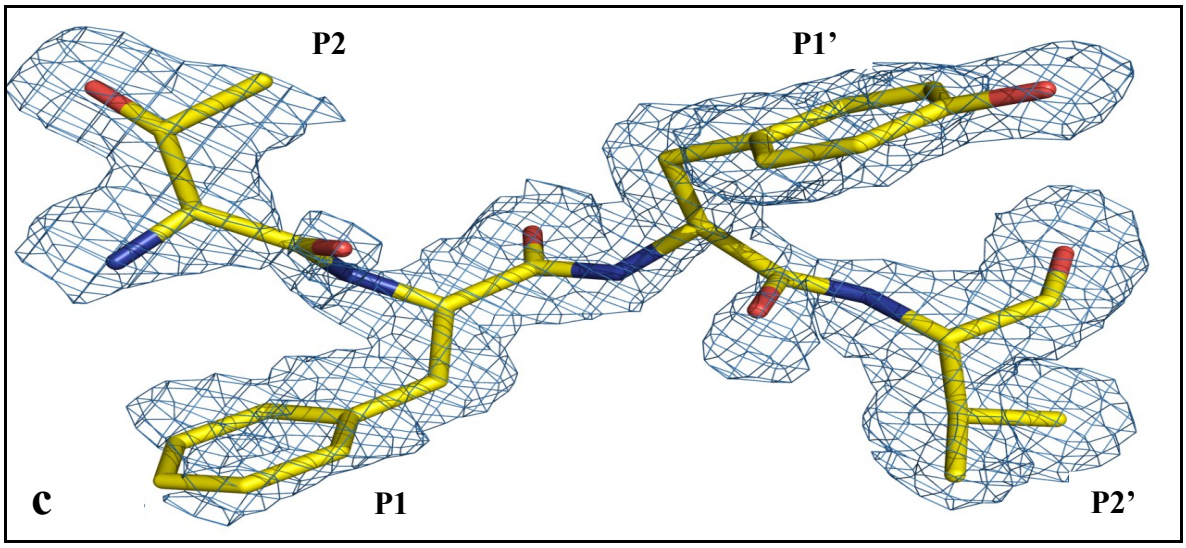


Figure 37: a) The cartoon diagram of HIV-1 protease with the SA omit map and substrate model occupying the active site cavity. A phosphate ion is also shown as orange sticks. **b)** SA omit with the refined substrate model. **c)** SA omit map in the P2-P2' region of the substrate. The maps are contoured at 1.5 σ. The refined nine residues of uncleaved substrate model are shown as yellow sticks where the substrate atoms are excluded from structure factor calculations.

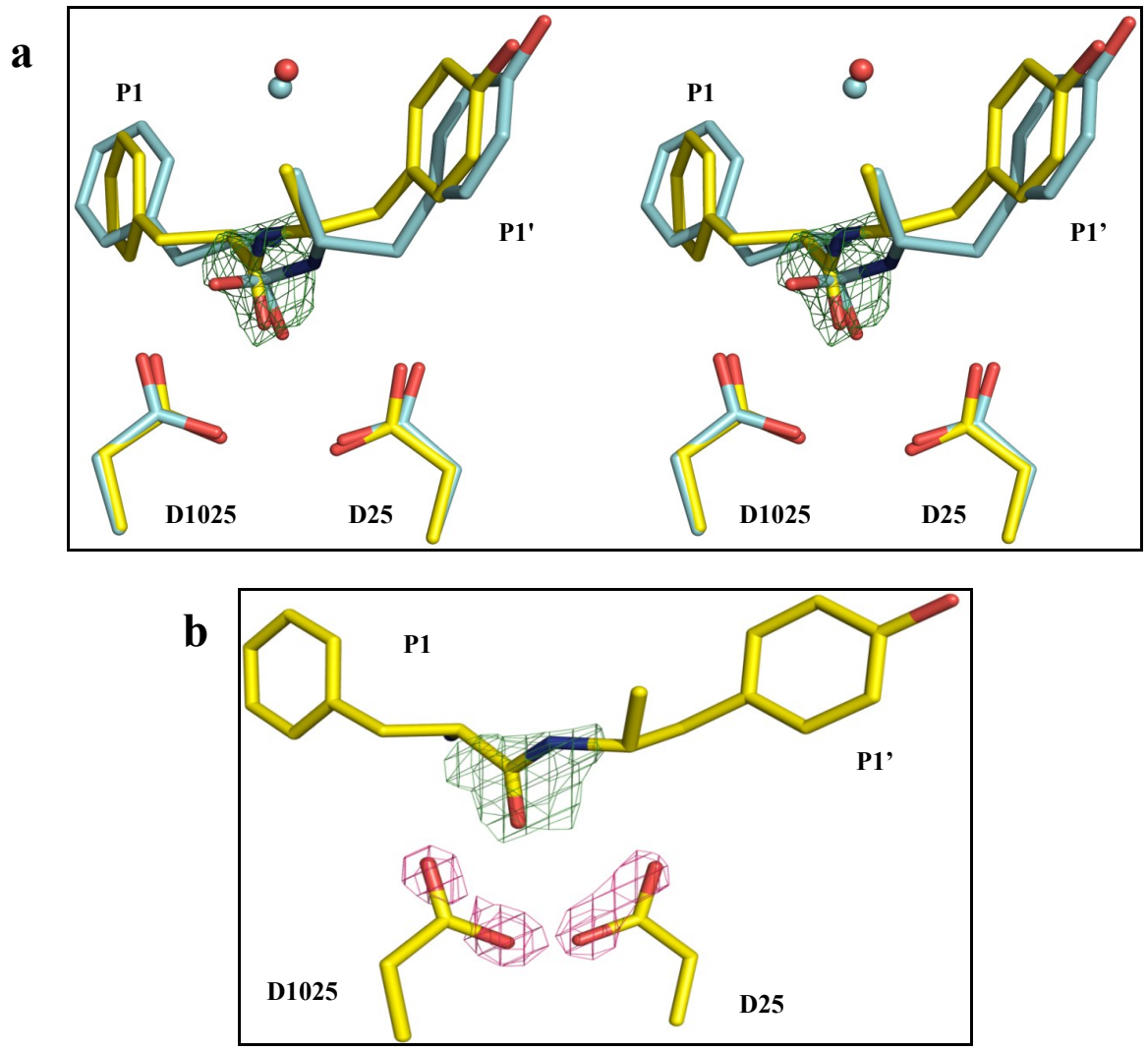


Figure 38: a) The refined tetrahedral intermediate, TI, (cyan sticks) is shown superposed on the regular peptide (yellow sticks) model. SA omit map contoured at 1.5 σ is shown, where the atoms of scissile peptide bond were excluded from map calculation. Note that the two hydroxyl groups of the tetrahedral peptide model are not well inside the electron density **b)** SA omit maps are shown where the scissile peptide bond (green, contoured at 1.5 σ) and the four oxygens of the catalytic aspartates (red, contoured at 2.7 σ) are omitted during map calculation.

Table 9: Refinement statistics comparing the TI and regular peptide models.

Ligand model refined:	**TI**	**Regular**
Resolution range (Å)	50.0 – 1.89	50.0 – 1.89
R_{work} (%)	20.17 (23.06)[#]	19.44 (22.39)[#]
R_{free} (%)[*]	23.59 (26.25)[#]	22.80 (24.84)[#]
Number of protein atoms	1514	1514
Number of solvent atoms	161	161
Number of ligand atoms	75	74
R.m.s deviation of bond lengths (Å)	0.014	0.013
R.m.s deviation of bond angles (°)	1.46	1.60
Ramachandran region (%) Favoured Allowed Outlier	97.0 3.0	97.5 2.5 0.0
Average B factor (Å^2) for Protein atoms Ligand atoms Water atoms	 24.4 59.9 41.6	 24.5 58.3 42.0

** The R_{free} was calculated using 5% of random reflections which were kept apart from the refinement throughout. # Highest resolution bin is 1.94 Å – 1.89 Å.*

Table 10: Real-space correlation coefficients for the nine substrate residues in the refined regular peptide and TI models.

Residue no.	Correlation coefficient	
	TI	Regular
1 (P4)	0.867	0.877
2 (P3)	0.772	0.845
3 (P2)	0.804	0.801
4 (P1)	0.864	0.911
5 (P1')	0.728	0.759
6 (P2')	0.730	0.771
7 (P3')	0.889	0.892
8 (P4')	0.836	0.877
9 (P5')	0.836	0.917
<4 & 5>	0.814	0.835
<All residues>	0.783	0.850

3.3.3.1 Hydrogen bonding interactions of regular peptide/HIV-1 protease in the active-site

The substrate binds in the active site in a slight asymmetric manner. The hydrogen bonding pattern in the active site cavity is shown in Figure 39. There are nine hydrogen bonds between the substrate and the enzyme. On the Pn side, the backbone of substrate forms two hydrogen bonds with the roof whereas in the Pn' side, two similar interactions are present with the base of the active site of HIV-1 protease. The flap water makes only one hydrogen bond with the substrate P1' carbonyl oxygen whereas the other hydrogen bond to the P2 carbonyl oxygen, which is found in earlier two substrate complexes, is absent (3.3 Å, shown as red dots in Figure 39). The hydrogen bonding interactions of the substrate with the catalytic aspartates, D25 and D1025 are shown in Figure 40. The scissile peptide carbonyl oxygen, O1 makes a hydrogen bond (2.5 Å) with the outer oxygen of the catalytic aspartate, D1025. This aspartate has lower B-value than the other aspartate ($<B_{D25}>$ = 17.7 $Å^2$ vs $<B_{D1025}>$ = 19.6 $Å^2$). The OD1 atoms of D25 and D1025 are separated by 2.9 Å. The inner oxygens (OD1) of the catalytic aspartates make hydrogen bonds with the nitrogens of G27/G1027 (2.7 Å and 2.8 Å). The pseudo-torsional angle between the carbonyl atoms of D25 and D1025 was measured and it was found that the planes of the two aspartates are rotated along Cβ-Cγ bond by 31°. The scissile peptide nitrogen of the substrate is located 3.4 Å and 4.1 Å away from the outer oxygens of D1025

and D25, and therefore does not make any hydrogen bond with the enzyme as observed in TI/HIV-1 protease complex at pH 2.0.

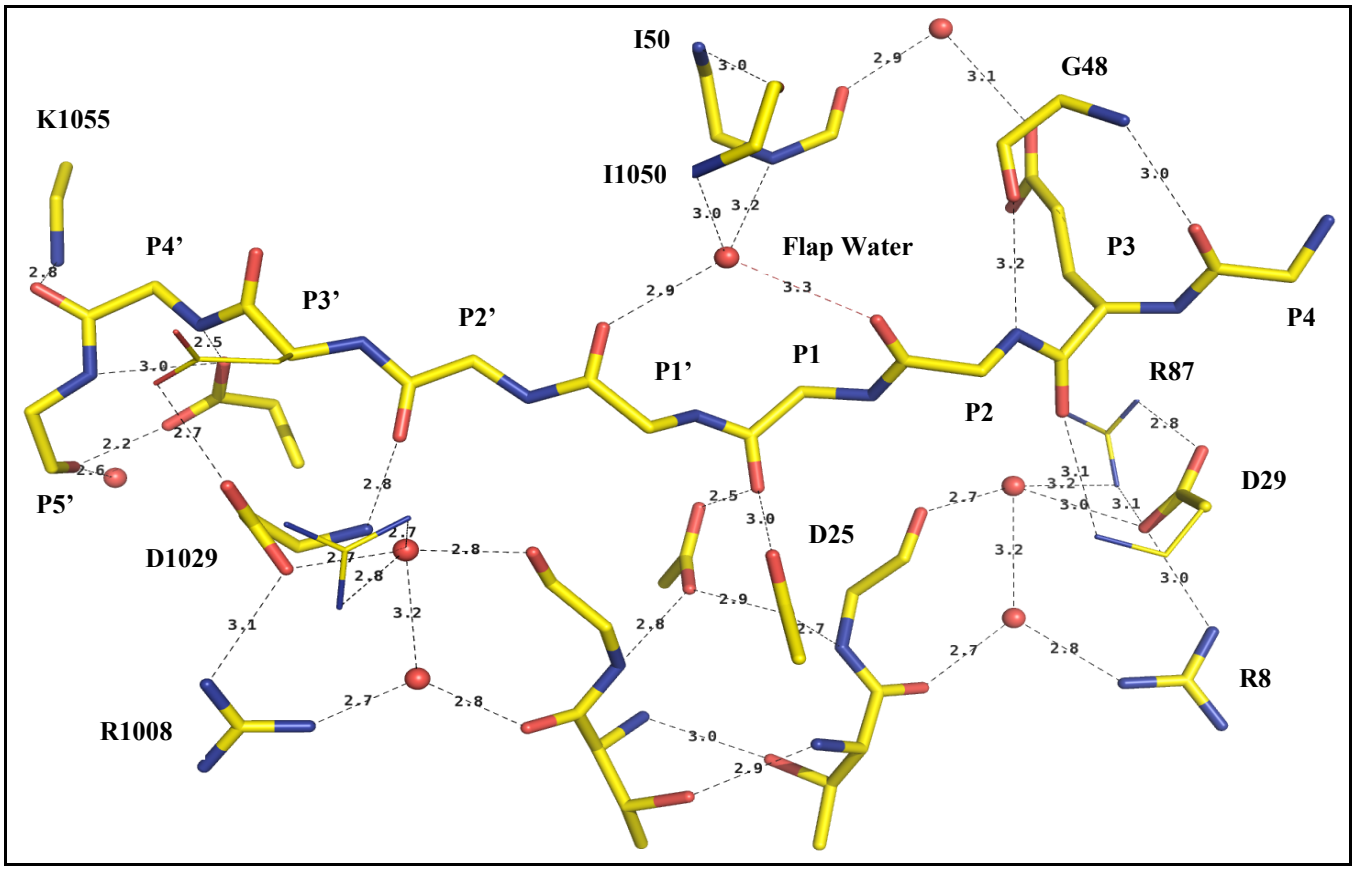


Figure 39: Hydrogen bonding interactions of the uncleaved peptide in the active-site of HIV-1 protease.

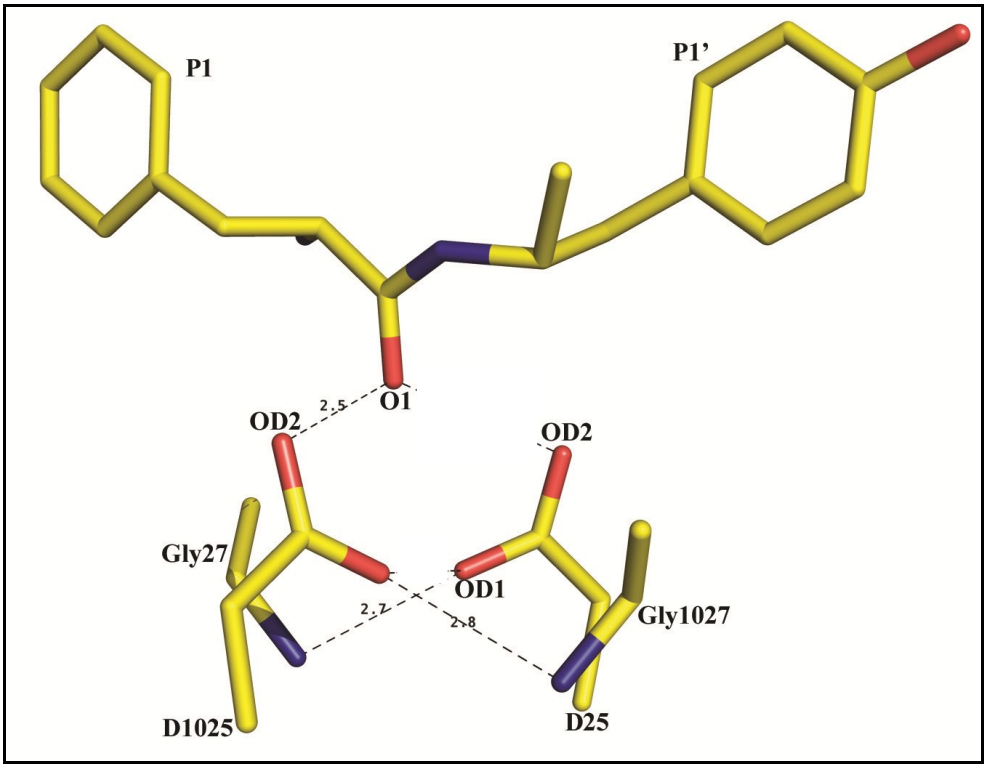


Figure 40: Hydrogen bonding at the catalytic center of regular peptide/HIV-1 protease complex.

3.3.3.2 The protonation state of the aspartates in the regular peptide/HIV-1 protease complex

In the regular peptide/HIV-1 protease complex at pH 8.0 the catalytic water molecule is absent. The scissile peptide carbonyl oxygen, O1 makes a strong hydrogen bond (2.5 Å) with OD2 of D1025. The interaspartate distance between inner oxygens (OD1 --- OD1) is 2.9 Å. Assuming this to be a hydrogen bond would imply the catalytic dyad to be di-protonated, which is unlikely because of the higher pKa_2 of catalytic dyad at pH 8.0. We suggest that the repulsion between the unprotonated OD1 atoms is overcome by their strong hydrogen bonds (2.7 Å and 2.8 Å) with the nitrogens of G27/G1027, a suggestion made earlier [146]. The catalytic aspartic acid dyad is thus monoprotonated and the lone proton should be located on the outer oxygen of either D25 or D1025. The probability of the proton residing on the outer oxygen of D1025 is higher by virtue of very strong hydrogen bond (2.5 Å). This inference is consistent with the fact that pseudo-angle C(D1025) - OD2(D1025) --- O1 (108.1°) being closer to the tetrahedral value as compared to C(D25) --- OD2(D25) - O1 (104.6°). Thus the proton is located on the outer oxygen of D1025 making inner oxygen of D25 anionic. It is worth mentioning that based on the higher angular deviation (22°) of the C(D1025) - OD1(D1025) - OD1(D25) angle and lower deviations (9.9° and 11°) of the Cα(G27/G1027) - N(G27/G1027) - OD1(D25/D1025) from tetrahedral value, the inner oxygens cannot have a proton. A di-anionic aspartate dyad, having two units of negative charge at this higher pH, will be unstable in the non-polar environment. The pKa_2 of the aspartic acid dyad has been shifted to higher values by the presence of the substrate, as effective pKa values can be perturbed by the microenvironment, electrostatic, hydrogen bonding interactions and solvent accessibility [455]. The localization of the proton on the D1025 OD2 in an uncleaved substrate bound to HIV-1 protease is a very important result from mechanistic point of view.

3.3.3.3 The strained regular peptide

The regular peptide in the present complex binds in single orientation. The scissile peptide between P1 and P1' residues is positioned between the two catalytic aspartates (Figure 40) with *trans*-peptide bond geometry. The peptide omega angle is 140° as against the normal 180° required for *trans* peptide geometry. This shows that the bound uncleaved peptide is under torsional strain.

At higher pH, the enzymatic reaction is either halted or reduced drastically as observed from the activity vs pH curve [143]. Since the rate of reaction is reduced above pH 7.0, the substrate has been

successfully trapped uncleaved. The trapping of a regular peptide substrate reported here is the first structure of unprocessed enzyme/substrate complex. Since different enzymes exhibit wide ranges of cleavage rates there is always a finite probability that substrates may bind in the active site without yielding any products. In bi-molecular reactions, the formation of the near attack conformation in the Michaelis complex for product formation requires a narrow conformational space depending upon proximity, stereo-chemical orientation and steering of orbital of the reactants [102]. Unproductive mode of substrate binding is a statistical phenomenon, which becomes dominant, when the pH is increased to 8.0. Further analysis was done by comparing secondary hydroxyl containing inhibitor/HIV-1 protease complexes with the present uncleaved substrate complex (Figure 41). The

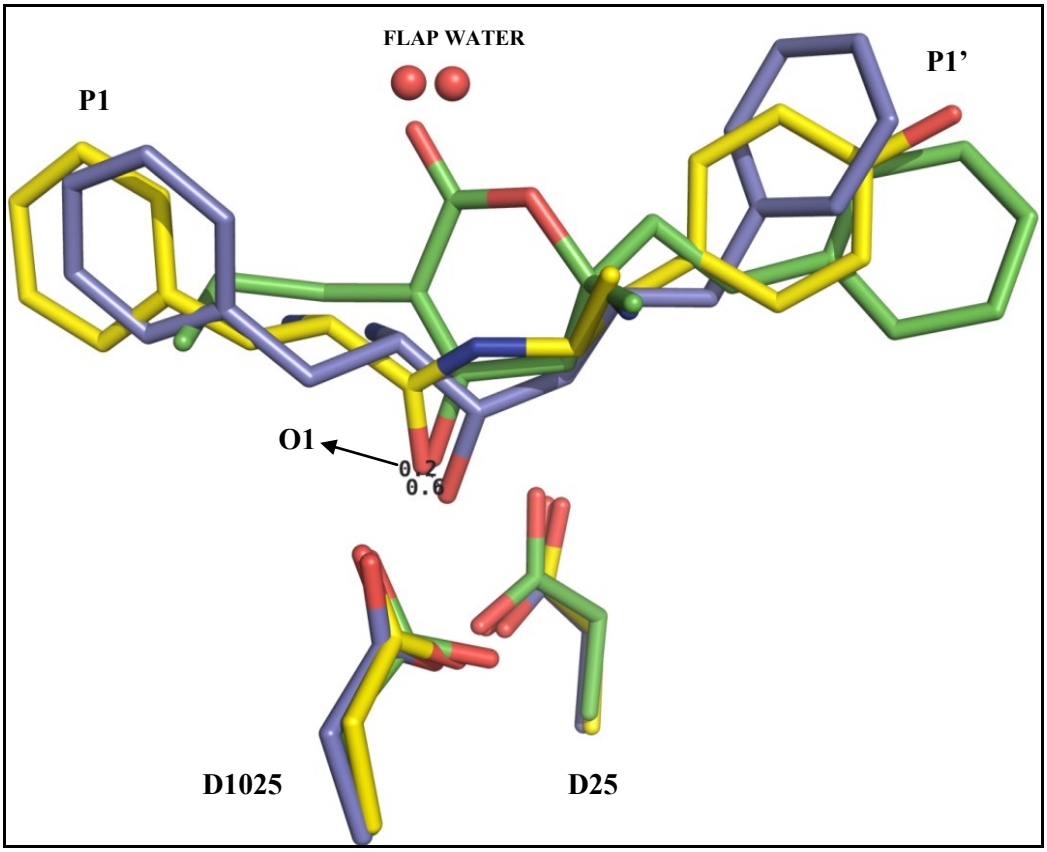


Figure 41: The regular peptide model (yellow sticks) superposed onto the FDA approved hydroxy inhibitors, ritonavir (blue sticks) and DMP323 (green sticks). Note the peptide carbonyl oxygen, O1 is near to the hydroxyl oxygens of the two different inhibitors.

catalytic water molecule is also absent in all the these complexes, as its position is occupied by the central secondary hydroxyl group of the inhibitors. The scissile carbonyl oxygen, O1 superposes very near to the secondary hydroxyl oxygen of existing FDA approved drugs, ritonavir and DMP323 (rmsd based on protein Cα atoms are 0.344 Å and 0.481 Å respectively) (Figure 41). The distances between O1 of the regular peptide and the inhibitor hydroxyl oxygen are 0.6 Å and 0.2 Å respectively.

3.3.3.4 Comparison of the regular peptide/HIV-1 protease and TI/HIV-1 protease complexes

Figure 42 shows the superposed models of *trans* regular peptide/HIV-1 protease and TI/HIV-1 protease complexes. The deviation from values of the peptide omega angles in TI and regular peptide are 74° and 40° respectively. The scissile carbonyl oxygen in the present structure is closer to the O2 oxygen of the TI *gem*-diol, which is derived from the catalytic water molecule, and makes a strong hydrogen bond (2.5 Å) to the OD2 of D1025. Instead this scissile carbonyl oxygen is farther away from the other oxygen of the TI *gem*-diol, which is derived from the peptide carbonyl of the substrate. Thus the regular peptide binds in a conformation where the scissile carbonyl oxygen is close to the catalytic water position and displaces it. Both these observations along with the fact that scissile carbonyl oxygen is also very close to the inhibitor hydroxyls (Figure 41) confirm its non-productive binding of the substrate in the present structure.

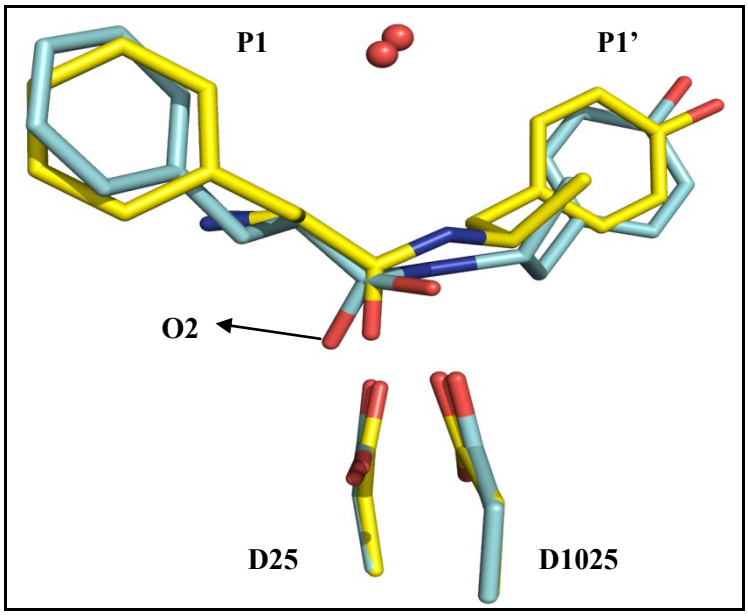


Figure 42: The regular peptide model (yellow sticks) superposed onto the TI/HIV-1 protease complex [PDB ID: 3MIM] (cyan sticks). The flap water molecules are shown as red spheres.

There are several differences in the side-chain conformations of the acidic and basic residues of both the substrate and enzyme in the two structures at pH 8.0 and pH 2.5. The active site has a combination of basic and acidic residues and depending upon the pH protonation states changes. The residues D29, D30, R8, P3'Asp and P3Glu adopt different conformations. An example of such difference in the P3'Asp side-chain positions can be seen in Figures 33 and 39. At pH 2.5 it makes a direct hydrogen bond with G48 carbonyl wheras at pH 8.0, it hydrogen bonds to a water molecule.

3.3.3.5 Comparison of the regular peptide complexed to active HIV-1 protease and inactive D25N HIV-1 protease

Figure 43 shows the comparison of the present structure with regular peptide/inactive D25N HIV-1 protease complex. In both, the peptide carbonyl oxygen hydrogen bonds (2.5 Å and 2.7 Å respectively) with the outer oxygen of the same catalytic aspartate. However, in the inactive D25N HIV-1 protease/regular peptide complex, the orientation of the peptide bond is different with respect

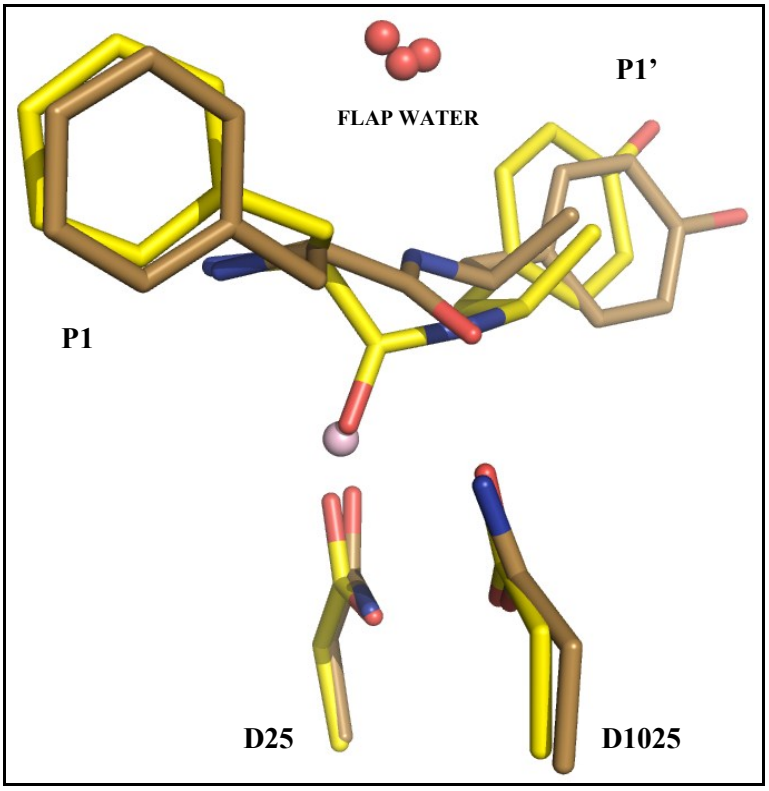


Figure 43: The regular peptide model (yellow) superposed onto the same peptide/D25N inactive HIV-1 protease complex (PDB ID: 1KJG) (brown) and apo-HIV-1 protease. The catalytic water molecule present in apo-enzyme is shown as pink sphere, which is near the scissile carbonyl oxygen of the regular peptide.

to the two catalytic aspartates (Figure 43). The plane of the scissile peptide bond in active enzyme complex (present structure) is almost parallel to the carboxylate planes of the two catalytic asparates, and binds similarly to inhibitors having a secondary alcoholic moiety (Figure 43). In contrast this peptide plane is rotated and is almost perpendicular to the carboxylate planes in the inactive D25N enzyme. The catalytic water molecule is absent in both the complexes. Thus the substrate binds HIV-1 protease in a non-productive conformation at pH 8.0 and the catalytic water molecule is displaced by carbonyl group. It is seen that the scissile carbonyl oxygen is very close to the catalytic water and not appropriately positioned for attacking the carbonyl carbon atom. As a result the substrate is trapped in the active site in an uncleaved state.

3.3.4 Structural snapshots of the cleavage reaction by HIV-1 protease

An enzyme crystal is a highly ordered three-dimensional arrangement of protein molecules with a number of well defined and stable contacts. The lattice contacts are similar to, but less extensive than, contacts between subunits in oligomeric proteins and may have a similar effect on protein function in the crystal as subunit-subunit contacts have on oligomeric proteins in solution. The lattice allows for some flexibility and protein molecules may undergo limited conformational changes. Protein crystals contain ~50 % water representing a somewhat denser environment than the cytosol. Many enzymes retain their activity in the crystal while few enzymes (e.g. papain) are even believed to be more active in the crystal than in solution [461]. In native HIV-1 protease crystals the three-dimensional network of solvent channels provide access to the enzyme active site for the diffusing substrate molecules. In the present study, the decapeptide substrate, AETFVDGAA diffuses into the HIV-1 protease active site cavity, increasing its occupancy in a time-dependant manner. Since the enzyme active site has both hydrophobic and hydrophilic residues, the substrate binding depends on the pH of the soaking solution and the nature of the substrate side-chain residues. Thus rate of reaction depends upon both pH and soak time. These two variables were modulated to successfully trap the substrate at three different stages of the cleavage reaction occurring within the crystals. The P1 and P1' residues of the substrate lie between the two catalytic aspartates, D25 and D1025, in all the structures. The scissile peptide bond between P1 and P1' residues is: a regular peptide bound in non-productive mode at pH 8.0, *gem*-diol intermediate generated *in situ* just after water attack at pH 2.5, and *in situ* cleaved bi-product at pH 6.2. These structures provide the structural snapshots of three different stages of cleavage reaction.

The hydrogen bonding patterns derived from these high resolution structures provide accurate insight into the enzyme mechanism. In some computer simulation studies on oligopeptide hydrolysis by HIV-1 protease, have led to the conclusion that electrostatic stabilization, such as hydrogen bonding, is more important for enhancing enzyme reaction rates than dynamic effects [162]. In the three structures presented here, the substrate is bound in an extended β-sheet conformation. The substrate establishes extensive contacts with both protease monomers; however, the contacts between monomers are not equally distributed. The asymmetric nature of decapeptide breaks the symmetry of

protease dimer in complexes by forming more contacts with one monomer than the other. Protein Cα atoms superpositions of the three structures indicate that the TI/HIV-1 protease and regular/HIV-1 protease structures superpose (rmsd is 0.22 Å) very well with each other (Figure 46). The cleaved peptide/ HIV-1 protease structure has a higher rmsd of protein Cα atoms with regular/HIV-1 protease (rmsd is 0.39 Å) and TI/HIV-1 protease (rmsd is 0.32 Å) structures.

The substrate/protein hydrogen bonds are nine each in the regular peptide and TI complexes, but increases to thirteen in the bi-product complex. This is due to linear displacement between the post-cleavage P- and Q-fragments along the active site tunnel making more contacts away from the catalytic site. In contrast to backbone atoms, the side chain atoms are much more flexible in these complexes and the deviation is larger at both termini. The substrate is more flexible than the protein as inferred from the B-factor values. The residues beyond P2/P2' are more flexible, especially the side chains, due to expansion in the active site cavity of the enzyme representing the entry and exit sides. On an average, the three states of the substrate superpose very well except in the P1 to P1' region (Figure 46). Thus the overall substrate conformation is conserved within the active site cavity through a wide range of pH values between the three states.

3.3.5 The proposed catalytic mechanism based on the HIV-1 protease/substrate complexes

The strategy of low pH and varying soaking times used in the present thesis has allowed accumulating and trapping *in situ* reaction intermediates within the crystals by conventional protein crystallographic methods.

The structures presented here are suggestive of the following mechanism of cleavage by HIV-1 protease. The catalytic mechanism of HIV-1 protease consists of two parts: 1) formation of a TI and 2) collapse of the TI to products on protonation of the scissile nitrogen atom [143, 145, 147-152]. However, there is no proof for the sequence of events that actually happen at the atomic level. In some proposals, the two processes happen simultaneously resulting in a single-step cleavage, leaving no scope for observation of a TI [145]. The catalytic mechanism of HIV-1 protease involves a TI flanked on either side by two transition states, TS1 and TS2. The activation barrier for TS2 is calculated to be higher than that for TS1 [462]. Formation of TI is preceded by two reaction steps: 1) polarization of scissile carbon atom, and 2) activation of nucleophilic water. There is no consensus

on the enzyme-substrate interactions that represent these steps. While some invoke, for polarization of scissile carbon, a hydrogen bond from scissile nitrogen atom to aspartate [145, 161], others invoke hydrogen bond from scissile carbonyl oxygen with aspartate [143, 144]. In X-ray structures, hydrogen bonds shorter than 2.5 Å are classified as "low barrier hydrogen bond" or LBHB [456] and when charged it is called "short ionic hydrogen bond" or SIHB [453-455]. The role of hydrogen bonds in enzymatic reactions is a well known fact, but the proposal that LBHB's drive and contribute to enzymatic catalysis is currently gaining importance, and has been demonstrated experimentally in a number of enzymatic systems [168, 169, 456]. Such delocalization causes the strengths of LBHBs and SIHBs to increase up to 20 kcal/mol as compared to 4-5 kcal/mol in a normal hydrogen bond [168]. In the case of HIV-1 protease, Northrop has invoked formation of a LBHB between inner oxygens of catalytic aspartates to account for all experimental observations on catalysis by aspartyl proteases. Theoretical calculations of varying degrees of sophistication also point to existence of LBHB's in the active-site of HIV-1 protease [157, 161-165]. The SIHB between O1 and (D1025 OD2) and the absence of any hydrogen on (D25 OD2), suggest that hydrogen bonding to the scissile carbonyl oxygen, is one of the key interactions in the Michaelis complex prior to TI formation (Figure 44a). This hydrogen bond from D1025 polarizes the carbonyl group, while in the same Michaelis complex, a hydrogen bond to the inner oxygen of anionic D25 activates the nucleophilic water molecule. Nucleophilic attack on the scissile carbonyl carbon atom leads to formation of TI, concomitantly with conversion of the hydrogen bond between catalytic aspartate and the scissile carbonyl into an SIHB. Such conversions of normal hydrogen bonds into SIHB's or LBHB's are important components of enzyme catalysis [168, 453, 454]. There are more than 250 structures of HIV-1 protease/inhibitor complexes available in the literature [19]. Examination of these structures reveals that, the length of the hydrogen bond between the inner carboxylate oxygen atoms of the two catalytic aspartates ranges from 2.59 Å to 2.86 Å. In the three complexes these lengths vary from 2.3 Å to 3.1 Å and such modulations are necessary for energy reorganization amongst the transition states and intermediates in enzyme catalysis. The TI can contain either a neutral, zwitterionic or anionic *gem*-diol, and occurrence of all these types has been postulated in different mechanistic proposals [151, 152]. The TI has not been directly observed and characterized in experiments carried out in solution, using fast substrates. Based on theoretical calculations of energetics of peptide bond

hydrolysis, it has even been suggested that the lifetime of TI's are too short to be detected experimentally [463]. However, some success in crystallographic observation of reaction intermediates has been realized by using millisecond Laue crystallography [464, 465]. The success of

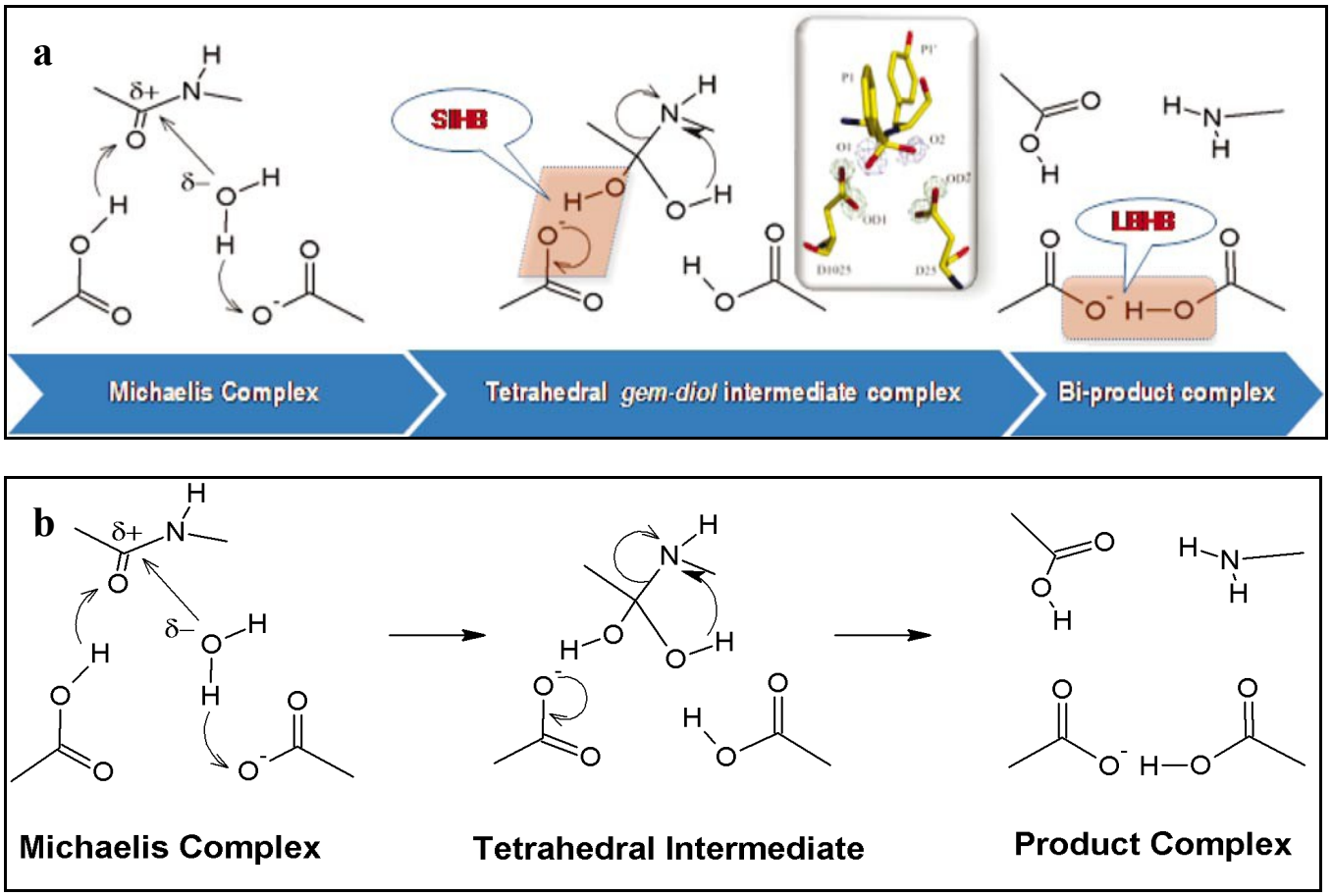


Figure 44: Schematic diagram of the proposed reaction mechanism based on the TI [PDB ID: 3MIM] and product [PDB ID: 2NPH] complexes of HIV-1 protease involving the same oligopeptide substrate. **a)** Shows the SIHB and LBHB in TI and product complexes respectively. **b)** Shows the proton shifts in Michaelis, TI and product complexes.

strategy adopted in the present thesis, in trapping the TI also implies directly that the rate-limiting step in peptide bond hydrolysis by HIV-1 protease is the collapse of TI. The collapse of TI is triggered by protonation of scissile nitrogen atom. The resulting TI is a neutral *gem*-diol, which hydrogen bonds with catalytic aspartates in a unique manner (Figure 34) which is not observed before. Given the protonation states of aspartates derived here, protonation of the scissile N-atom has to be carried out by the *gem*-diol hydroxyl (Figures 44a and 44b) through an intra-molecular proton transfer pathway. It is conceivable that this proton transfer may be facilitated, as the next step in the reaction pathway, through conformational changes in the enzyme and TI, which would steer the N-atom lone pair to receive the proton [102]. This TS2 activation barrier may be higher at lower pH. The TI collapses into product complex at pH 6.2, with the *gem*-diol converting into a planar carboxyl

group. The intermolecular SIHB from *gem*-diol is disrupted and in the product complex a new intra-enzyme LBHB is observed between the inner oxygen atoms of D25 and D1025 (Figures 44a and 44b) [460]. The TI structure obtained at pH 2.5 is similar to the F'T state but the tetrahedron is a neutral *gem*-diol with both the diol oxygens hydrogen bonding to the catalytic aspartates. The bi-product complex at pH 6.2 is similar to the F'PQ state in the mechanism proposed by Northrop (Figure 9, Chapter 1), but a LBHB is present between the inner oxygens of catalytic aspartates. It may be pointed out that this mechanism is at variance with earlier proposals: 1) conversion of normal hydrogen bond to a SIHB between enzyme aand substrate during TI formation, 2) scissile nitrogen protonation by the *gem*-diol rather than the aspartate oxygen and 3) no LBHB prior to TI formation but intra-molecular LBHB is formed between inner oxygens of catalytic aspartates upon product formation.

The importance of knowing the protonation state in the context of drug-design has been underscored till the recent work by Yu *et. al.* [466], who has developed a novel method of refining X-ray structures wherein, in the energy function used during refinement, the atoms in the active-site are modeled by quantum mechanics while other atoms are represented by molecular mechanics. This treatment, besides fitting atomic structures into experimental electron density, enables very accurate energy rankings of structures having different protonation states of the aspartates. Their calculations imply that if incorrect protonation state were used, substantial error will be introduced in predicting the binding affinities of different lead compounds.

3.4 Conclusions

The structures reported here represent the first X-ray snapshots of HIV-1 protease in action on a natural substrate and provides atomic level description of the peptide bond cleavage catalyzed by the enzyme, HIV-1 protease. The natural RT-RH cleavage site oligopeptide substrate has been trapped at three distinct stages of cleavage: as an uncleaved peptide at pH 8.0, as an *in situ* gem-diol at pH 2.5 and as *in situ* cleaved bi-product at pH 6.2. At these stages of the reaction, the aspartic dyad is monoprotonated, with the proton residing on the outer oxygen atom of D1025 at pH 8.0, inner oxygen atom of D25 at pH 2.5 and shuttling between inner oxygen atoms of D25/D1025 at pH 6.2. Binding of the *gem*-diol to catalytic aspartates is asymmetric, with O1 binding to anionic D1025 through one

hydrogen bond and O2 binding to neutral D25 through two hydrogen bonds. The hydrogen bond to D1025 is a SIHB. The inner oxygens of the catalytic aspartates are too far apart to form any inter-aspartate hydrogen bond. The structures show that the catalytic aspartates cannot protonate the scissile nitrogen atom in the TI. The hydrogen bond between the scissile peptide carbonyl oxygen and the outer oxygen of the catalytic aspartate is conserved in the three stages, and is necessary for initial substrate binding. Comparison with the bi-product complex reveals the exact reorganization of hydrogen bonds at the catalytic centre to satisfy different energy requirements during collapse of TI into products. The inferred protonation scheme and interactions at the active site provide inputs to *in silico* compound screening and for design of more effective and tighter binding substrate and/or substrate-envelope based inhibitors.

CHAPTER 4

Crystal structures of HIV-1 Protease complexed to a type-1 oligopeptide substrate - entrapment of *in situ* cleaved substrate with a *cis* peptide

4.1 Introduction

HIV-1 protease cleaves the *gag* and *gag-pol* polyproteins into mature structural and functional proteins, cleaving at nine specific sites [467, 468]. Out of the nine sites, three have proline residues at the C-terminus of the peptide bond to be cleaved. These three sites are called type-1 cleavage sites or junctions. One of these is the Phe-Pro cleavage site at the $p6^{pol}$-*PR* and *PR-RT* junctions flanking the protease sequence within the polyprotein precursor. The full-length transframe region (TFR) protease precursor, undergoes maturation via auto-cleavage to form a putative dimer [469-471], which is responsible for the appearance of mature-like catalytic activity of HIV-1 protease [471-473]. Mutations that prevent cleavage at the $p6^{pol}$-*PR* junction lead to the production of an N-terminally extended protease precursor species, and cause a severe defect in *gag* polyprotein processing and the complete loss of viral infectivity *in vivo* [474, 475]. In contrast, the C-terminal cleavage does not significantly affect either enzymatic activity or dimerization of the protease *in vitro* [476] or processing of HIV-1 precursor proteins, virus maturation, viability and morphology *in vivo* [477-479].

Structural model of the Phe-Pro cleavage site peptide was used to develop many HIV-1 protease inhibitors and, some of them are currently FDA approved drugs against AIDS [152]. The first attempts were made by Stroud and coworkers to co-crystallize HIV-1 protease with oligopeptide substrates [149]. These authors found that the substrate was cleaved during crystallization and that only one of the products was present in the crystal. Attempts to crystallize inactive D25N HIV-1 protease complexed to oligopeptide substrates containing Phe-Pro cleavage site have been unsuccessful [130]. A 2.0 Å structure of HIV-1 protease complexed to a type-1 substrate peptide, involving a different cleavage junction (Tyr-Pro) was reported from our laboratory [480]. Three-dimensional structure of Phe-Pro cleavage site substrate bound to HIV-1 protease is not yet available. Out of the two Phe-Pro cleavage sites ($p6^{pol}$–*PR* versus *PR-RT*) the $p6^{pol}$–*PR* cleavage site substrate, VSFNFPQITL was selected as it has the slowest rate of cleavage amongst the type-1 substrates (Table 2, Chapter 1). I wished to harness the slow kinetics of cleavage of the important $p6^{pol}$–*PR* junction, in order to trap the substrate at different stages of cleavage process, just the way it was obtained for the type-2 substrates discussed in Chapter 3. The structure solution of complex of

decapeptide substrate, VSFNFPQITL and HIV-1 protease will be the topic of discussion in this chapter. The interactions of the type-1 substrate will also enable a direct comparison with the type-2 substrate, and provide inputs to the mechanism of cleavage of type-1 substrates.

4.2 Methods

4.2.1 Crystallization and soaking experiments

Crystals of unliganded HIV-1 protease were grown according to procedures described in **Chapter 3**. Lyophilized powder of decapeptide, VSFNFPQITC, the chemical structure of which is shown in Figure 45, was dissolved in MilliQ water to a concentration of 5 mM and then the solution was filtered through 0.22 µm membrane. Soaking experiments were performed in the same way as described in Chapter 3. The crystals were left for incubation in the peptide solution at pH 6.2, from 6 hours to a week, at room temperature before one was taken out for data collection.

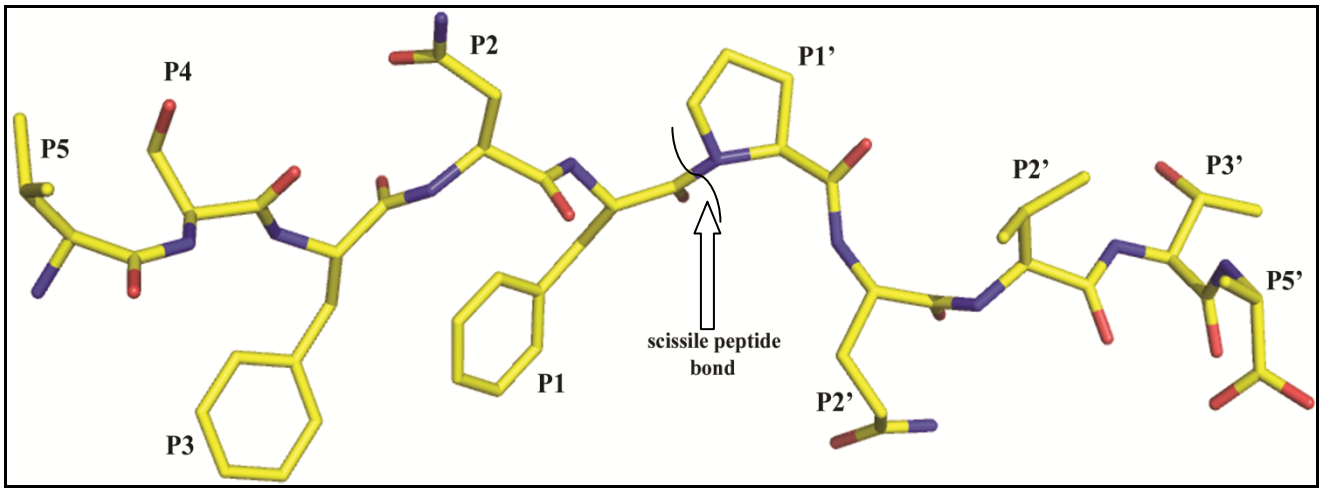


Figure 45: Molecular structure of the 10-residue type-1 peptide substrate (H_2N-VSFNF*PQITC-COOH) of HIV-1 protease that was used for soaking experiments. The scissile peptide bond between the Phe (P1) and Pro (P1') residues is indicated by the wavy line.

4.2.2 X-ray data collection

The soaked crystals were frozen under cryoconditions described in **Chapter 3** and screened for diffraction using the BM30A beamline at ESRF, Grenoble, France [439]. Two orthogonal 1° oscillation frames, each exposed for 20-30s were collected on the crystals. The crystals which diffracted beyond 2.0 Å were preserved in liquid nitrogen for subsequent data collection on BM30A and ID14-4 beamlines at ESRF. After many trials, two datasets of good quality were selected, one each collected on crystals soaked for 3 days and for 1 day in the peptide solution. Hereafter the 3 days soaked and 1 day soaked structures will be designated as "3d" and "1d" respectively. The diffraction

data on the crystal soaked for 3 days was collected remotely on BM30A beamline from Remote Data Collection Facility situated in HBNI, Mumbai. The 120 oscillation images collected on each crystal were separately indexed and integrated using the program *XDS*, and scaled using the program *XSCALE* implemented in the *XDS* program [253]. The crystals belonged to $P6_1$ space group. Crystal and intensity data statistics for the two structures presented in this chapter are given in Table 11.

Table 11: Reflection data statistics of two structures of type-1 oligopeptide substrate complexed to HIV-1 protease.

Parameters	**3d**	**1d**
Resolution (Å)	50.0-1.70	50.0-1.74
Unit cell parameters (Å)	a=b=62.49, c=82.39	a=b=62.44, c=82.24
Beamline, Exposure time	BM30A, 45s	ID14-4, 7s
X-ray wavelength (Å)	0.97945	0.9685
Total no. of reflections	235151	128747
No. of unique reflections	21960	17226
Completeness (%)	99.0 85.2 (1.74-1.70Å)*	95.3 88.5 (1.78-1.74Å)*
Mosaicity (°)	0.26	0.31
R_{merge} # (%)	8.2 58.0 (1.74-1.70Å)*	7.2 46.5 (1.78-1.74Å)*
<I/σ(I)>	21.43 3.10 (1.74-1.70Å)*	21.76 2.98 (1.78-1.74Å)*

* *The numbers between parentheses indicate the value in the highest resolution shell.*

$$\#\ Rmerge(I) = \left(\sum_i \sum_j |I_j - \langle I \rangle| \right) \Big/ \sum_i \sum_j \langle I \rangle$$

where I_j the intensity of the jth observation of reflection I, <I> is the mean of the intensities of all observations of reflection i, the summation $\sum_i$ is taken over all reflections and the summation $\sum_j$ is taken over all observations of each reflection.

4.2.3 Refinement of structures

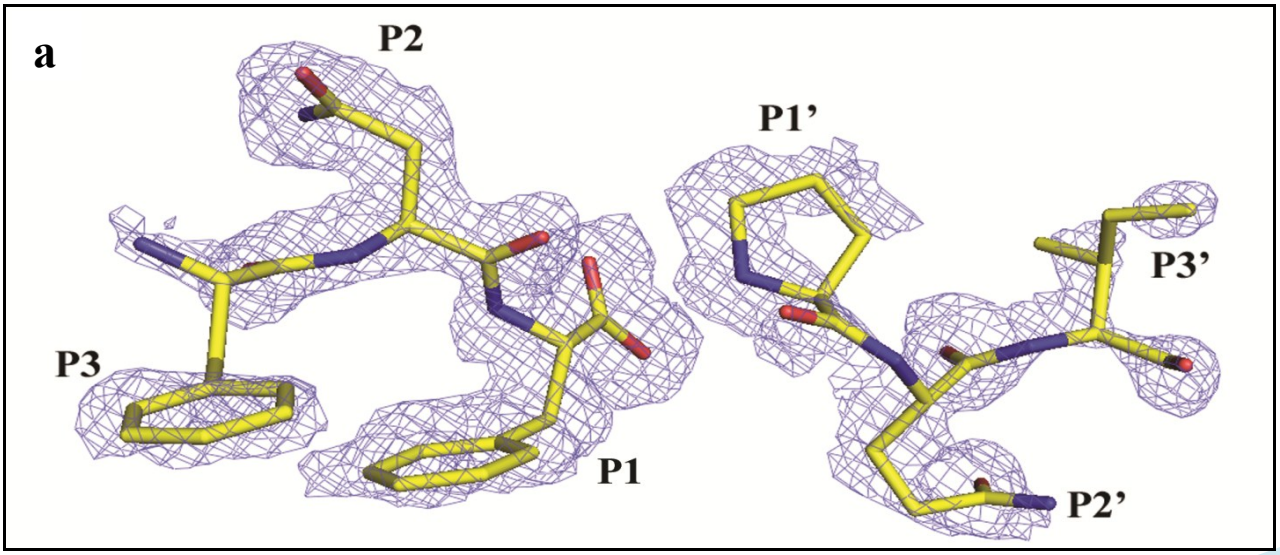
The structures were solved by the Molecular Replacement (MR) method [286]. Unliganded tethered dimer of HIV-1 protease (PDB ID: 1LV1) obtained from Protein Data Bank [412] was used as a search model after removing all the solvent molecules. The structure was refined in *CNS* using standard simulated annealing protocols and the amplitude based maximum likelihood target function [347, 370]. All reflections in the respective resolution ranges (Table 11) were included in the refinement. A composite omit map was calculated to enable unbiased interpretation of the electron

density map. Water molecules were added manually by examining environment around electron densities that were present in both *mFo-DFc* and *2mFo-DFc* maps.

4.3 Results and discussion

4.3.1 The type-1 substrate in the active site

Connected positive maxima of the *mFo-DFc* map (Figure 46a and 46b) in the active site region of the 3d and 1d indicated presence of the substrate. This map also indicated that the substrate is cleaved between P1 and P1' in both the complexes. Electron density for P2, P1, P1' and P2' residues of the peptide was clearly visible at a contour level of 2.0 σ and beyond P2/P2' it is relatively weak (Figures 46a and 46b). In both the structures, the shapes of omit densities at P1 and P1' positions were different, and were characteristic of Phe and Pro residues respectively. The molecular motif refined is a 1:1 complex between the HIV-1 protease and the bi-products, VSFNF and PQITC, formed as result of substrate cleaved *in situ* inside the crystal. The SA omit electron density map for the substrate was good for the P3'-P3 residues (Figure 46a and 46b) with a break in electron density between the Phe-Pro peptide bond, in both the complexes. Nevertheless, because of the pseudo two fold symmetry of the free enzyme active site, the substrate was modeled in two orientations and the occupancies were treated as variables during initial crystallographic refinement. It was found that the occupancy of one orientation increased from the originally assigned value of 0.5 in the two complexes, finally converging to 0.85 in 1d and 1.0 in 3d. The absence of any difference electron density corresponding to the residues of second orientation further confirmed the single orientation of substrate in the active-site. Similar single orientation has been observed in few structures of other decapeptide substrates complexed with the inactive D25N mutant of HIV-1 protease [130].



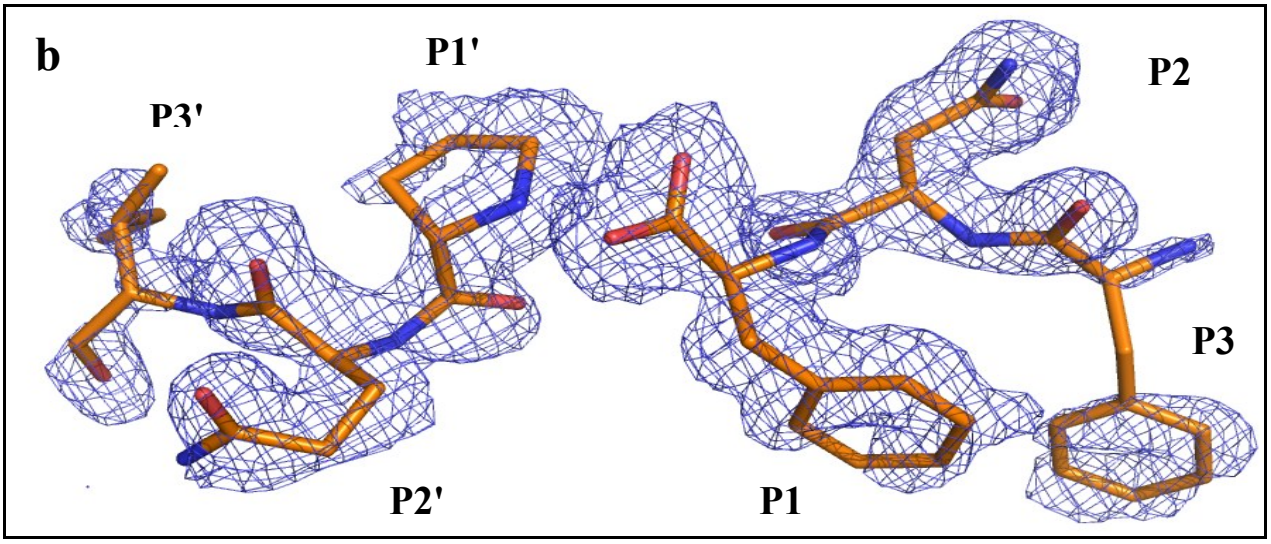


Figure 46: a) SA omit *mFo-DFc* map (light blue), contoured at 2.0 σ, level showing the refined substrate model (yellow sticks) which is cleaved between Phenylalanine and Proline residues in 1d. **b)** SA omit *mFo-DFc* map (blue), contoured at 2.0 σ, showing the refined substrate model (orange sticks) which is cleaved between Phe and Pro residues in 3d. The atoms of the substrate were not included during map calculations.

Since the crystals were of an active enzyme and the substrate was a cleavable substrate, the substrate is hydrolyzed *in situ* inside the active site cavity into two peptide fragments, cleaved between P1 and P1' residues. The refinement statistics for the two complexes are shown in Table 12. The correlation coefficients for the eight refined residues of the substrate are given in Table 13. The correlation coefficients averaged over all the eight residues, <All res> is higher for 3d probably because of higher occupancy for the substrate (Table 13). The structure therefore is of a complex between HIV-1 protease and the two product peptides.

Table 12: Refinement statistics of the type-1 substrate/HIV-1 protease complexes.

Model Refined for ligand	***in situ* bi-product model, 3d**	***in situ* bi-product model, 1d**
Resolution Range (Å)	50 – 1.70 (1.744-1.70)*	50 – 1.74 (1.785-1.74)*
R_{work} (%)	17.6 (23.2)*	17.1 (20.6)*
R_{free} (%) #	21.9 (34.8)*	20.3 (22.9)*
Number of protein atoms	1515	1521
Number of solvent atoms	220	211
Number of ligand atoms	69	69
r.m.s. deviation of bond lengths (Å)	0.018	0.018
r.m.s. deviation of bond angles (°)	1.77	1.65
Average B factor (Å^2)		
Protein atoms	26.8	32.2
Substrate atoms	49.0	47.7
Water molecules	23.3	27.5

 # The R_{free} was calculated using 5% of random reflections which were kept apart from the refinement during the whole process.
** The values for the last resolution shell are given in bracket.*

Table 13: Real-space correlation coefficients (CC) for eight residues of the bi-product type-1 substrate in 3d and 1d.

Residue no.	*in situ* bi-product model, 3d	*in situ* bi-product model, 1d
	CC	CC
1	0.674	0.701
2	0.729	0.667
3	0.872	0.803
4	0.887	0.790
5	0.705	0.741
6	0.853	0.725
7	0.605	0.543
8	0.623	0.725
<All res>	0.744	0.712

4.3.2 Hydrogen bonding in active site cavity

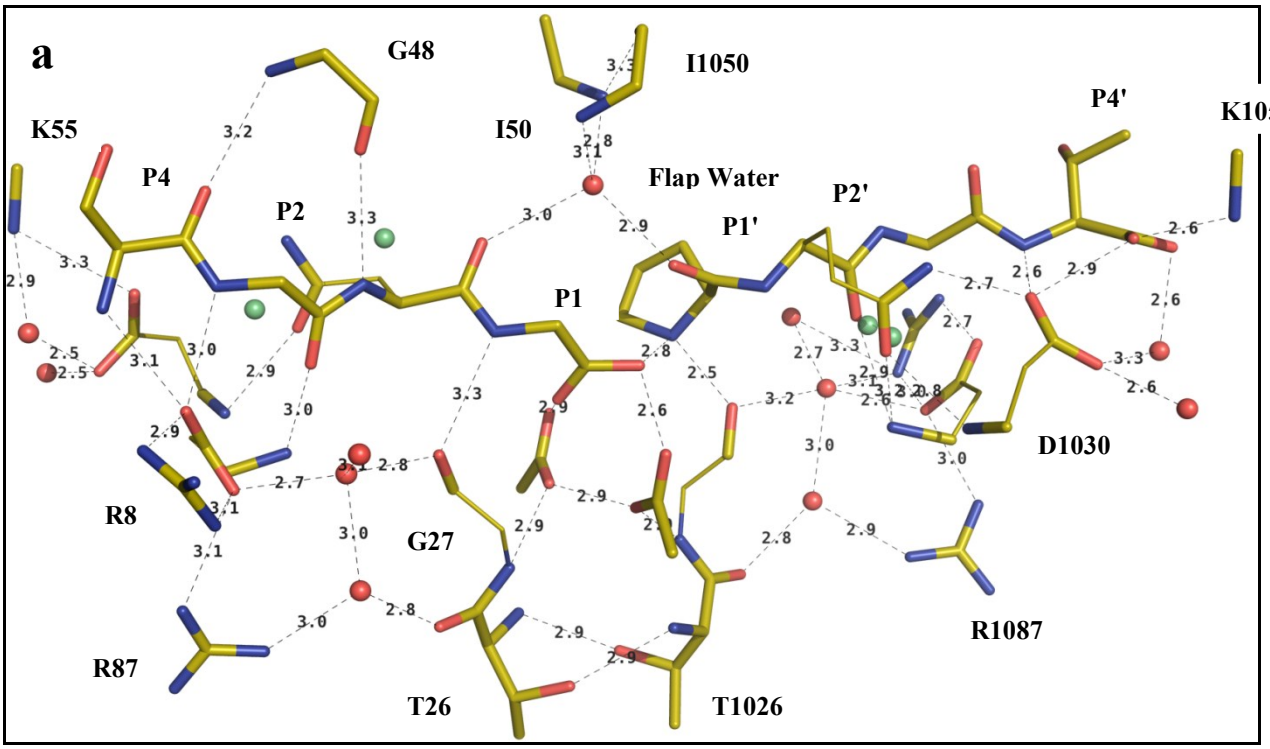
In both 1d and 3d, the type-1 substrate adopts extended β-sheet conformation, similar to that adopted by the type-2 substrate reported in Chapter 3. The flaps of the A and B subunits adopt a closed conformation, and they interact with the P- and Q-product fragments of the substrate. The substrate makes a parallel β-sheet with flap A and an anti-parallel β-sheet with flap B. Four hydrogen bonds are mediated through the flap water molecule, two are involved in anchoring the two flaps and other two are to the carbonyls of P2 and P1' of the substrate. Since the substrate binds primarily in β-strand conformations, most of the hydrogen bonds were observed mainly between the backbone of the protease and the backbone of the substrates (Figure 47a and 47b).

In 1d, the substrate binds by virtue of eighteen hydrogen bonding interactions, which are assymetrically distributed between the primed and unprimed halves of the active site (Figure 47a). Twelve hydrogen bonds are between the substrate peptide backbones and the protein. Of these, eight are on the unprimed side involving atoms D25 OD2, D1025 OD2, G27 O, D29 OD2, D29 NH, G48 O, G48 NH atoms of the first monomer, and four are on the primed side involving G1027 O, D1029 NH, D1030 OD2, K1045 NZ atoms of the second monomer. In the P-product, OD1 of P2 Asn makes a single hydrogen bond with nitrogen atom of D1030 (2.9 Å). On the other hand in the Q-product, side chain of P2' Gln makes three strong hydrogen bonds with the OD2 (2.7 Å) and N (2.8 Å) atoms of D1030 and nitrogen (2.6 Å) atom of D1029. There are five water molecules on either side of the active site cavity, and these are held through multiple hydrogen bonds with D29, R87, G27 residues

of the protein and to each other. There are two additional water molecules on each side making hydrogen bonds with D30/D1030 OD1s. It is significant that no water molecule is found near the catalytic aspartates.

In 3d, the substrate binds by virtue of twenty hydrogen bonding interactions to the protein (Figure 47b). Fourteen hydrogen bonds are between the substrate peptide backbones and the protein and these are symmetrically distributed: seven on the unprimed side involving residues D25 OD2, D1025 OD2, G27 O, D29 OD2, D29 NH, G48 O, G48 NH atoms of the first monomer, and seven on the primed side involving D1029 NH, D1030 OD2, K1045 NZ atoms of the second monomer. In the P-product, OD1 of P2 Asn makes two hydrogen bonds, one with nitrogen atom of D30 (3.1 Å) and the other with OD2 atom of D1029 (3.1 Å). In the Q-product, side chain of P2' Gln makes two strong hydrogen bonds with the OD2 (2.4 Å) and nitrogen (2.6 Å) atoms of D1030. There are three water molecules on each side of the monomer which makes multiple hydrogen bonds with D29, R87 and G27 residues of the protein and to each other. Overall the hydrogen bonding interactions are symmetrically distributed in 3d.

The backbone hydrogen bonds are more or less conserved whereas those involving the side chains of the substrate are more flexible. Importantly, in both 1d and 3d, the polar atoms of Gln at P2' in the substrate are within hydrogen bonding distance of the carboxylate oxygen of D1030 of the protease.



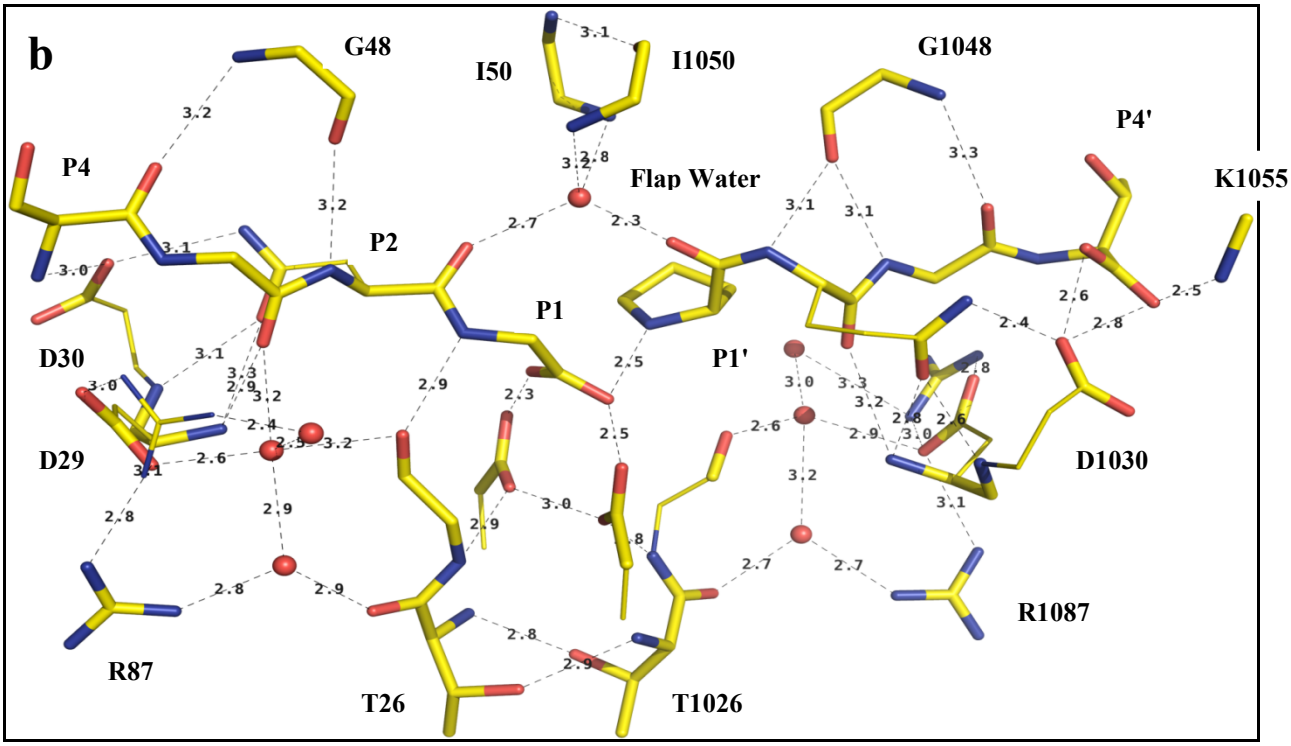
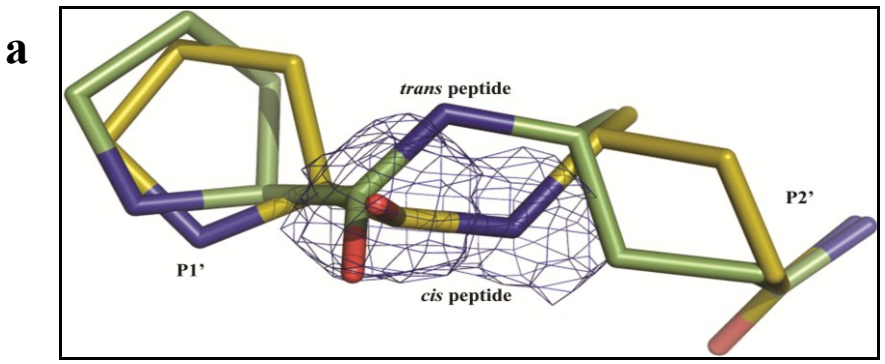


Figure 47: Hydrogen bonding in the active site of HIV-1 protease in the **a)** 1d; the partial water molecules are shown in green spheres, and **b)** 3d. Some side-chain atoms are not shown for clarity. Hydrogen bonding interactions are shown as black dotted lines. Note the changes in orientations of P1 Phe carboxylates and P1' prolines in the two complexes.

This interaction suggests that the enzyme-substrate complex is additionally stabilized. The importance of this interaction is emphasized by the absence of polymorphisms of D30 (or D1030) in the protease and variants of P2' Gln in the $p6^{pol}$-*PR* cleavage site [481].

4.3.3 Conformations of the trapped type-1 bi-product substrate

Figures 48a and 48b show the SA omit maps for the P1'-P2' peptide bond of the bi-product complexes observed in 1d and 3d. Careful observation showed that the standard *trans* conformation for the peptide between P1' and P2' residues did not fit into the SA omit map, whereas a *cis* peptide could be fitted better to the electron density. This is supported also by the lower R-free value for the *cis* peptide (22.9 %) model as compared to that for the *trans* peptide (23.1 %) model. The actual values of the peptide bond (Cα-N-C-Cα) torsion angle between P1' and P2', in 1d and 3d are ~50° and ~20° respectively.



b

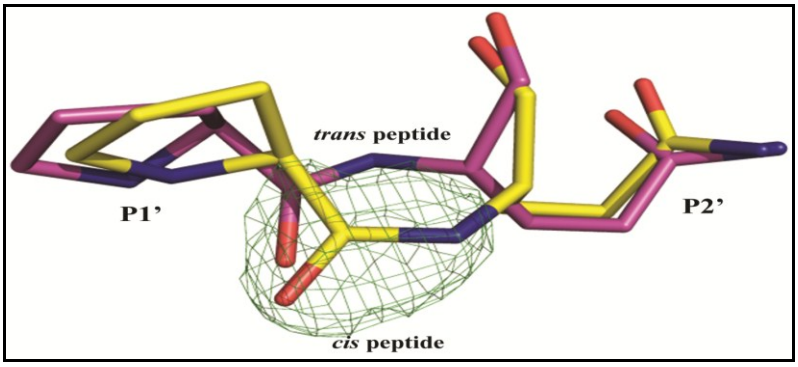


Figure 48: a) SA *mFo-DFc* map (blue) with the peptide bond between P1' and P2' of the refined substrate modeled as *cis* peptide (yellow sticks) and *trans* peptide (green sticks) in 1d. **b)** SA *mFo-DFc* map (green) with the peptide bond between P1' and P2' of the refined substrate modeled as *cis* peptide (yellow sticks) and *trans* peptide (magenta sticks) in 3d. The maps were contoured at 2.4 σ.

In protein and peptide structures, the peptide conformation is found to be *trans* in the vast majority of the cases [414]. However for peptide bonds between any amino acid and Pro (Xaa-Pro), an appreciable fraction occurs in the *cis* conformation. A survey found only 0.03 % - 0.05 % of all Xaa-nonPro [482, 483], but 5.2 % - 6.5 % of all Xaa-Pro peptide bonds to occur in the *cis* conformation [484]. Weiss *et. al.* [483] observed that at high resolution, the number of Xaa-Pro and Xaa-non Pro *cis* peptide bonds are about two and four times as high as that at medium and low resolutions. Experimentally it was found that the model compound N-methylacetamide occurs at about 1.5% in the *cis* form [485, 486] and in Pro-containing peptides a *cis* content of about 10-15 % was reported [487].

The occurrence of non-Pro *cis* peptide bonds has been associated with steric strain in proteins [488] similar to the occurrence of residues with unfavorable phi/psi-angles, and it has been speculated that the location of these *cis* peptide bonds is often special with respect to the function of the molecule [489, 490]. It was reported that many non-proline *cis* peptide bonds contain an aromatic residue and the presumed reason for this was the occurrence of an aliphatic–aromatic interaction from C–H to the π-system [491]. Structural analyses of several proteins have identified that conversion of non-proline peptide bonds from *trans* to *cis* conformations can be induced by external structural factors, such as metal binding, substrate binding or cofactor interaction that result in isomer-specific functional states [492–495]. Non-proline *cis* peptide between A207-D208 is seen in a 0.94 Å concanavalin A structure [496]. Indeed similar Pro-Gln *cis* peptide was found by Leesong *et. al.* between P31-N32 residues of *β*-hydroxydecanoyl ACP dehydrase structure [497]. In hypoxanthine-guanine phosphoribosyltransferase (HGPRT) the L78–K79 peptide bond changes its geometric status during

the course of catalysis, in which *cis*/*trans* isomerization of this peptide bond is related to the substrate binding (*trans* to *cis*) and product release (*cis* to *trans*), respectively [493, 494]. The authors proposed that a portion of the energy released upon substrate binding to the apoenzyme is used to drive the L78–K79 peptide bond into the *cis* configuration, and that the energy released upon isomerization of the *cis* peptide bond back to the *trans* ground state, helps to propel the substrate out of the active site [494, 495]. These studies indicate *cis*/*trans* conversion of non-proline peptide bond is governed by intermolecular binding events involving substrates, metal ions or cofactors. In a coarse grained molecular dynamics study, the product release pathway in HIV-1 protease was investigated. It was found that the polyprotein launching into the active site, requires an open flap conformation but the separation and release of the two product fragments can take place in closed flap conformation [498]. In the present complexes, similar *cis* conformation exists involving P1' (Pro)-P2' (Gln) peptide product bound to HIV-1 protease. These sites of strain at *cis* peptide act as energy reservoir for the system. Sites in the *gag* and *gag-pol* polyproteins are each cleaved with specific unique enzyme kinetics and in a tightly regulated manner [499] and the present cleavage is necessary to produce an active protease, occurring late in the viral life cycle during maturation of HIV-1. Since the K_m and rate of cleavage of the $p6^{pol}$-*PR* substrate used, is not the slowest amongst the cleavage sites, the post-cleavage product release steps could be rate determining. The *cis* peptides observed in the current complexes indicate, that the energy required to release the products out of the active site, is provided by the strained state of the peptide.

4.3.4 Comparison between 1d and 3d

The longer and shorter soaked complexes superpose with rmsd 0.32 Å using C*α*s of the protein atoms. Figure 51 shows the superposed substrate atoms in the two complexes. The Phe carboxylates formed after cleavage were positioned differently in the two complexes. In 3d, the carboxylate was closer to the two catalytic aspartates as compared to the 1d (Figure 49). The planes of the carboxylates have rotated in 3d to form symmetrical hydrogen bonds with both outer oxygens of catalytic aspartates (Figure 49). The Pro residues in the two complexes are differently oriented with respect to the nearby Phe carboxylate and catalytic aspartates (Figure 50). The P-product Cα's superpose very well whereas in the Q-product side there is a lot of variation. The P1 Phe residue moves slightly closer to the plane of the two catalytic aspartates whereas the P1' Pro residue moves

away. Due to the lateral shift of the Q-product in the longer soaked complex, the hydrogen bonding pattern changes (Figures 47a and 47b). The P3' Gln adopts different rotameric form in the two complexes whereas the P3 Asn superposes well. The Gln side chain, in 3d, makes stronger hydrogen bond with D1030 as a consequence of this shift (Figures 47a and 47b). The rotameric conformation of D1030 side chain is same in both the complexes. The hydrogen bonding pattern at the catalytic center is shown in Figure 50. The inter aspartate hydrogen bonds are weak. The P1 carboxylate is closer to the catalytic aspartates in the 3d. The hydrogen bond between the scissile nitrogen atom and the outer oxygen of the catalytic aspartates is absent in both structures. Assuming substrate cleavage happens within the first day, the differences show the rearrangements that take place to facilitate product release. Interestingly the Q-product is more strained in 3d and is accompanied by more hydrogen bonds.

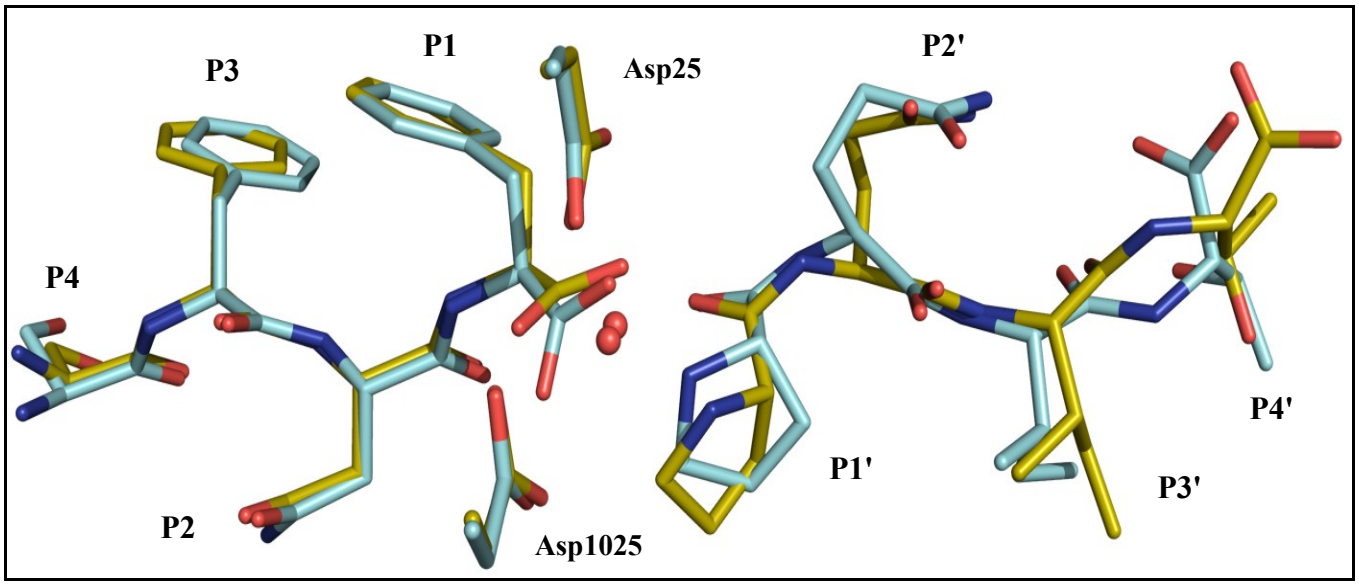
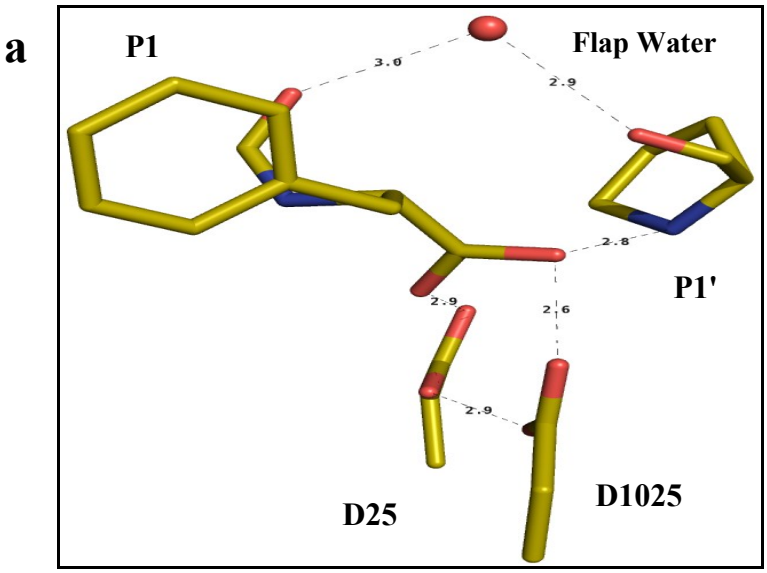


Figure 49: Superposition of refined coordinates of 1d (yellow sticks) and 3d (cyan sticks). Note that in 3d, P1' residue has moved away from and P1 has moved closer to the plane of the catalytic aspartates as compared to 1d.



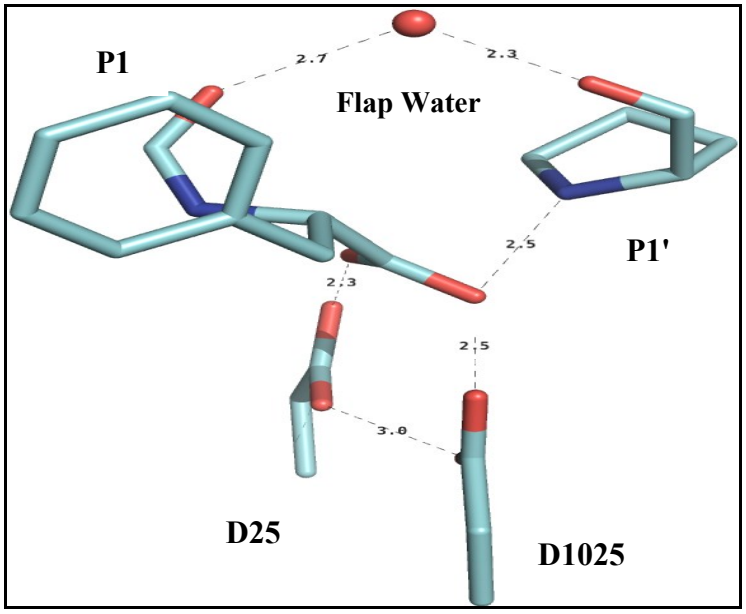


Figure 50: Hydrogen bonding in the catalytic center of HIV-1 protease/type-1 substrate complex for **a)** 1d (yellow sticks), and **b)** 3d (cyan sticks). Note that there is no hydrogen bond between P1' nitrogen atom with the enzyme.

The product release pathway has been reported by theoretical studies [498]. There are proposals that the P-product may be released first from the active site [149]. Comparison of a uncleavable reduced substrate (PDB ID: 2AOH), having the same sequence, with the present complexes with HIV-1 protease is shown in Figure 51 (rmsd with 1d is 0.46 Å and 3d is 0.44 Å).

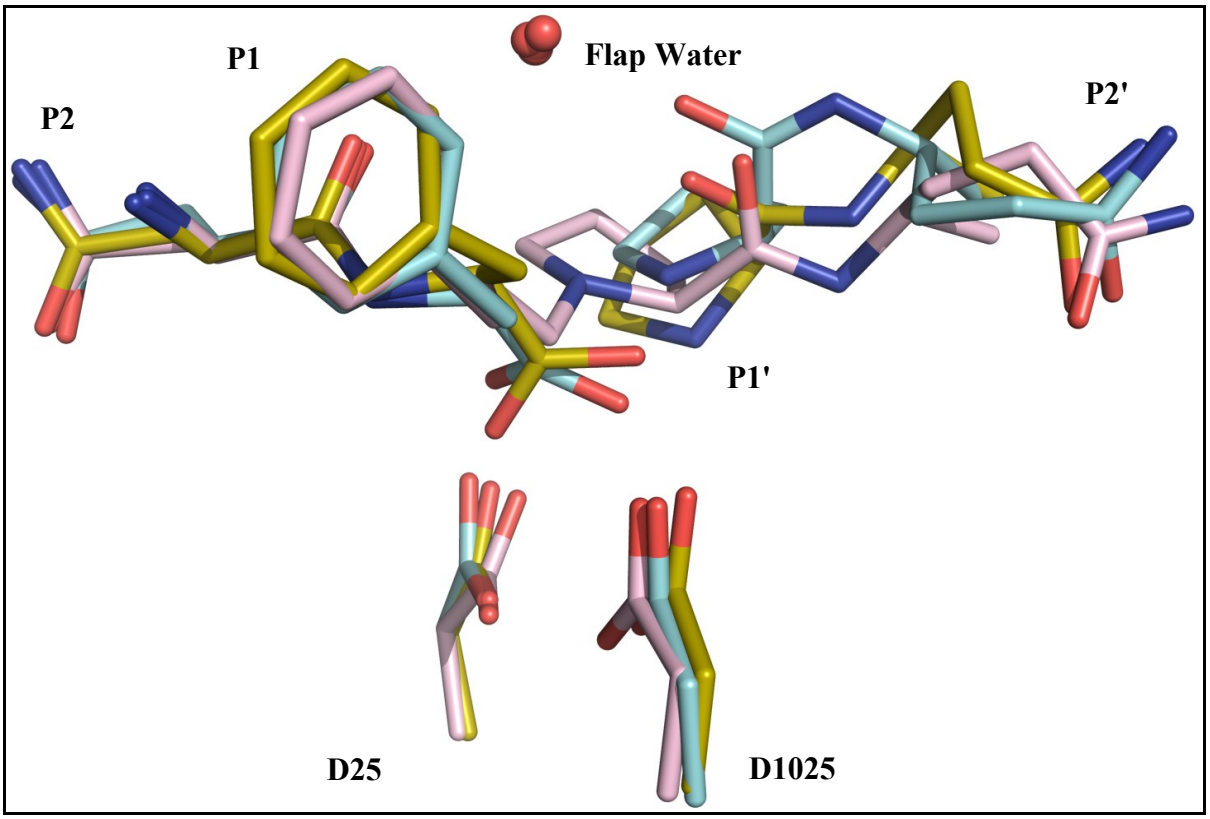


Figure 51: Superposition of refined coordinates of the peptide soaked in 1d (yellow sticks), 3d (cyan sticks) and uncleavable reduced substrate (pink sticks; PDB ID: 2AOH) complexed to active HIV-1 protease.

It is evident that the P1 Phe Cα's superpose perfectly whereas the P1' Cα's of Pro residue shows a lateral shift towards the active site exit. Comparison with the D25N HIV-1 protease / regular peptide complex indicates that for type-1 substrates the Q-product is released first. This can be seen in Figure 51 where the lateral shifts of the P1' Pro residue is more in 3d and less in 1d providing molecular level snapshots of product release. Thus the true substrate complexed to active enzyme exhibits more complex behavior manifested by its intrinsic flexibility.

4.3.5 Comparison between type-1 and type-2 bi-product complexes

Figure 52 shows the comparison between the *in situ* cleaved bi-product complexes of type-1 and type-2 substrates with HIV-1 protease (rmsds between 2NPH [460] and 1d is 0.27 Å and 3d is 0.40 Å). It is seen that 1d superposes better than 3d with the 3 days soaked type-2 complex [2NPH]. In the three complexes, P1 Cα-P1' Cα distances are 5.0 Å in 1d (type-1) and 4.6 Å in 3d (type-1) and 5.1 Å in type-2 [2NPH] structures, P1 C-P1' N distances are 3.6 Å in 1d (type-1) and 2.7 Å in 3d (type-1) and 2.7 Å in type-2 [2NPH] structures, and the D25 OD2 (D1025 OD2) - P1' N distances are 3.7 (4.8) Å in 1d (type-1) and 4.3 (4.9) Å in 3d (type-1) and 2.9 (3.1) Å in type-2 [2NPH] structures. The distances between the P1' Pro Cδ and D25 OD2 atoms are 3.1 Å in 1d and 4.0 Å in 3d. In 3d, this clash is relieved due to the lateral shift of the P1' residue. As the Q-fragment moves out, the P-fragment comes closer to the catalytic aspartate dyad. Thus the scissile bond, P1 C - P1' N distances

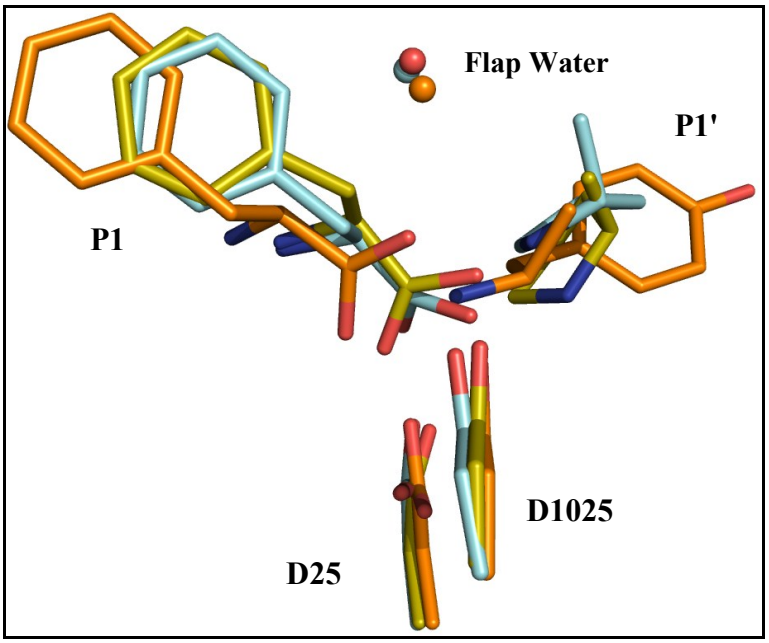


Figure 52: Superposition of refined coordinates of bi-product complexes of type-1 peptide, 1d (yellow sticks), 3d (cyan sticks) and the type-2 peptide substrate (orange sticks; PDB ID: 2NPH) at the catalytic centre of HIV-1 protease. Note the differences in the Cα positions between the type-1 and type-2 substrates.

in 1d becomes less than that of 3d, as P1 Phe carboxylate in the latter, makes stronger hydrogen bond with the catalytic aspartates (Figure 50). Such steric clash between the Pro Cδ is not applicable in type-2 product complexes. Thus the catalytic aspartates can rotate about Cβ-Cγ bond to have co-planarity, observed in the type-2 complex, modulating the hydrogen bond lengths. The D25 OD2 (or D1025 OD2) - P1' N distances gives an estimate of how much the Q-product has moved out of the active site, with respect to the plane of the catalytic center, after the products are formed. Indeed it is seen that the OD2-P1' N distances in type-2 product complex is least. As expected this distance in the 1d structure is lower than the 3d structure. Thus both the products (P and Q) have shifted laterally in the same direction with respect to the catalytic aspartates in the type-1 complex. The product release pathways are thus different in the type-1 and type-2 substrates. It is clear from Figure 52 that the P1 residue moves away from the catalytic aspartates more than the P1' residue in type-2 substrate. This is in contrast to the type-1 product release pathway where Q-product is released first.

4.3.6 Comparison of Phe-Pro and Tyr-Pro bi-product complexes soaked for 3 days

Comparision of 3d having Phe-Pro clevage site with an earlier 3 days soaked structure having Tyr-Pro clevage site (peptide sequence: SQNY*PIV; PDB ID: 3DOX) is shown in Figure 53. The two structures superpose with rmsd 0.26 Å. The Proline ring at P1' is rotated and the P1 Phe is also slightly rotated. As a result the carbonyl oxygen of Proline residue makes 2.3 Å hydrogen bond with the flap water in Phe-Pro complex whereas in Tyr-Pro this is weak (3.1 Å). The Tyr carboxylate is positioned across the two catalytic aspartates in the Tyr-Pro complex, each carboxylate oxygen making two hydrogen bonds with both inner and outer oxygens of the nearby catalytic aspartates. In the present Phe-Pro complex the orientation of the Phe carboxylate is almost perpendicular to the Tyr carboxylate (Figure 53) and makes one hydrogen bond with the outer oxygen of catalytic aspartates (Figure 50b). The bulk of para-hydroxyl oxygen atom in Tyr residue could be responsible for the change in orientation. The Proline nitrogen makes hydrogen bond with the carboxylate oxygen of Phe (2.5 Å) and Tyr (2.2 Å) in the two complexes. The interaspartate hydrogen bonds are similar, 3.0 Å and 2.8 Å in Phe-Pro and Try-Pro complexes respectively.

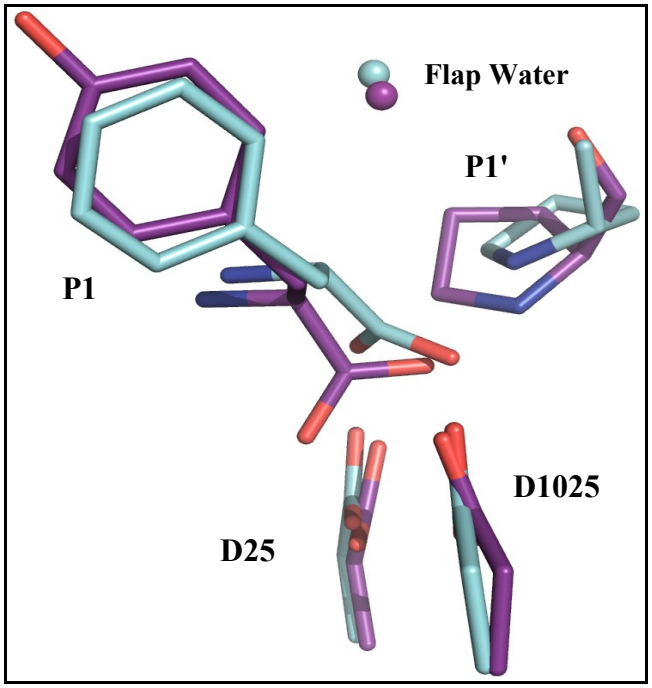


Figure 53: Superposition of refined coordinates of bi-product complexes of type-1 substrate peptide, 3d having Phe-Pro (cyan sticks) and Tyr-Pro (purple sticks; PDB ID: 3DOX) at the catalytic centre of HIV-1 protease.

4.3.7 Implications on the mechanism of type-1 substrate cleavage

Structures of enzyme substrate complexes provide atomic level mapping of the sequence of events that actually happen during the process of bond cleavage followed by product separation. The structures reported here reveal the atomic rearrangements at the catalytic center after product separation. In Figure 51 the sequence of the cleavage events can be envisaged. The reduced uncleavable peptide represents position of the substrate in the first step in the cleavage process. It is seen that after bond cleavage the P1 carboxyl group moves close to the catalytic aspartates in the present complexes (Figure 51). The displacement of the proline residue in P1' position is substantial. These facts suggest that positioning of the substrate in the active site is different in type-1 and type-2 substrates. In order to avoid the steric clashes involving the Cδ and Cγ atoms of proline with the outer oxygen of D1025, lateral shifts occur. The type-1 substrate P1' residues are Proline where the nitrogen is a secondary amide and cannot have a proton in the *Michaelis* complex and the tetrahedral intermediate. One of the important step in the mechanism is the protonation of the scissile nitrogen atom of the P1' residue, necessary for the C-N bond cleavage. The hydrogen bond between scissile nitrogen to the outer oxygen of the catalytic aspartates is absent in the present type-1 product complexes (Figure 50). Comparing type-1 and type-2 product complexes it is seen that the inter-product hydrogen bond between the oxygen of the newly formed P1 carboxylate and nitrogen

atom of P1' residue is present in both. Scissile nitrogen atom protonation mechanism proposed in Chapter 3, based on the type-2 substrate complexes, is also consistent with type-1 substrates. Interestingly, the cleavage mechanism is the same irrespective of whether the amino acids at the cleavage site are proline (type-1) or non-proline residues (type-2).

Another observation relevant to the mechanism is the separation between scissile carbon and nitrogen atoms. In 1d, this separation is 3.6 Å implying normal van der Waal separation without steric repulsion. In this complex the corresponding hydrogen bonding network including the interaspartate hydrogen bonds are relatively weak. Hower in 3d, the corresponding separation is 2.7 Å, which is indicative of steric repulsion. To overcome this repulsion few shorter and stronger hydrogen bonds are observed at the catalytic centre in 3d, but not in 1d. Interestingly in type-2 bi-product (PDB ID: 2NPH) complex, the repulsion due to 2.7 Å scissile carbon and nitrogen separation is compensated by the formation of intra-enzyme LBHB (2.3 Å) between the inner oxygens of the catalytic aspartates. In 3d, the inter-product steric repulsion is compensated in a different way by a intramolecular LBHB (2.3 Å, Figure 51) formed between the substrate P1 carboxylate oxygen with outer oxygen of the catalytic aspartate. Since these three structures are closely related and were obtained using almost identical experimental methodology, this correlation may be significant. It is therefore suggested that, in the reaction mechanism of HIV-1 protease, the hydrogen bond modulates differently, making up for inter-product steric repulsion, at different stages of the cleavage reaction.

4.4 Conclusions

This chapter reports two crystal structures, 1d and 3d, of *in situ* type-1 substrate bi-product/HIV-1 protease complexes. Both the structures indicate that the substrate adopts a *cis*-peptide conformation between the P1'- P2' residues, just next to the cleaved peptide bond. These structures however do not suggest whether the energetically higher *cis* conformation is adopted by the substrate prior to or subsequent to cleavage. The peptide ω-angle is lower by 30° in 3d suggesting that the formation of *cis* peptide may be a post-cleavage phenomenon. The more strained 3d has more number of hydrogen bonds than 1d. The Q-product having secondary amide is released first and P-product having the newly formed carboxylate binds more tightly in the active site of HIV-1 protease when the substrate is of type-1. This is in contrast to the type-2 product release pathway where P-product is released

first. Through these structures it is confirmed that the protonation of the scissile nitrogen is not by the outer oxygen of catalytic aspartates may occur intramolecularly via *gem*-diol of the tetrahedral intermediate. The structures of the type-1 and type-2 complexes, and especially the tetrahedral intermediate, will provide the best definition of the surface envelope [500] that would engulf all substrates at the catalytic centre. Future drug design projects should utilize both the interactions of tetrahedral intermediate complex and this substrate envelope, to design inhibitors which are less likely to give rise to drug resistance, "deadly" for HIV-1.

CHAPTER 5

Crystal structures of unliganded HIV-1 protease at different pHs

5.1 Introduction

As mentioned in Chapter 2, the pH activity profile of HIV-1 protease is bell shaped with optimum activity between pH values of 4.5 to 5.5. The expected difference in the protonation state of catalytic aspartates from being diprotonated at low pH to dianionic at high pH is thought to be responsible for pH-activity profile of HIV-1 protease [143]. Northop has proposed one of the mechanisms for activity of HIV-1 protease [150]. A critical feature of Northop's kinetic *iso*-mechanism is the postulate of a LBHB between the inner oxygens of the co-planar carboxyl groups of the two catalytic aspartates in the unliganded state. However, the atomic level structure of free HIV-1 protease is not known as function of low and high pH values. X-ray structures can suggest position of hydrogen atoms, only when determined to high resolution. With these facts in mind, crystal structures of wild-type HIV-1 protease were solved at three pH values of 2.0, 6.2 and 7.5. These three structures refined to 1.63 Å, 1.39 Å and 1.85 Å resolutions are reported in this chapter.

The positions of protons can be unambiguously determined by using thermal neutrons as probes [501]. But the lower neutron flux at research reactors requires crystalline samples to be much larger in size. Despite technological upgradation of neutron diffractometers on reactor beamlines, the volume required is still about 0.1 mm^3. Results of the attempts to grow large and deuterated HIV-1 protease crystals will aslo be given in this chapter.

5.2 Methods

The expression, purification and crystallization of HIV-1 protease have been discussed in Chapter 2. Using standard concentration of the protein, native crystals were obtained between pH 4.9 to pH 7.2 but they appeared as very thin and long fragile needles at the pH values away from pH 6.2. At other pH values there was precipitate without any crystals. Therefore crystals of free enzyme grown at pH 6.2, were soaked for 24 hours into buffer solutions having pH 2.0 and pH 9.0. This duration of soak is enough to equilibrate the bulk water protons with that of the active site in a crystal by physical exchange and *Grotthuss* mechanism [502-506]. The crystals were intermittently monitored under a microscope. At the extremes of pH values, most of the crystals developed cracks and fissures and many of them turned opaque. Few of the larger sized crystals were intact but fragile. They had to be handled gently during cryo-protection, prior to freezing and data collection.

X-ray diffraction data were collected using the method described in earlier chapters. The crystals above pH 7.5, and many crystals between pH 2.0 and pH 7.5 did not survive the whole data collection. However, one dataset each, at pH 2.0 and pH 7.5, could be successfully collected using soaked crystals. The X-ray datasets at pH 2.0 and pH 7.5 were collected on ID29 beamline using ADSC Quantum Q315r detector whereas the pH 6.2 dataset was collected on ID14-1 beamline using ADSC Quantum Q210 detector at ESRF. Crystal and intensity data statistics for these are given in Table 14. All the crystals belonged to $P6_1$ space group.

Table 14. Reflection data statistics of the unliganded HIV-1 protease at pHs 2.0, 6.2 and 7.5.

pH values :	**2.0**	**6.2**	**7.5**
Resolution range (Å)	50.0-1.63	50.0-1.39	50.0-1.85
Unit cell parameters (Å)	a=b=62.28, c=81.87	a=b=62.32, c=82.55	a=b=62.31, c=81.86
X-ray wavelength (Å)	0.98142	0.93340	0.98142
Total no. of data frames	120	120	120
Total no. of reflections	163139	268379	111456
No. of unique reflections	22439	35986	15423
Completeness (%)	100.0 99.7 (1.67-1.63)*	98.5 98.4 (1.55-1.39)*	99.9 100 (2.04-1.85)*
Mosaicity (°)	0.286	0.220	0.380
R_{merge} [#] (%)	6.2 47.9 (1. 67-1.63)*	10.5 72.3 (1.55-1.39)*	13.7 58.4 (2.04-1.85)*
$\langle I/\sigma(I) \rangle$	26.91 3.30 (1. 67-1.63)*	21.51 2.56 (1.55-1.39)*	10.97 2.82 (2.04-1.85)*

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

[#] *R_{merge} is given by $\Sigma_{hkl} \Sigma_i |I_i(hkl) - \langle I(hkl) \rangle| / \Sigma_{hkl} \Sigma i\, I_i(hkl)$, where $I_i(hkl)$ is the intensity of a reflection and $\langle I(hkl) \rangle$ is the average intensity.*

5.2.1 Structure solution and refinement

The structure was solved by the Molecular Replacement (MR) method [286, 287] and was refined using the protocol described in earlier chapters. The water molecules were added at the maxima common to both the *mFo-DFc* and *2mFo-DFc* maps. The occupancies of the active site water molecules were determined using methods described in earlier chapters after providing a starting value of 0.5. In the catalytic center the *3mFo-2DFc* maps and *5mFo-4DFc* maps were examined and their peaks coincided with the *2mFo-DFc* and SA omit maps.

5.2.2 Preparation of crystals for neutron diffraction

In order to obtain good signal:noise in neutron diffraction data, very large protein crystals ~1 mm^3 and perdeuterated sample are necessary [507-509]. Since perdeuteration is costly, simpler technique was used to equilibrate exchangeable protons with deuteriums. The salts and sodium acetate buffer containing DTT used for crystallization setup, were dissolved in 99.97 % D_2O and distilled and under vacuum at 45 °C. These steps, with the semi-dried salt, were repeated four times. The final stock solution in D_2O was kept in a dessicator in an air tight enclosure containing D_2O vapours. The refolded protein was exchanged with the deuterium exchanged sodium acetate buffer at pH 4.6 containing 1 mM DTT. The D_2O exchanges were done five times and the protein was concentrated to less than 1 mg/ml in each preparation. Higher concentrations resulted in slight precipitation which increased upon storage. Immediately several crystallization setups were made using hanging drop vapour diffusion method. The pD value in D_2O is obtained in the following way: pD = pH + 0.41 with pH corresponding to the value read with an electrode calibrated in H_2O buffer. The pD of reservoir solution consisting of deuterium exchanged phosphate-citrate buffer and concentration of saturated ammonium sulphate were varied in order to obtain few large crystals in a crystallization drop. The optimum condition for used for crystallization had pD 6.2 and 2-3 % ammonium sulphate. To obtain the required crystal size for neutron diffraction seeding was used. Seeding provides a template crystal which grows in size when placed a non-nucleated droplet. Macroseeding techniques introduce a single crystal into a pre-equilibrated protein solution [510, 511]. In microseeding, submicroscopic crystals are introduced into the protein droplets either by streaking or dilution methods [512, 513]. Further increase in size of the crystals in D_2O environment was achieved by both micro- and macro- seeding techniques.

5.3 Results and discussion

5.3.1 Crystals of deuterated HIV-1 protease

Figure 54a shows a crystallization drop containing typical unliganded wild-type HIV-1 protease crystals used in the pH studies. The crystals of HIV-1 protease in D_2O environment appeared after a week and inherently the number of crystals in each drop was consistently less than that in H_2O environment. The strength of deuterium bonds are stronger than hydrogen bonds by 0.1 - 0.2 kcal/mol

due to lower zero point vibration energy of the X-D bond (where X = O, N, S) attributed to the higher mass of the deuterium [514]. The enhanced structural rigidity in D_2O can be accounted for by both stronger intrapeptide D-bonds over H-bonds or formation of extra intramolecular bonds at the expense of bonds with the solvent. The greater propensity of D_2O, relative to water, to form

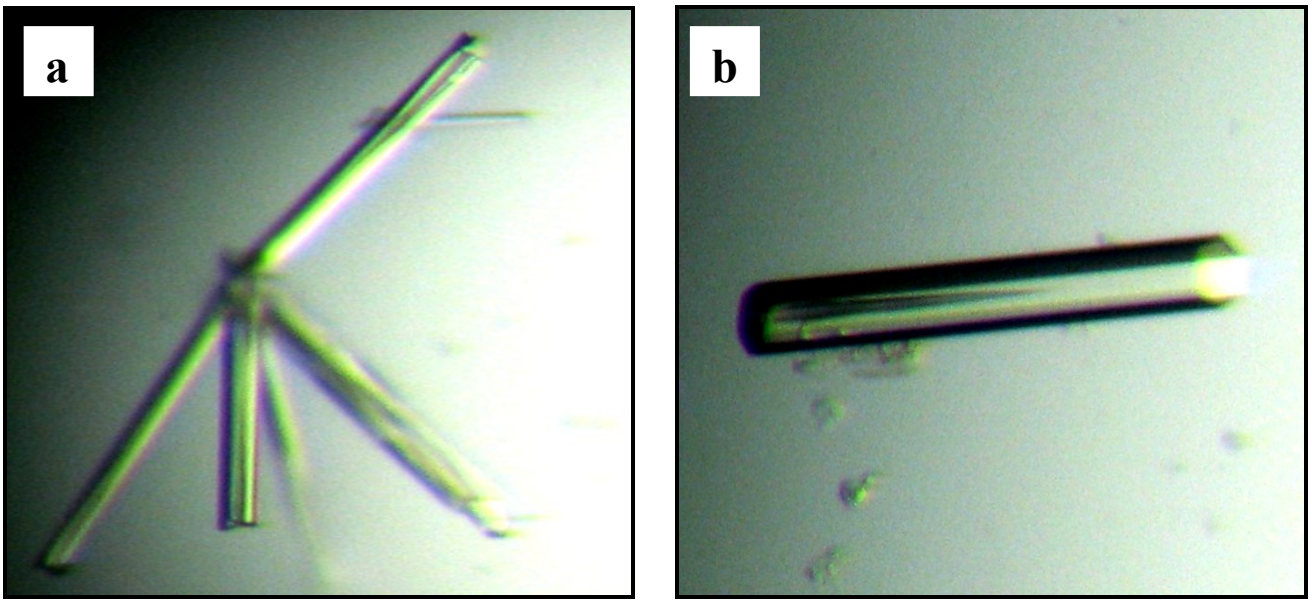


Figure 54: The crystals of wild-type HIV-1 protease grown in **a)** H_2O having 2 mg/ml protein, and **b)** D_2O having 0.89 mg/ml protein.

solvent-solvent D-bonds makes an important contribution to the latter process. Thus stabilizing effect of D_2O is attributed to the increase in hydrophobic interactions through additional folding/compaction of the native HIV-1 protease structure. These are the reasons for precipitation upon higher concentration and late appearance of crystals during crystallization of D_2O exchanged HIV-1 protease. Current neutron spallation sources having maximum flux, requires a minimum macromolecular crystal volume 1 mm^3 (~ 10^{15} ordered unit cells) in order to collect reasonable data. Efforts are underway to further increase the size of crystals in D_2O using other alternative techniques [515]. Figure 54b shows the largest crystal of deuterated HIV-1 protease grown to a volume of ~0.2 mm^3. Even medium resolution neutron diffraction data can determine H/D positions taking advantage of the sign reversal of scattering cross-section of H and D atoms [516-519]. Combined X-ray and neutron data can be used to accurately determine the proton positions during refinement [520, 521]. In fact, recently upgraded monochromatic D16 neutron beamline at the Institut Laue Langevin (ILL), has been used for low resolution data collection. This mode of data collection is under commissioning and will use cold neutrons to collect neutron diffraction data on crystals of 0.1 mm^3 or smaller [522].

5.3.2 Unliganded HIV-1 protease at different pHs

Table 15 shows the refinement statistics of the unliganded HIV-1 protease structures determined at pHs 2.0, 6.2 and 7.5. The structure at pH 6.2 is refined to very high resolution of 1.39 Å. Although the reflections extended beyond 1.35 Å with more than 90 % completeness, the highest resolution was cut off at 1.39 Å due to high R-merge values in the highest resolution shells. The stereochemistry of the refined models is good as seen from the RMSDs of bond lengths and bond angles.

Table 15: Refinement statistics comparing the unliganded HIV-1 protease at three pHs.

pH values :	**2.0**	**6.2**	**7.5**
Resolution range (Å)	50.0 – 1.63	50.0 – 1.39	50.0 – 1.85
R_{work} (%)	17.1 20.0 (1.67-1.63)*	16.6 18.2 (1.43-1.39)*	18.4 23.4 (1.89-1.85)*
R_{free} (%)#	21.0 23.8 (1.67-1.63)*	20.3 23.4 (1.43-1.39)*	20.9 26.6 (1.89-1.85)*
Number of protein atoms	1515	1515	1515
Number of solvent atoms	214	280	167
r.m.s. deviation of bond lengths (Å)	0.016	0.012	0.015
r.m.s. deviation of bond angles (°)	1.62	1.50	1.52
Average B factor (Å^2) for Protein atoms Water atoms	 30.2 47.2	 26.1 37.2	 32.1 49.1

 # R_{free} was calculated using 5% of randomly chosen reflections which were not used in the refinement throughout.
** Highest resolution bin used.*

5.3.3 Water molecules at the catalytic center

The continuous *2mFo-DFc* electron densities in the two flap regions were observed at the three pH values, indicating that the flaps were well-ordered within the crystal. In fact the flaps are found to adopt closed conformation between pHs 2.0 and 7.5, even in the absence of ligands. The discrete electron densities, common to *mFo-DFc* and *2mFo-DFc* in the active site of the structures at three pH values, indicated the presence of several water molecules. These water molecules, include the five that are found to be conserved in all unliganded structures [523, 524, 130]. Figures 55a-55c shows the omit electron densities for water molecules near the catalytic aspartates.

In the 1.39 Å structure at pH 6.2 (Figure 55b), two partially occupied water molecules, with occupancies of 0.65 and 0.35 and related by approximate 2-fold axis of symmetry, were modeled into the positive maxima of the *mFo-DFc* and *2mFo-DFc* electron density maps. The peak heights in

electron density maps correspond to their relative occupancies. These two modeled partial waters are separated by 2.0 Å, a distance which is too short to allow simultaneous existence of both water molecules (Figure 55b). Thus at pH 6.2 there is one catalytic water molecule at the catalytic center per HIV-1 protease dimer. A water molecule in an analogus position (the catalytic water) is observed in other unliganded closed-flap (PDB ID: 1LV1 and rmsd of protein Cαs is 0.3 Å) and open-flap (PDB ID: 2G69 and rmsd of protein Cαs is 0.57 Å) structures of HIV-1 protease. The water molecule having major occupancy in the present structure overlaps to within 1.4 Å and 1.0 Å from the corresponding water molecule in 1LV1 and 2G69 structures respectively. This water will be considered for hydrogen bonding interactions later in the text.

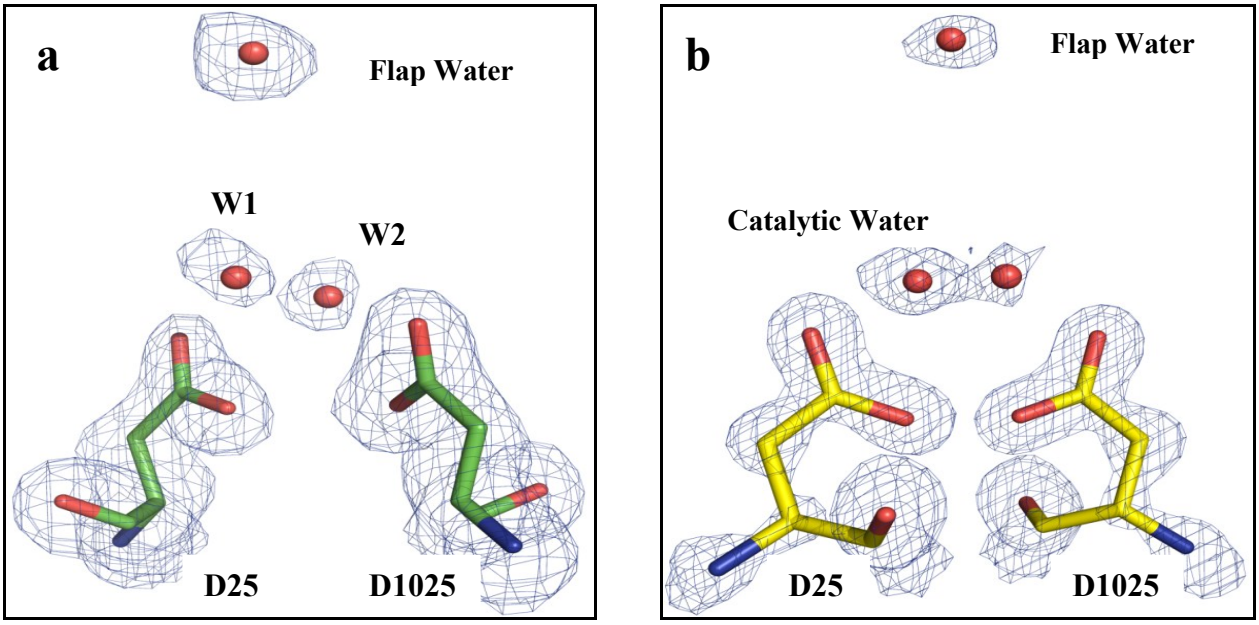
Similarly in the pH 2.0 and pH 7.5 structures, discrete and isolated peaks were present at the catalytic centre (Figures 55a and 55c). Interestingly there were two distinct positive peaks, and with respect to the nearby catalytic aspartic acids, they were positioned differently compared to what was found at pH 6.2. Occupancy refinement indicated full occupancy for these water molecules. The genuineness of these positions was confirmed by calculations of the SA difference maps with coefficients *3mFo-2DFc* and *5mFo-4DFc*. Thus in contrast to a single water molecule near the catalytic aspartates at pH 6.2, there are two fully occupied water molecules near the catalytic aspartates of HIV-1 protease at both pH 2.0 and pH 7.5 (Figures 55a and 55c).

The number and positions of water molecules around the D29/D1029 and D30/D1030 residues were also different at pH 2.0 and pH 7.5 (Figures 56a and 56c). At pH 2.0 the Asp's are protonated and can act as both donor (C-OH) and acceptors (C=O and C-OH) whereas at pH 7.5 the Asp's are anionic and can act as acceptors of hydrogen. At pH 2.0 there are more water molecules near the Asp residues due to the extra hydrogen on the Asp's. To explore how co-planar the carboxyl groups of the catalytic aspartates were, the planes were least squared fitted to different carboxyl groups. Table 16 shows the distances of the water molecules at the catalytic centre from the least squares planes (denoted by P_{D25}, P_{D1025} and $<P_{D25\text{-}D1025}>$) at the three pHs. The atoms C, OD1 and OD2 of D25 or D1025 or from both are used for least squares plane calculations. The least squares planes were calculated using software *Mercury* [525]. The aspartate planes are almost coplanar at pH 2.0 and the waters, W1 and W2 are close to the P_{D25} and P_{D1025} respectively. These waters are also closer to the average plane, and $<P_{D25\text{-}D1025}>$ between the two aspartates. This indicates that D25, D1025, W1 and W2 are almost

coplanar. The catalytic water molecule at pH 6.2 is not co-planar with any of the three planes in contrast to Northrop's mechanism. Similarly, at pH 7.5, W1 and W2 are quite far from P_{D25} and P_{D1025} and are situated at a similar distance from the average plane.

Table 16: The distances of the water molecules at the catalytic centre from the least square planes of the individual catalytic aspartates and average plane between them at the three pHs are shown.

pH → / Distance from Plane in Å ↓	2.0		6.2		7.5	
	W1	W2	Major Water	Minor Water	W1	W2
P_{D25}	0.158	1.104	0.754	0.900	1.003	1.363
P_{D1025}	1.092	0.094	1.235	1.014	0.099	1.937
$\langle P_{D25\text{-}D1025} \rangle$	0.220	0.047	0.838	1.234	1.111	1.100
Angle between P_{D25} and P_{D1025} (°)	4.58		19.13		27.00	
D25 OD1 --- D25 OD1 separation	3.1		2.8		2.9	



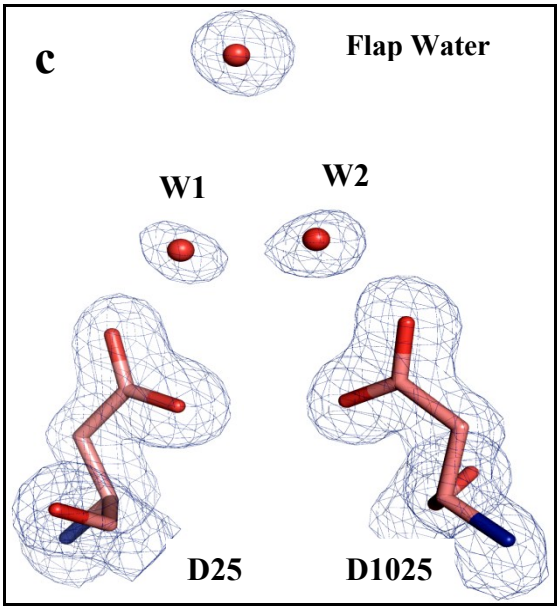


Figure 55: The SA omit map of the two catalytic aspartates and associated water molecule/s (W1 and W2) in the refined structures of unliganded HIV-1 protease at **a)** pH 2.0 (green sticks) contoured at 2.1 σ, **b)** pH 6.2 (yellow sticks) contoured at 2.5 σ and, **c)** pH 7.5 (salmon sticks) contoured at 2.0 σ. The atoms of D25, D1025 and the water molecules shown were not included in the structure factor calculations.

5.3.4 Hydrogen bonding interactions in the active site and comparison of unliganded HIV-1 protease at three pHs

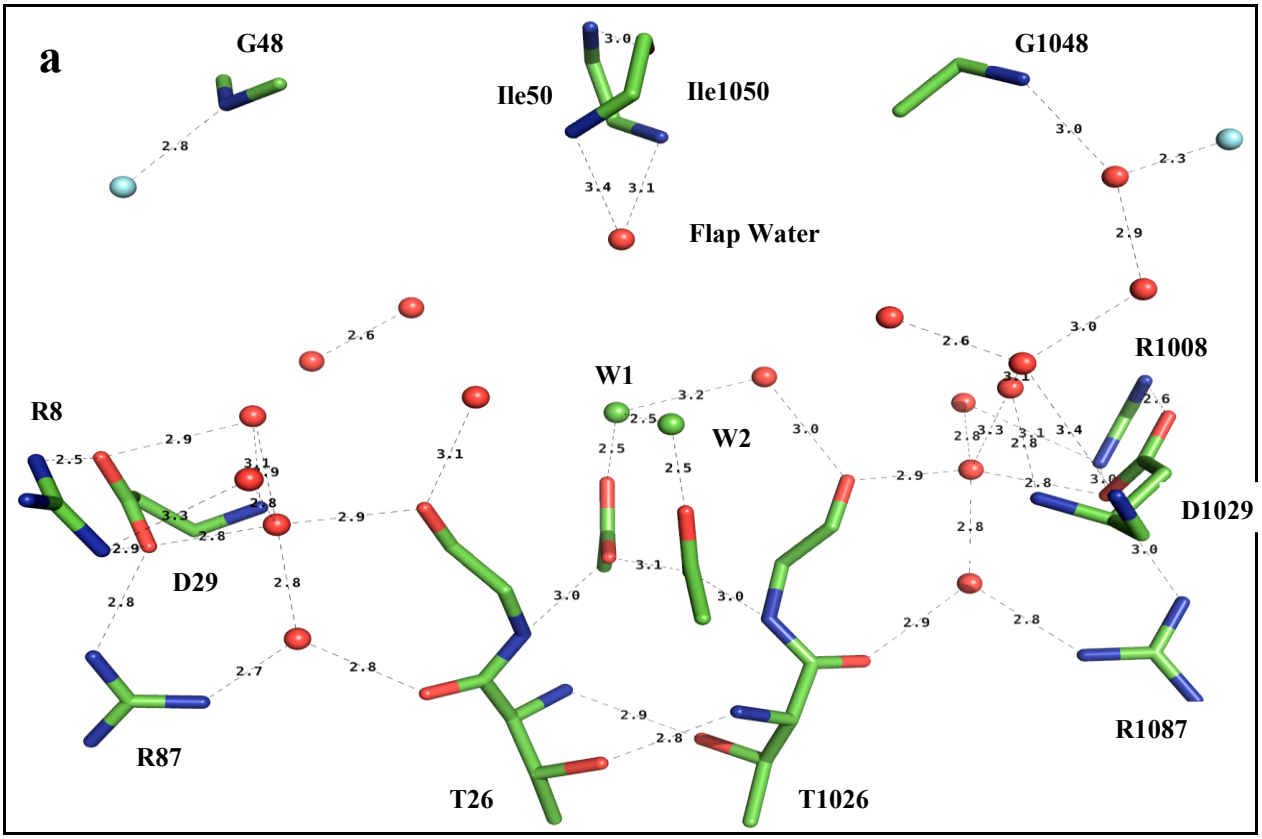
The hydrogen bonding interactions in the active site of unliganded HIV- 1 protease at pH 2.0 are shown in Figure 56a. There are twelve and nine hydrogen bonds from water molecules to protein backbone and side chains respectively. The flap water molecule makes two hydrogen bonds with the nitrogen atoms of the I50/I1050 (3.4 Å and 3.1 Å). The two water molecules (W1 and W2), shown as green spheres in Figure 57a, are strongly hydrogen bonded to each other (2.5 Å) and to the outer oxygens of both catalytic aspartates (both 2.5 Å). The positions of these waters are 1.3 Å / 1.4 Å and 1.9 Å / 1.5 Å away from the catalytic water molecule observed in other unliganded structures of HIV-1 protease, 1LV1 and 2G69 respectively. These two waters (W1 and W2) are coplanar with the carboxylate planes of the aspartates (W1-D25 OD2-C and W2-D1025 OD2-C) and are symmetrically placed. The carboxylate groups of two catalytic aspartates are also coplanar having only 4.6° deviation. The interaspartate distance is 3.1 Å which is weak hydrogen bond. D30/D1030 interacts with two nearby water molecules.

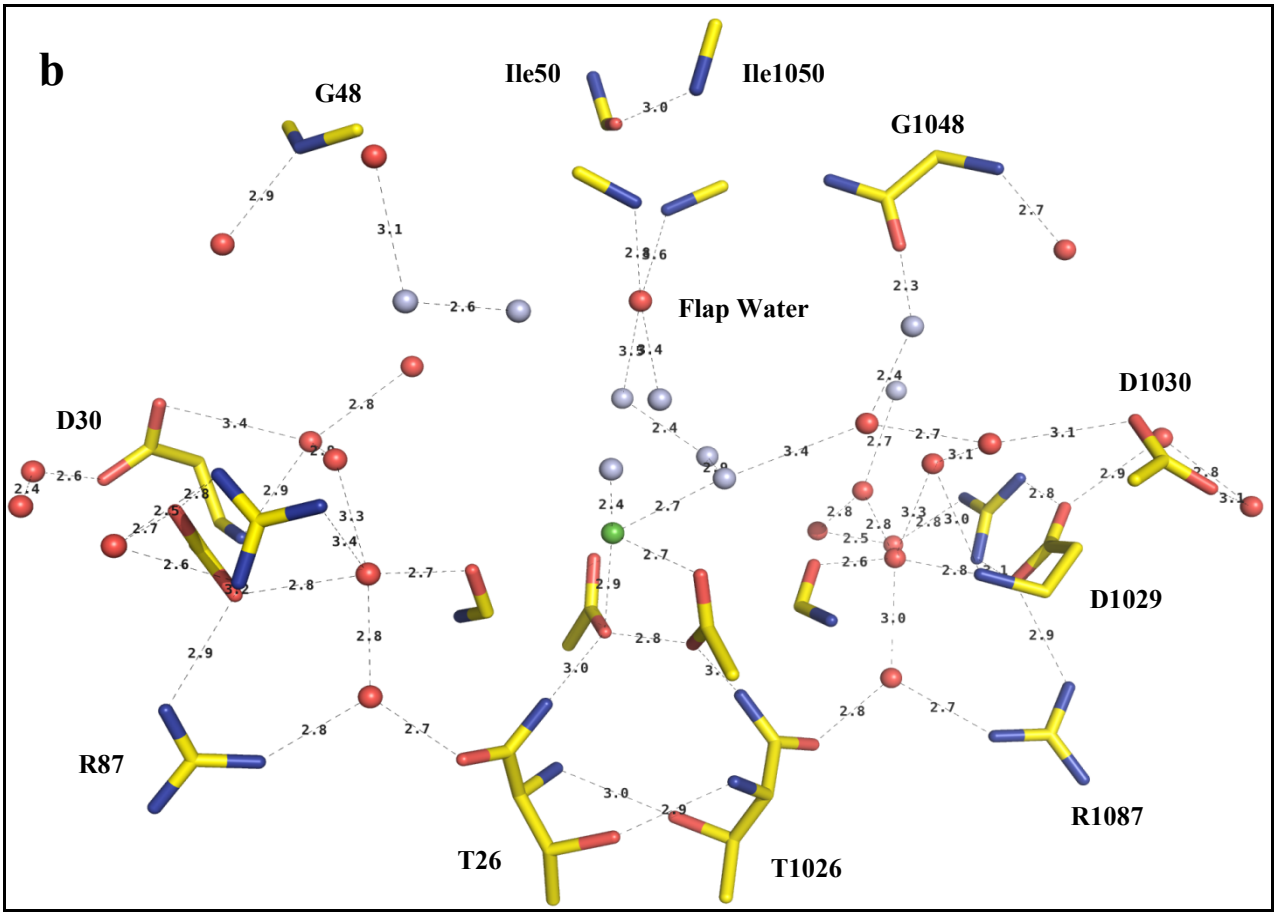
The water molecules in the active site of the pH 6.2 structure form a network of hydrogen bonds shown in Figure 56b. The flap water molecule forms a hydrogen bond with the nitrogen atom of the I1050 (2.8 Å) but there is no hydrogen bond with the nitrogen of I50 (3.6 Å) residues of the protein. The fully occupied water molecules in the active site superpose very near to water molecules present

in the 1LV1 and 2G69 structures (1.2 Å and 1.1 Å respectively). This indicates that the occupancies of individual water molecules during refinement are accurate and correct. The flap water is asymmetrically placed in a 2-fold symmetric active site. There are nine and ten hydrogen bonds between water molecules and protein backbone and side chains respectively. The catalytic water molecule having major occupancy lies on one side of the plane containing the oxygens of catalytic aspartates (Figure 57b). It forms two hydrogen bonds, one with the outer oxygen of Asp25 and the second with the inner oxygen of D1025. The interaspartate hydrogen bond length is 2.8 Å. The two carboxylate planes of catalytic aspartates are not co-planar but deviate by 19°.

In the pH 7.5 unliganded structure, there are fifteen fully occupied water molecules in the active site. The hydrogen bonding interactions in the active site are shown on Figure 56c. There are two water molecules (W1 and W2) near the catalytic aspartates strongly hydrogen bonded to each other (2.4 Å). They are situated across the planes of carboxylate groups of catalytic aspartates (Figure 57c). Each water molecule forms hydrogen bonds with the outer oxygen of the two catalytic aspartates (2.6 Å and 3.2 Å), although one of them is a very weak hydrogen bond. Thus these water molecules are asymmetrically placed with respect to the catalytic aspartates. The positions of these waters are 1.2 Å / 1.6 Å and both 1.2 Å away from the catalytic water molecule observed in other unliganded structures of HIV-1 protease, 1LV1 and 2G69 respectively. The interaspartate hydrogen bond length is 2.9 Å. The two carboxylate planes of catalytic aspartates are not co-planar but deviate by 27°.

The salt bridges between the D29/D1029 and R8/R1008 residues are present at all the pHs. At the catalytic center the unliganded HIV-1 protease structures at pH 6.2 and pH 7.5 are more similar than that at pH 2.0. While the rest of the structure remains the same (average rmsd of Cα 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 and each one, in turn, hydrogen bonds separately to the outer oxygen atoms of the two catalytic aspartates. The positions of the water molecules are distincty different at these three pHs as illustrated in Figure 57d.


a
G48
Ile50
Ile1050
G1048
Flap Water
R1008
W1
W2
R8
D1029
D29
R87
T26
T1026
R1087


b
G48
Ile50
Ile1050
G1048
Flap Water
D1030
D30
D1029
R87
T26
T1026
R1087

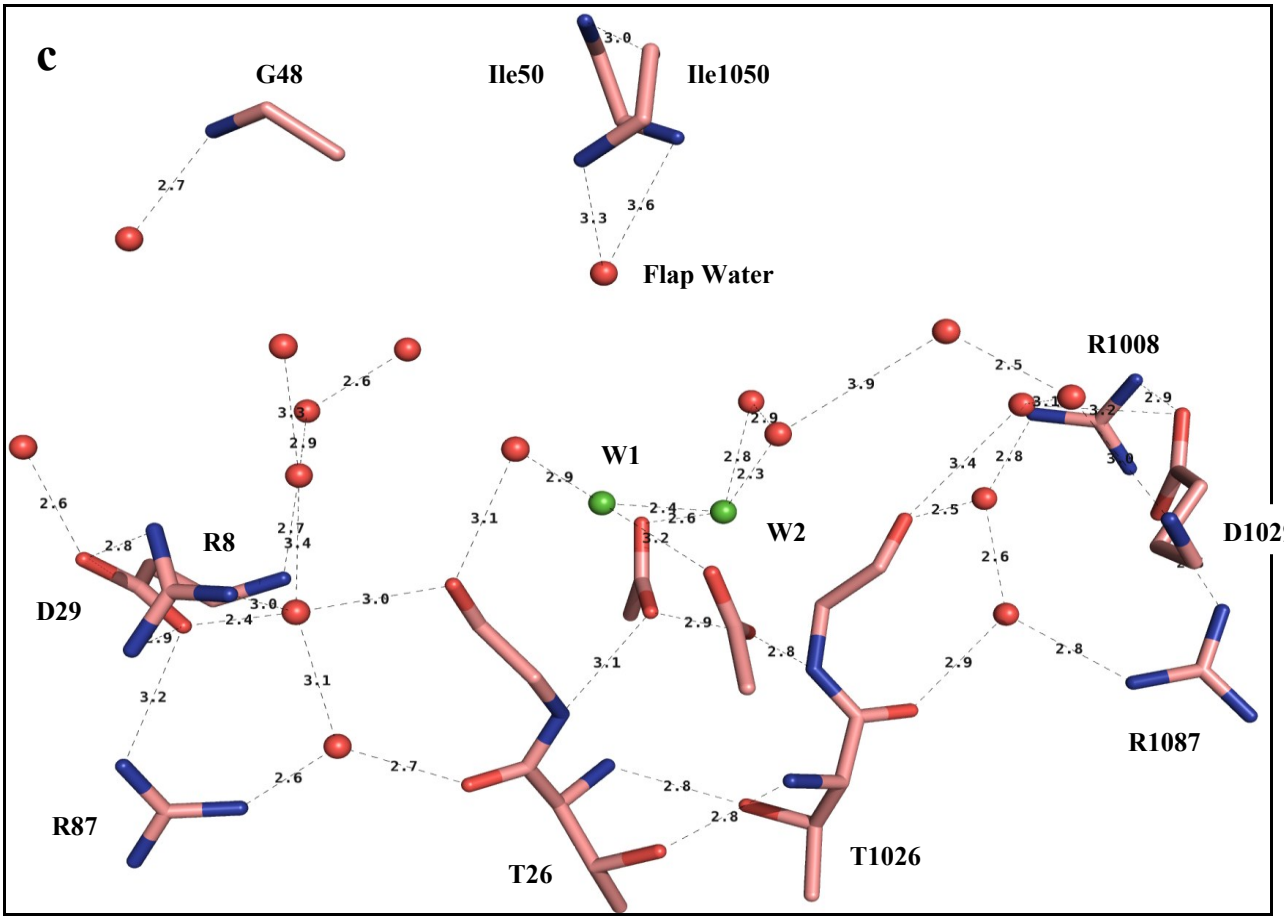
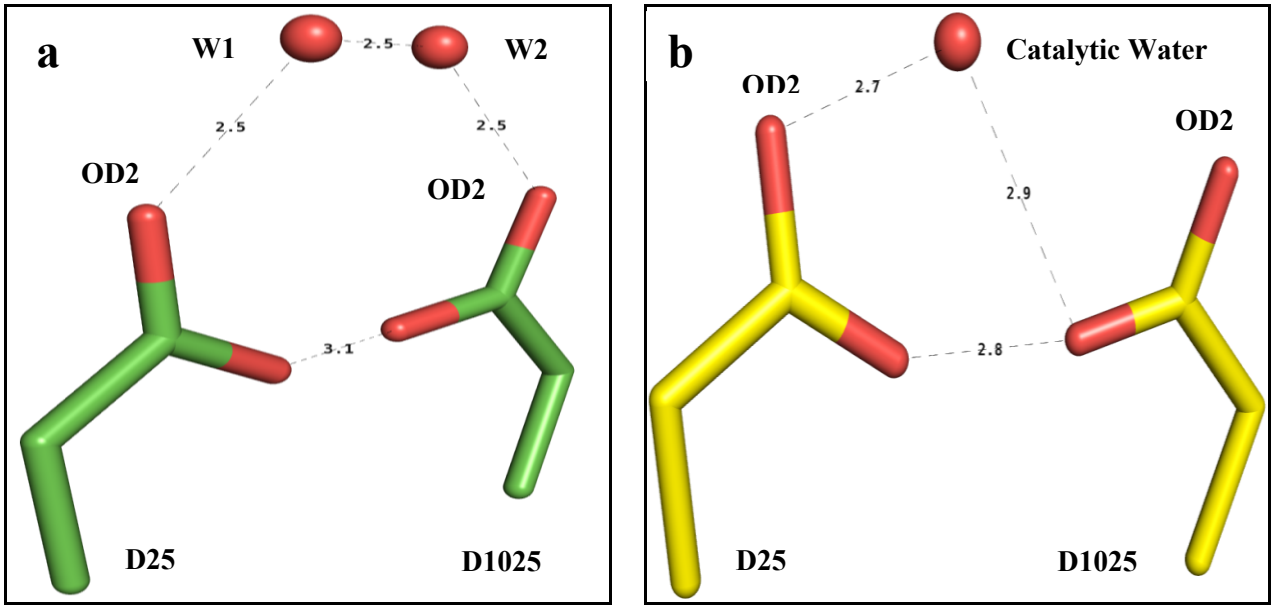


Figure 56: The hydrogen bonds in the active site of unliganded HIV-1 protease at **a)** pH 2.0, protein atoms are shown as green sticks, **b)** pH 6.2, protein atoms are shown as yellow sticks and catalytic water having major occupancy is shown, and **c)** pH 7.5, protein atoms are shown in salmon sticks. Waters at the catalytic site are shown as green spheres, fully occupied water molecules as red spheres and partial waters as cyan or light blue spheres.



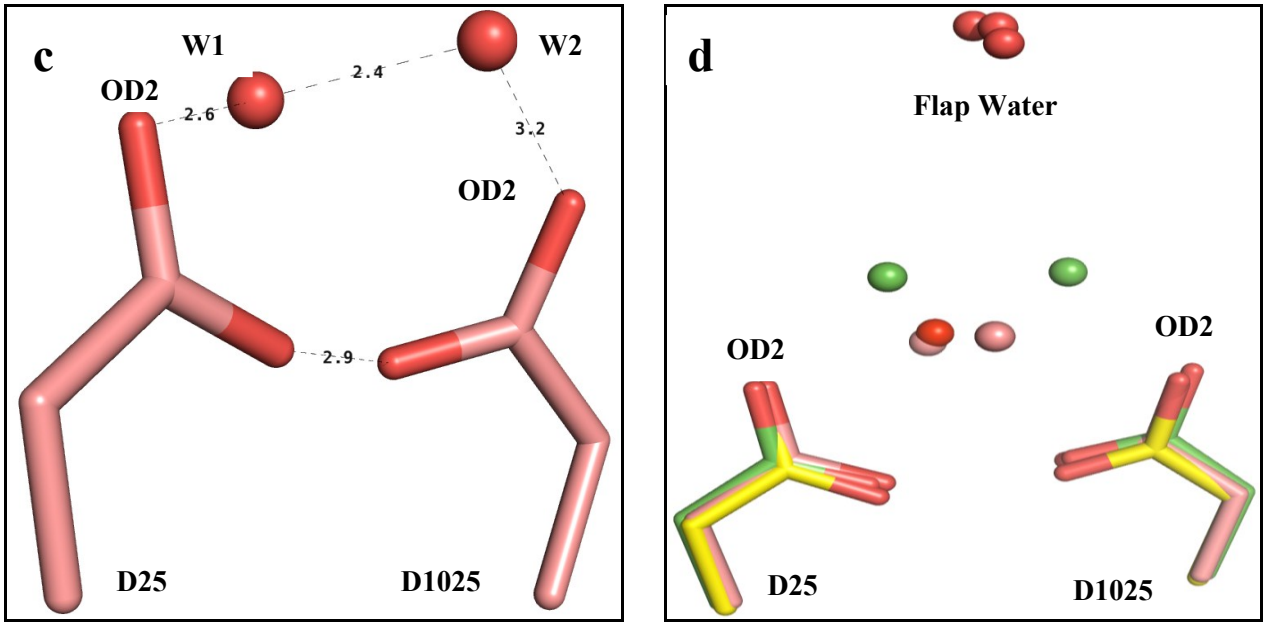


Figure 57: The hydrogen bonds at the catalytic center of unliganded HIV-1 protease at **a)** pH 2.0 (green sticks), note the two water molecules (W1 and W2) at pH 2.0 are coplanar to the catalytic carboxylates, **b)** pH 6.2 (yellow sticks), and **c)** pH 7.5 (salmon sticks). **d)** Superposed refined structures at the three pHs. Note that the water pairs (W1 and W2) at pH 2.0 and pH 7.5 are almost perpendicular to each other.

5.3.5 What are the protonation states in the catalytic centre ?

The protonation states of the catalytic aspartates can be determined using the same methodology based on hydrogen bond distances, as described in earlier chapters. A proton should be present between the short hydrogen bonds.

In the pH 2.0 structure (Figure 57a), there should be one proton between the two waters (W1 and W2), W1 and OD2(D25) and W2 and OD2(D1025). The short distance observed between W1 and W2 at pH 2.0, indicates a positive charge-assisted hydrogen bond [526]. Thus the homonuclear water dimer observed should be positively charged at pH 2.0 resulting in a hydronium-water cation, $(H_2O\text{---}H\text{---}OH_2)^+$. The 2.5 Å W1-W2 distance correlates well with theoretically calculated positive charge-assisted hydrogen bond distance [527]. The proton between W1 and W2 equilibrates dynamically converting W1 or W2 to water and hydronium cation, resulting in a symmetric hydrogen bonding pattern (Figure 57a). Since the pKa of hydronium cation (say W1) is lowest and pKa of water (say W2) is higher than catalytic aspartates, the protons would be attached to W1 and W2 rather than to OD2 atoms of catalytic aspartates. Thus there cannot be a proton on OD2 atoms having short distances to W1 and W2. It is also energetically not feasible to have inner OD1 atoms of both D25 and D1025 protonated. Therefore, we believe that the catalytic aspartate dyad is monoprotonated at pH 2.0. This proton should be between the inner oxygens of catalytic aspartates. One of the outer

oxygen of catalytic aspartates (OD2-D1025) exists as neutral oxygen and the other outer carboxylate oxygen (OD2-D25) is anionic. The hydronium cation (W1) donates its protons to W2, OD2 of anionic D25 and to a third water molecule (Figures 56a and 58). The water molecule (W2) donates a proton to OD2 of neutral D1025 (Figure 58, pH 2.0).

In the pH 6.2 structure (Figure 57b), the catalytic aspartates are monoprotonated with the proton situated between the inner oxygens. The two hydrogens of the catalytic water are associated with the outer oxygen of D25 and inner oxygen of D1025. Thus inner oxygen of D25 has the proton whereas D1025 exists as carboxylate (Figure 58, pH 6.2). Interestingly the catalytic water molecule has its hydrogen bond saturated so there are no full occupancy water molecules near it.

In the pH 7.5 structure (Figure 57c), there is a proton between the two water molecules (W1 and W2) and between the inner oxygens of catalytic aspartates. D25 is anionic having the negative charge on the outer oxygen which forms a 2.6 Å hydrogen bond with the water molecule (W1). This water molecule (W1) donates the other hydrogen to an active site water molecule (2.9 Å) (Figures 56c and 58). D1025 is protonated having the hydrogen on the inner oxygen. The water molecule (W2) near D1025 donates its hydrogens, one to the other water molecule (W1) and the other to outer oxygen of D1025 (Figure 58, pH 7.5). The water molecule (W2) makes two more bifurcated hydrogen bonds to two active site water molecules (Figures 56c and 58).

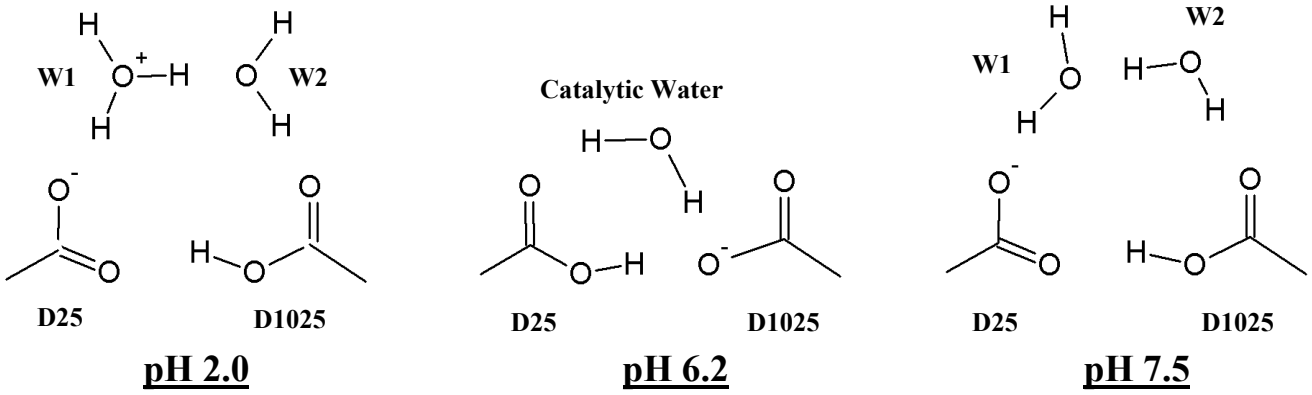


Figure 58: The protonation states at the catalytic center of unliganded HIV-1 protease at pH 2.0, pH 6.2 and pH 7.5.

5.3.6 Mechanism of rate reduction at low and high pHs

The crystallographic studies on HIV-1 protease/substrate complexes, presented in earlier chapters, were used to get better insights in the mechanistic pathway of the cleavage reaction. Crystallographic study of HIV-1 protease in complex with hydroxyl-based inhibitors showed that the central hydroxyl

group is situated between the catalytic D25 and D1025 residues, effectively replacing the putative catalytic water molecule present in the unliganded enzyme [155, 528, 529]. This hydroxyl represents the position of one of the oxygen atoms of the tetrahedral intermediate that is formed during catalysis. As described in Chapter 1, the peptide bond cleavage by HIV-1 protease is a multi-step process. These steps comprise of initial fitting of the substrate in active site, fine tuning the substrate-protease interactions via flap closure, proper orientation of cleavable peptide to have the near attack conformation, conformational rearrangement leading to collapse of TI and stepwise separation of the product fragments. All these steps have their own microscopic rates which affect the global rate of reaction.

It is generally believed that the non-optimal protonation state of the catalytic aspartates at lower and higher pH values is responsible for the reduction in enzyme activity. The present structures provide an alternative explanation, since the protonation states are found to be similar at these pH values. At the high and low pH values, there are two water molecules tightly bound to the catalytic aspartates and are located in altered positions compared to the single water molecule present at pH 6.2. One of the factors affecting the near attack conformation leading to *Michaelis* complex is proper stereochemical orientation of the scissile peptide bond relative to the attacking water molecule and the catalytic aspartates. Figure 59 below shows the expected trajectory taken by the attacking water molecule in nucleophilic addition reactions to sp^3, sp^2 and *sp* carbon atoms. These directions lead to maximal HOMO-LUMO overlap in the transition states [530-533]. LUMO is π^* molecular orbital (MO) of C=O and HOMO is provided by nucleophile. A productive scissile peptide bond requires the following simultaneous stereochemical requirements in order to form the *Michaelis* complex (see step 1 of proposed reaction mechanism in Chapter 3 and Figure 44): a) the scissile carbonyl oxygen should form a hydrogen bond with one of the outer oxygen of catalytic aspartic acid (polarizing the carbonyl), b) re-positioning of the catalytic water below the carbonyl carbon, having Burgi-Dunitz angle of 105° before tetrahedral intermediate formation (for proper HOMO-LUMO overlap, Figure 59) [534-537] and form a hydrogen bond with the inner oxygen the other catalytic aspartate (Figure 60). When the substrate occupies the active site, the hydrophobicity near the catalytic site increases, and the hydrogen bonds involving the water pair strengthens making them more rigid. Thus the presence of two rigid water molecules cause steric hindrance and the stereochemical requirements

(mentioned above) necessary for optimal orbital steering and overlap, in order to form the first transition state in HIV-1 protease catalysis, are not satisfied. Hence the rate of the reaction is drastically reduced at pH 2.0 and above pH 7.5, which is also experimentally observed [135]. The increased bulk due to these waters causes steric hindrance to the incoming substrate scissile peptide bond (Figures 61a and 61b). As the scissile peptide approaches towards the catalytic aspartates, in order to polarize the carbonyl bond, it repels these two water molecules resulting in expulsion of these water molecules from the catalytic site. Thus the nucleophilic water molecule required to form the Near Attack Conformation (NAC), which directly leads to the corresponding transition state, does not happen at higher pHs. It is also conceivable that the rigid water structure in the catalytic centre at pH 8.0, which is different from that at pH 6.2, guides the incoming substrate scissile peptide to bind in an unproductive fashion, enabling entrapment of uncleaved substrate at pH 8.0 as reported in Chapter 3. Thus the water structure could also contribute to non-productive binding of the substrate at higher pHs. In a productive reaction, the catalytic water molecule has to be appropriately placed with respect to the scissile carbonyl bond. Further evidence that the catalytic water at pH 6.2 is properly placed is derived from Figures 61a and 61b. In fact the catalytic water molecule at pH 6.2 is in close proximity (0.74 Å) to the O2 oxygen (derived from attacking water molecule) of the TI, the indicating its proper placement below scissile carbonyl group (1.9 Å from the carbon atom and the angle with the carbonyl plane is 94°). But the waters at high and low pHs are positioned differently from the TI oxygens (Figures 61a and 61b) and hence are not appropriately placed to form the TI. Comparing Figure 60 and Figure 61 it is seen that at pH 6.2, in the presence of the scissile peptide, the catalytic water is displaced towards the inner oxygen of D1025 and scissile carbonyl oxygen forms a hydrogen bond with the outer oxygen of D25 to form the *Michaelis* complex. This is evident from the decrease in hydrogen bond distances from 2.9 Å, between the inner oxygen of D1025 and catalytic water in native HIV-1 protease at pH 6.2 (Figure 57b), to 2.6 Å, between O2 atom of TI and the inner oxygen of D1025 in the TI complex. It may be noted that the D25 resonates to a form where the outer oxygen of D25 is protonated in order to form the crucial enzyme-substrate hydrogen bond (Figures 60 and 58, pH 6.2 and step 1 of proposed reaction mechanism in Chapter 3).

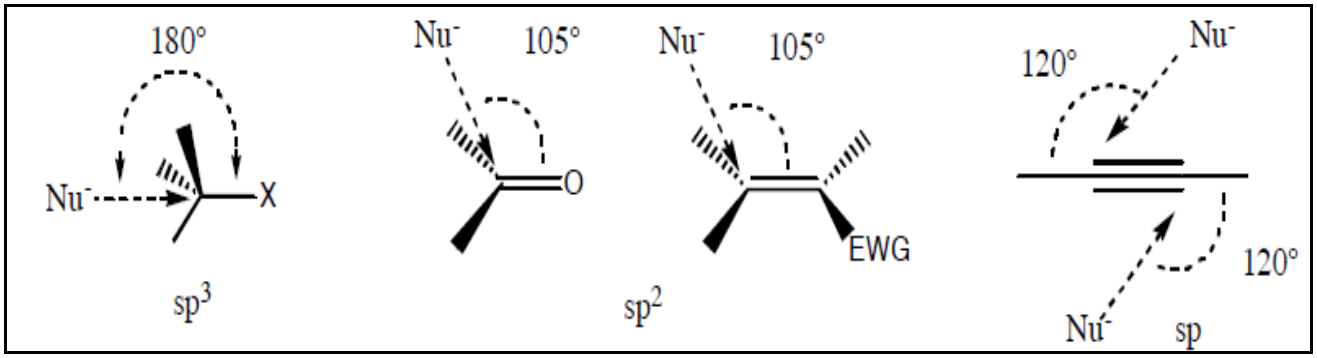


Figure 59: The trajectory for nucleophilic attack on sp^3, sp^2, and sp carbons are shown. In HIV-1 protease catalysis, the catalytic water molecule approaches the δ^+ polarized sp^2 carbon of the scissile carbonyl to form the TI. The 105° angle in addition reactions to carbonyl is called Burgi-Dunitz angle. EWG stands for electron withdrawing group.

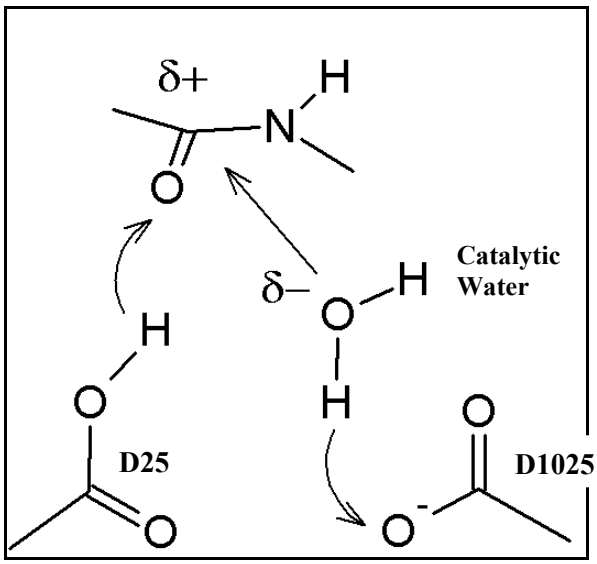


Figure 60: The Michaelis complex showing the protonation via hydrogen bonds (curved arrows) and catalytic water attacking the peptide polarized carbonyl bond (straight arrow).

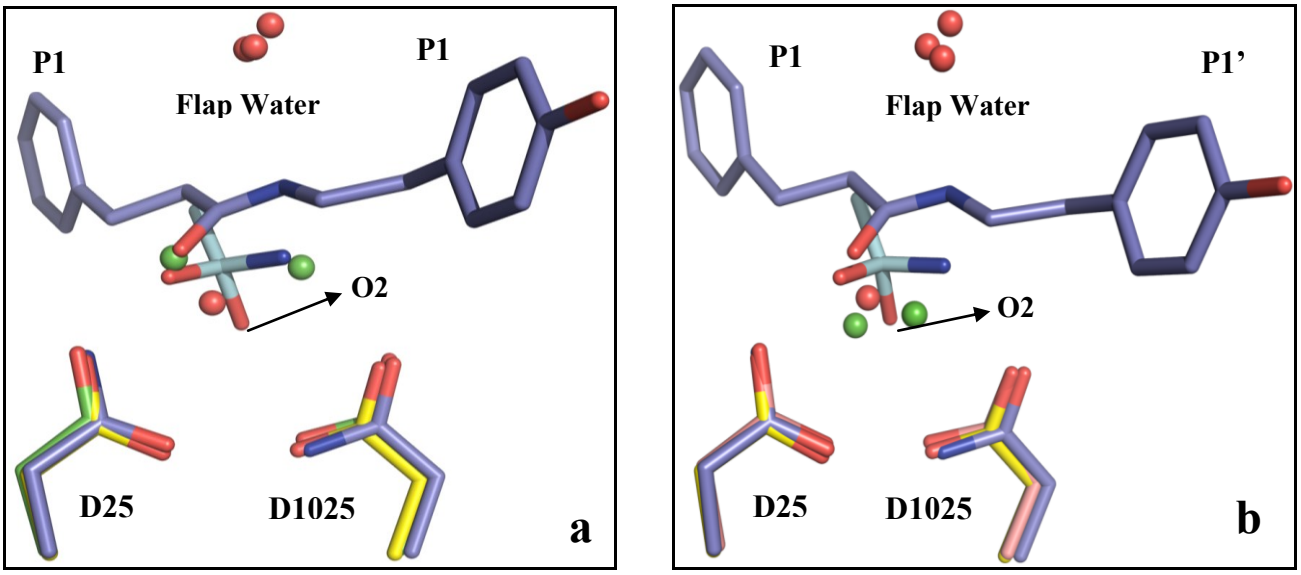


Figure 61: Superposed structure of unliganded HIV-1 protease (yellow sticks) at pH 6.2 onto **a)** pH 2.0 (green sticks), 1KJG (blue sticks) and TI (light blue sticks) structures, and **b)** pH 7.5 (salmon sticks), 1KJG (blue sticks) and TI (light blue sticks) structures. The pair of water molecules at pHs 2.0 and 7.5 are shown as green spheres and the *gem*-diol oxygen derived from catalytic water is labeled O2. Only carbonyl carbon tetrahedron is shown for TI model.

5.4 Conclusions

The structures of unliganded HIV-1 protease at a low pH of 2.0 and high pH of 7.5 reveals two water molecules in between the catalytic aspartates. The water dyad occupying the space near the catalytic aspartates cause steric hindrance to the incoming scissile peptide bond and disrupts the stereochemical requirements for water attack. These waters are observed for the first time and provide a new insight into the molecular level description of HIV-1 protease catalytic rate reduction as a function of pH. For example, these strongly bound water molecules may not be easily displaced, and may in fact force the incoming substrate to bind in a non-productive manner. The LBHB invoked by Northrop for the free enzyme is not observed at pH 6.2, where the enzyme is active.

CHAPTER 6

Crystal structures of wild-type and drug-resistant mutants of HIV-1 protease complexed to Ritonavir

6.1 Introduction

As described in earlier chapters, drug-resistant viral strains develop very quickly and these have been evolving continuously to become more drug-resistant through mutations, either in the protease sequence or in the cleavage-site sequence [126, 174, 538-540]. Some of the protease inhibitors developed earlier as drugs against HIV/AIDS are: saquinavir, ritonavir, indinavir and nelfinavir [541]. Ritonavir, developed by Abbott Laboratories, was the second HIV-1 protease inhibitor that the FDA approved in 1996 for the treatment of HIV-1 infection [542, 543].

The main driving force for the binding of these four inhibitors to the wild-type HIV-1 protease (WT) is a large positive entropy change originating from the burial of a significant hydrophobic surface upon binding [544, 545]. Since the inhibitors are pre-shaped to match the geometry of the binding site, their conformational entropy loss upon binding is small, a property that contributes to their high binding affinity. At 25°C, the binding enthalpy is unfavorable for all inhibitors except ritonavir, for which it is slightly favorable (-2.3 kcal/mol). Ritonavir is derived by optimization of the pharmacokinetic properties of a series of C2 symmetry-based and peptidomimetic protease inhibitors [546-549] and is currently used in combination with other drugs in Highly Active Antiretroviral Therapy (HAART). Genotypic analysis of the HIV-1 protease gene isolated from patients undergoing ritonavir monotherapy revealed a stepwise accumulation of multiple mutations at nine different codons, and the chart below shows these mutations [550]. Shown in larger bold font, in the chart below, are the major mutations along with other minor mutations.

	L	K	V	L	M	M	I	I	A	V	***V***	***I***	L
Ritonavir	10	20	32	33	36	46	50	54	71	77	***82***	***84***	90
	FI RV	MR	I	F	I	IL	V	VL	VT	I	***AF TS***	***V***	M

The major mutations appear at the 82nd and 84th residues of HIV-1 protease. The initial mutation at position 82 was consistently observed in all patients, and this mutation appeared to be necessary for the primary loss of antiviral effect [185, 551]. In earlier reports it has been shown that binding of ritonavir is reduced by 90-fold to V82F mutant, and by 10-fold to I84V mutant [552]. Circulatory

Recombinant Form_01 (CRF_01) A/E strain of HIV-1 selects V82F mutation in its protease, to overcome all the drugs used in the study [553].

The development of resistance involves multiple mutations in the enzyme's active site and non-active site [554]. The underlying mechanisms driving the evolution of drug resistance in the non-active site of HIV-1 protease are only partially understood [545, 555, 556]. A genotype trial identified 21 mutations at 16 positions: L10V, I13V, K20MRV, L33F, E35G, M36I, K43T, M46L, I47V, I54AMV, Q58E, H69K, T74P, V82LT, N83D, and I84V [557]. An updated trial excluded I13V, K20MRV, E35G, and H69K mutations and reclassified I47V, I54AMV, Q58E, T74P, V82LT, and N83D as major mutations; L10V, M36I, K43T, M46L, and I84V as minor mutations; and included L24I, I50LV, I54L, and L76V as mutations likely to improve drug susceptibility and virological response [558]. M36I mutation is observed in increasing proportions in non-subtype B clades of the HIV-1. Forty-nine percent had mutations in the hinge (M36I, R41K, H69K) and alpha-helix (L89M) regions of the C-virus protease, which has been linked to increased catalytic activity among HIV-1 isolates from treatment-naive individuals in North India [559]. In another study M36I mutation appeared in almost all the samples associated with drug resistance against ritonavir, and appeared after active site mutations as a minor drug resistant mutation [560, 561]. The structural basis for the exact role of M36I mutation as compensatory mutation is not clear. This is because the 36th residue of protease is not located at the active site of protease and has no direct interaction with any substrates or any protease inhibitors.

In the absence of crystal structure of V82F/ritonavir complex, several computer modeling studies have investigated effects of the V82F substitution on the structure of HIV-1 protease, and also on the interactions of HIV-1 protease with ritonavir [126, 548, 549]. In one molecular dynamics simulation study the flaps opened farther and were more flexible in V82F mutant than in the WT, and this dynamics was cited as the reason behind drug-resistance [562, 563]. In another molecular modeling study, the F82 residue was found to be sterically clashing with the benzyl rings at P1/P1’ sites. It was suggested that adjustments to protein conformation required to relieve this steric clash would likely lead to further decreased interactions with the P1’ benzyl group [564]. Results of molecular modeling and simulations depend critically on the correctness of the starting structure, and if simulations are carried out with an incorrect starting structure, the errors, instead of getting corrected, would also

propagate with long simulation runs needed to understand functionality. Therefore crystal structures of these complexes are necessary to understand the mechanism of drug resistance.

In this chapter, crystal structures of complexes between wild type, active site V82F mutant, and non-active site M36I mutant of HIV-1 protease and ritonavir will be presented.

6.2 Methods

6.2.1 Cloning

A pET11a-protease wild type (WT) tethered dimer construct was used for site specific mutagenesis as described in Chapter 1.

Using the site-directed mutagenesis procedure of Quickchange (Stratagene Inc.), Val82 to Phe82 amino acid substitution, was introduced to prepare the pET11a-protease-V82F (V82F) gene. The following primers were used:

1. V82F_sense

5'-gtgggcccgactccg***ttt***aacattatcggcc-3'

2. V82F_antisense

3'-cacccgggctgaggc***aaa***ttgtaatagccgg-5'

The underlined codon codes are for Phe at the 82nd amino acid position.

Similarly, M36I point mutation was introduced in the WT gene to prepare pET11a-protease-M36I gene (M36I), following the same procedure. The following primers were used:

1. M36I_sense

5'-tgtactggaggag***ata***tctctcccgggcc-3'

2. M36I _antisense

3'-acatgacctcctc***tat***agagagggcccgg-5'

The underlined codon codes are for Ile at the 36th amino acid position.

The mutations were confirmed by sequencing the plasmid DNA.

6.2.2 Expression and purification

The codons were incorporated to prepare V82F and M36I plasmids, which were inserted into *E. coli* BL21 (DE3) expression cells. Procedures similar to those described earlier [88, 95, 424, 460] were used to express and purify the WT, V82F and M36I proteins. The V82F and M36I mutant protein

yields were less than and similar to the WT protein respectively. The proteins were concentrated to approximately 5-8 mg/ml before crystallization. Purity was checked by sodium dodecyl sulfate-polyacrylamide gel electrophoresis analysis.

6.2.3 Crystallization

Crystal of WT/RTV complex were prepared by both soaking and co-crystallization methods. Crystals of WT protein, used for soaking, appeared under conditions reported in **Chapter 3**. Ritonavir was dissolved in reservoir solution containing 10 % DMSO to prepare a 1 mM soaking solution. Crystals were soaked for 3 days at room temperature using the method described in **Chapter 3**. All mutant complexes were generated through co-crystallization. Stock solution of ritonavir prepared in dimethyl sulfoxide (DMSO) was added, in 8-10 molar excess, to WT, V82F and M36I proteins present in 20 mM sodium acetate, pH 5.0. The mixtures were incubated on ice for 2 hours before crystallization trials using ammonium sulfate dissolved in sodium citrate/disodium hydrogen phosphate buffer as the precipitant. A reservoir volume of 1 ml and a drop volume of 2 μL (1:1 and 2:1 for mixture of protein and reservoir solutions in V82F and M36I respectively) were used. The optimal reservoir conditions for V82F were found to be 63 mM sodium citrate and 126 mM disodium monohydrogen phosphate, pH 6.4 with 5 % saturated ammonium sulfate. The M36I co-crystals appeared in 63 mM sodium citrate and 126 mM disodium monohydrogen phosphate, pH 6.0 with 2.5% saturated ammonium sulfate. Finer grids were made to grow larger crystals of the mutants. Long rod-shaped crystals suitable for diffraction measurements, typically grew within 7-10 days (Figure 62).

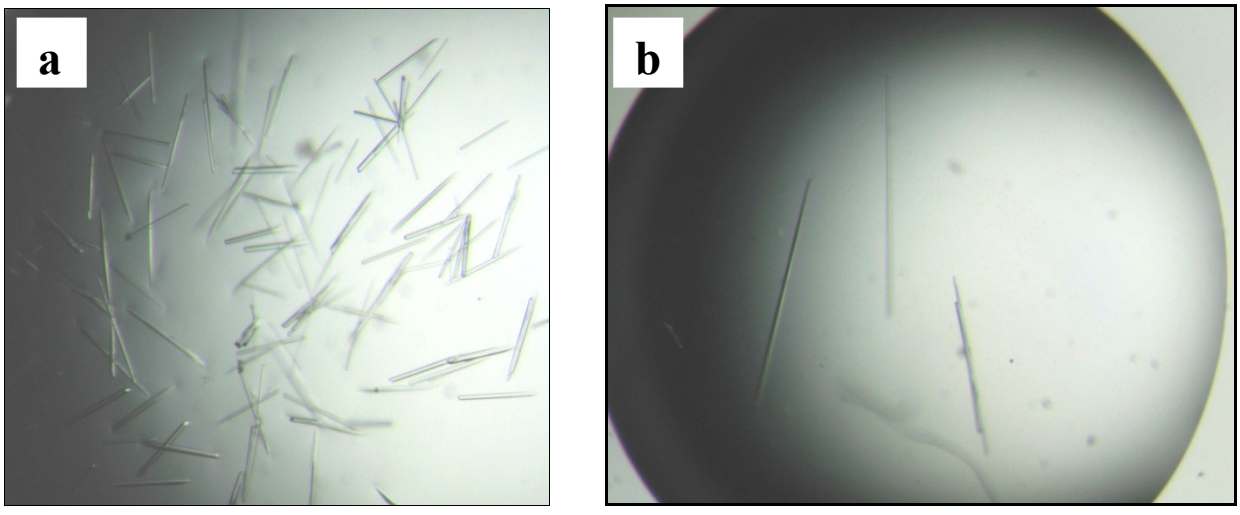


Figure 62: Crystals of **a)** V82F/ritonavir and **b)** M36I/ritonavir complexes obtained by co-crystallization.

6.2.4 X-ray data collection and processing

WT/RTV complex

A WT complex crystal, prepared by the soaking method, mounted inside a 0.3 mm quartz capillary was used for diffraction data collection on a RAXIS-IIC diffractometer installed on a RU200 X-ray Cu-Kα generator. The data was collected by the oscillation method, in which each frame had 1° oscillation range and exposure time of 30 minutes. Crystals were found to be sensitive to radiation, since diffraction intensities progressively decreased on exposure of the crystal to X-rays. The first 30 images were included in the data integration and data scaling using *HKL* program [241].The crystals were found to give useful diffraction to 2.5 Å resolution.

Crystals of the complex between WT protein and ritonavir obtained by co-crystallization were screened for their diffraction quality on FIP BM30A beamline [439] at ESRF tuned to 0.979966Å radiation. An ADSC two-by-two mosaic CCD detector set to the 1-K binned mode was used to record the images, while the cold stream from Oxford Cryosystems model 600 liquid nitrogen cooler maintained the mounted crystal at 100K. On one of the good crystals of the WT/RTV complex, a dataset was collected using ID14-4 beamline at ESRF. Complete diffraction data was recorded to 1.60 Å resolution using 120 contiguous frames. The images were processed with *XDS* suite [252, 253].

V82F and M36I mutant/RTV complexes

After similar screening of the co-crystals of V82F and M36I mutant complexes, final datasets were collected at FIP BM30A beamline at ESRF. The diffraction data on M36I mutant complex was collected remotely on BM30A beamline from Remote Data Collection Facility situated in HBNI, Mumbai. Complete diffraction data were recorded as 120 contiguous oscillation frames. The images were processed with *XDS* suite [252, 253]. All the crystals exhibited symmetry consistent with space group $P6_1$. The statistics for the processed diffraction data are given in Table 17.

Table 17: Summary of crystal and intensity data statistics of WT/RTV, V82F/RTV and M36I/RTV complexes.

Parameters	WT/RTV (soaking)	WT/RTV (co-crystal)	V82F/RTV (co-crystal)	M36I/RTV (co-crystal)
Resolution range (Å) Outer resolution shell	50.0–2.50 (2.66–2.50)	50.0–1.60 (1.64–1.60)	50.0–1.90 (1.95-1.90)	50.0–1.60 (1.64–1.60)
Unit Cell (Å)	a=b=63.34, c=83.21	a=b=61.87, c=81.60	a=b=62.50, c=82.63	a=b=62.35, c=82.95
Wavelength (Å)	1.5418	0. 96850	0.979996	0. 97975
No. of total measured reflections (unique)	21871 (5547)	140826 (23169)	106039 (14477)	210826 (22662)
Completeness (%)	83.8 (78.8)*	94.7 (86.9)*	94.2 (80.2)*	93.9 (76.5)*
I/σ (I)	6.8 (0.8)*	15.7 (3.0)*	12.9 (2.8)*	26.2 (3.0)*
R_{merge} [#] (%)	8.7 (92.1)*	6.9 (59.6)*	11.0 (54.2)*	7.75 (60.4)*

**The numbers between parentheses indicate the value in the outer resolution shell.*
[#] *R_{merge} is given by* $\Sigma_{hkl}\Sigma_i |I_i(hkl) - \langle I(hkl)\rangle|/\Sigma_{hkl}\Sigma i\, I_i(hkl)$*, where* $I_i(hkl)$ *is the* i*th measurement of the intensity of a reflection and* $\langle I(hkl)\rangle$ *is the average intensity*.

6.3 Results and discussion

Since the present crystals are isomorphous to those of the unliganded protein structure (PDB ID: 1LV1), atomic coordinates were extracted from this structure [412, 423] and used as the starting molecular replacement model in the software package *Phaser* [324]. The results are summarized in the Table 18. The atomic coordinates were extracted from the top molecular replacement solution and used as the starting model for refinement. In the V82F and M36I mutants, the Val residues at the 82 and 1082 positions and Met residues at the 36 and 1036 positions of the tethered dimer were mutated to Phe and Ile respectively, using *O* [366]. The resultant protein coordinates after molecular replacement were used for rigid body refinement using the software package *CNS*, employing standard simulated annealing (SA) protocols to minimize the amplitude based maximum likelihood target function [347 370]. A total of 5 % of randomly selected reflections were set aside for cross validation [393]. The coordinates of the drug molecule in 1RL8 were extracted from PDB. The restraint and the topology used for refinement were obtained using *PRODRG* [441]. In the initial stages of simulated annealed (SA) refinement and during calculation of SA omit maps, the model was heated to a temperature of 3000 ^{o}K, and then annealed at a cooling rate of 100 ^{o}K per iteration. The

Table 18: The MR solutions for V82F and M36I mutants using 1LV1 as search model.

****

Phaser Module 2.1.4 for V82F mutant complexed to ritonavir

SPACegroup HALL P 61 #P 61

SOLU SET 1

RFZ=17.4 TFZ=30.8 PAK=0 LLG=1097 LLG=1744

SOLU 6DIM ENSE ensemble1

EULER 267.090 0.615 332.944

FRAC -0.00623 0.00530 & -0.33402

SOLU SET 2

RFZ=16.9 TFZ=35.5 PAK=0 LLG=1098 LLG=1743

SOLU 6DIM ENSE ensemble1

EULER 212.223 179.379 152.292

FRAC 0.00507 -0.00642 & 0.33749

****

****

Phaser Module 2.1.4 for M36I mutant complexed to ritonavir

SPACegroup HALL P 61 #P 61

SOLU SET 1

RFZ=18.5 TFZ=30.7 PAK=0 LLG=1321 LLG=1962

SOLU 6DIM ENSE ensemble1

EULER 277.166 179.872 157.171

FRAC -0.00597 0.00112 & 0.33743

SOLU SET 2

RFZ=16.9 TFZ=35.5 PAK=0 LLG=1098 LLG=1743

SOLU 6DIM ENSE ensemble1

EULER 263.835 0.127 336.172

FRAC -0.00489 0.00109 & -0.33393

****

relative weighting between geometric and X-ray terms in the target function was determined automatically in *CNS*. Water molecules were added manually by examining unaccounted electron densities present in both *mFo-DFc* and *2mFo-DFc* maps. Composite omit maps were calculated by leaving out 3 % of the amino acid residues at a time. In the final stages of refinement, *Refmac* and *phenix.refine* implemented in *CCP4* [376] and *PHENIX* [375] respectively, were used with *TLS* [407, 408] refinement on the protein model. The entire model building and the structural superpositions were carried out using the softwares *O* [366] and *Coot* [367, 442] and figures were made using *Pymol* [415] and *Chimera* [416]. When comparing different complex structures, only protein Cα atoms were used in the calculation of the superposition matrices. The volume of the active site cavity was calculated using a spherical probe of radius 1.4 Å in the software *CASTp* [565]. The refinement statistics for the three structures are given in Table 19.

Table 19: Refinement statistics of the WT/RTV, V82F/RTV and M36I/RTV complexes.

Refinement Parameters	**WT:RTV (soaking)**	**WT:RTV (co-crystal)**	**V82F:RTV (co-crystal)**	**M36I:RTV (co-crystal)**
R_{work} (%) Outer resolution shell	18.7 (29.9)* (2.61-2.50 Å)	17.6 (29.3)* (1.64–1.60)	19.6 (24.2)* (1.95-1.90)	19.5 (22.5)* (1.64–1.60)
R_{free} (%) #	25.1 (35.1)* (2.61-2.50 Å)	21.0 (34.1)* (1.64–1.60)	21.9 (31.8)* (1.95-1.90)	21.9 (31.8)* (1.64–1.60)
No. protein atoms	1514	1515	1522	1516
No. solvent atoms	140	154	205	231
No. drug atoms	100	50	50	100
rmsd of bond lengths (Å)	0.012	0.016	0.014	0.016
rmsd of bond angles (°)	1.69	1.63	1.91	1.92
Average B factor (Å^2) Protein atoms Drug atoms Water atoms	 33.3 75.7 74.5	 37.6 39.7 23.3	 43.9 31.9 43.5	 19.0 15.2 17.4
Ramachandran plot favoured allowed	 192 (99%) 2 (2%)	 192 (99%) 2 (2%)	 192 (99%) 2 (2%)	 192 (99%) 2 (2%)

 # The R_{free} was calculated using 5% of reflections that were kept apart from the refinement during the whole process.

** The numbers between parentheses indicate the value in the outer resolution shell.*

6.3.1 Ritonavir bound in active site of V82F and M36I mutants of HIV-1 protease

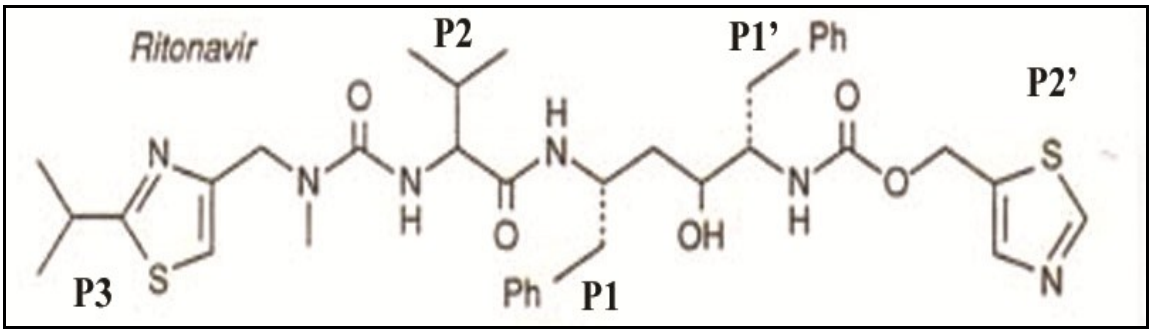
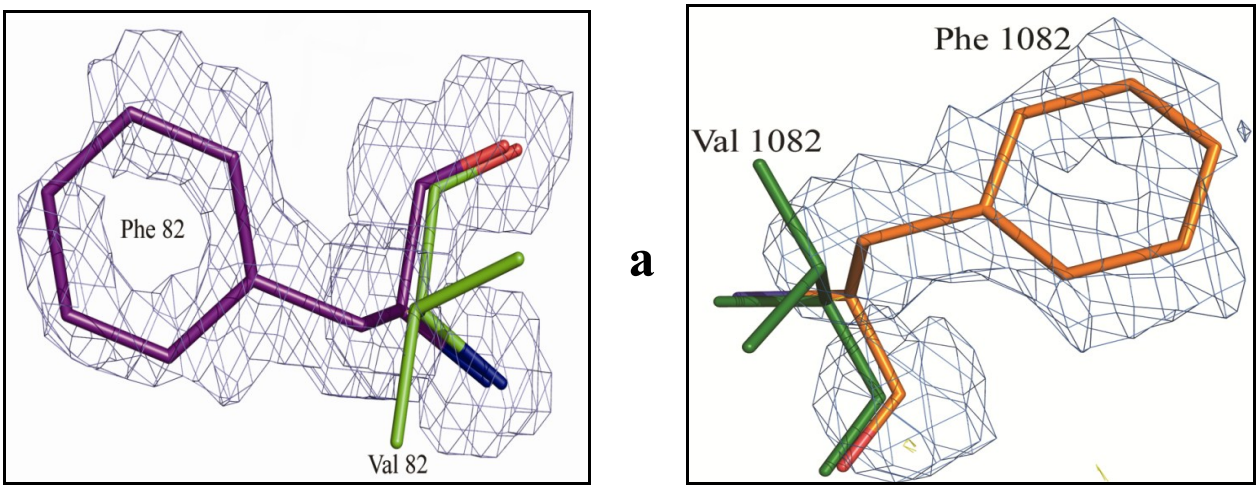


Figure 63: The chemical structure of HIV-1 protease inhibitor, ritonavir.

The chemical structure of ritonavir is shown in Figure 63. Figures 64a and 64b (left) show the simulated annealed *mFo-DFc* maps calculated around 82nd and 36th residues. The aromatic ring of F82 and aliphatic side chains of I36 residues fit well into the omit electron density maps, while V82 and M36 do not fit, confirming V82F and M36I mutations. Similar electron densities are also observed at positions 1082 and 1036 (Figures 64a and 64b, right).



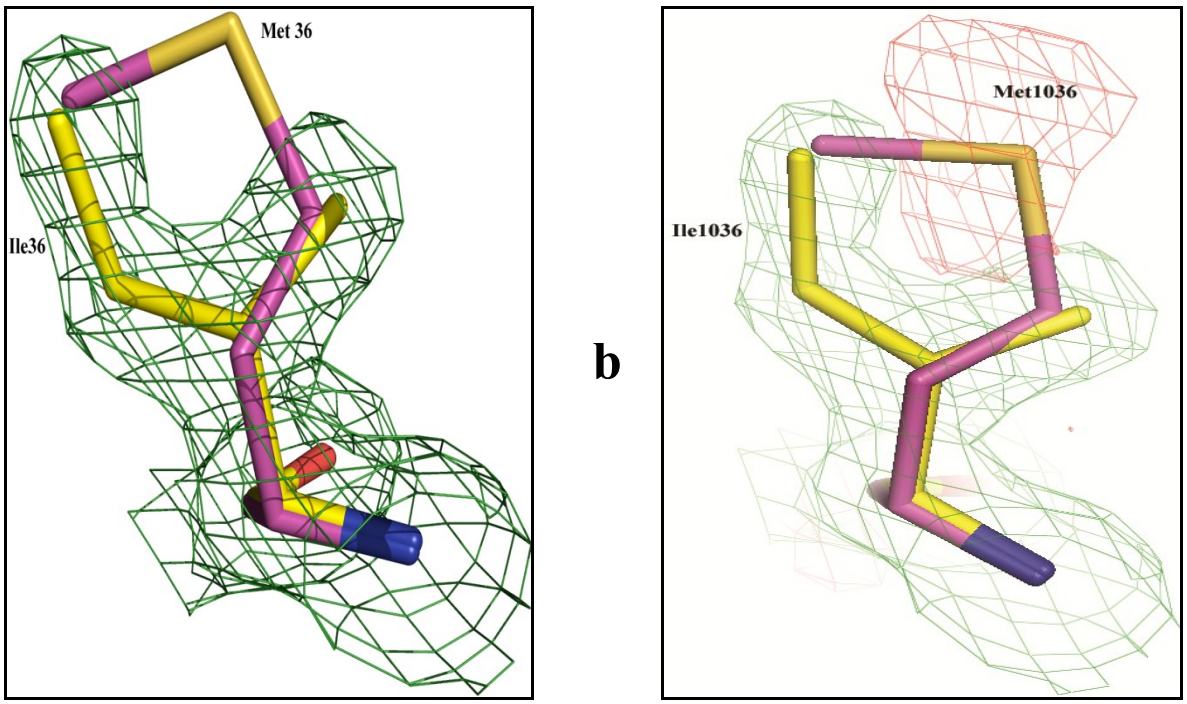
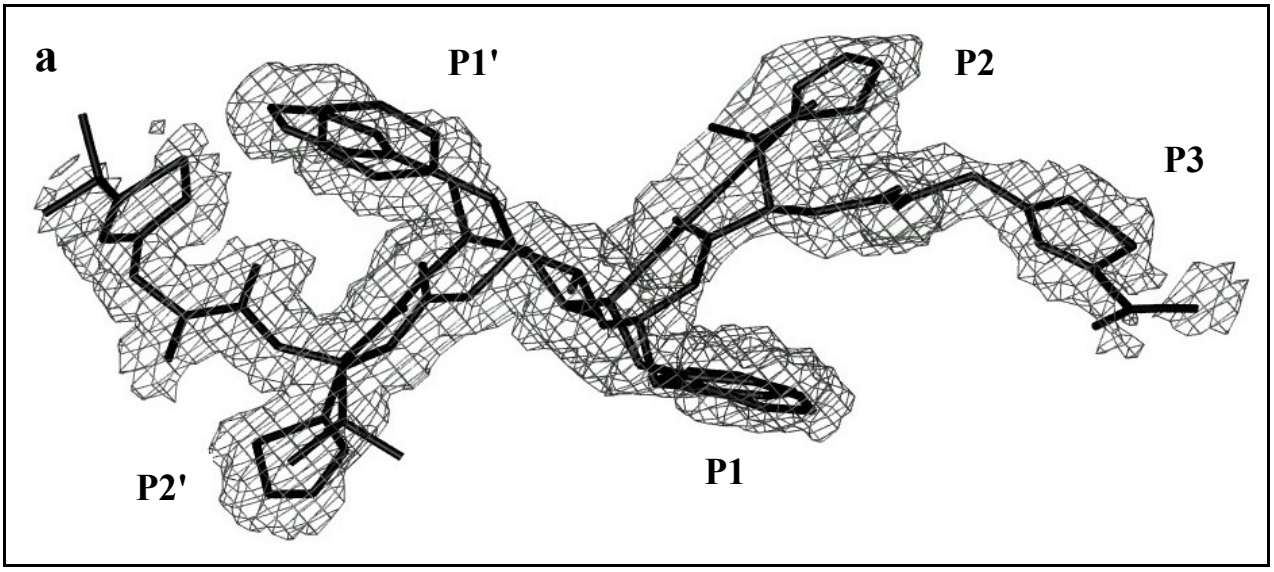
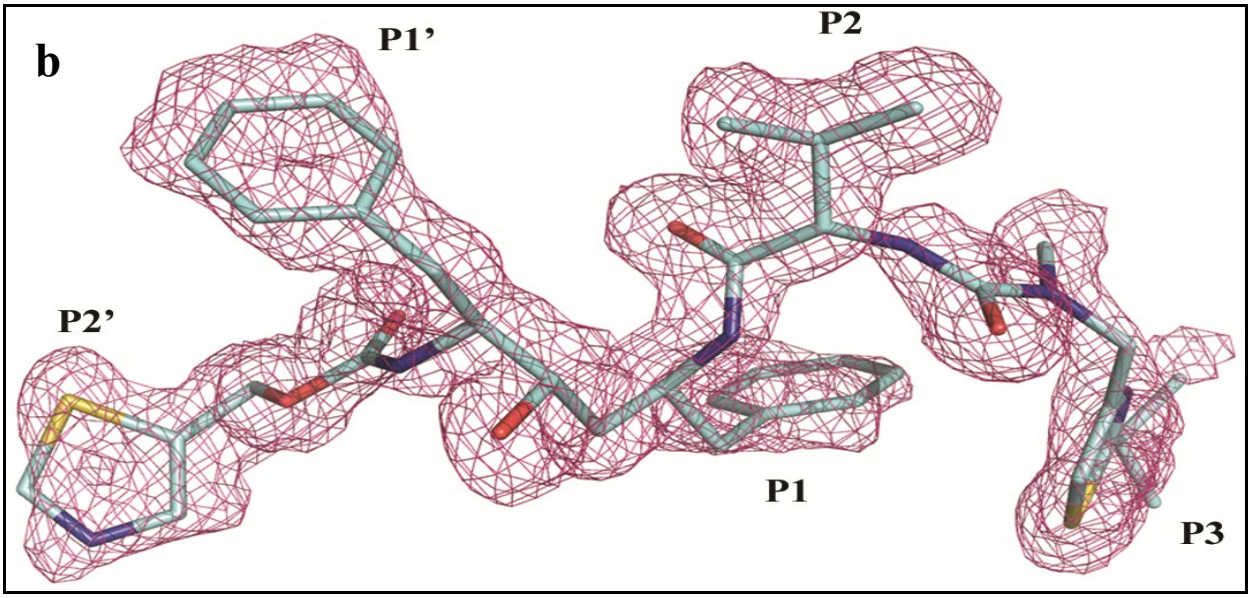


Figure 64: The F82/F1082 residues in V82F/ritonavir and I36/I1036 residues in M36I/ritonavir were not included in map calculations. **a)** Simulated annealed *mFo-DFc* map (blue) contoured at 2.5 σ showing F82 and F1082 residues fits well. The refined F82 (purple sticks) and F1082 residue (orange sticks) in V82F mutant and the V82 residue (green sticks) in WT HIV-1 protease models are superposed for comparison. **b)** Simulated annealed *mFo-DFc* map (green) contoured at 2.5 σ showing I36 and I1036 residues fitting well. The refined I36 and I1036 residues (yellow sticks) in M36I mutant and the M36 residue (magenta sticks) in WT HIV-1 protease models are superposed for comparison. The red *mFc-DFo* map indicates that M1036 residue is incorrect.

Connected positive maxima in the active site of the simulated annealed omit map indicated the presence of the inhibitor (Figures 65a-d). The extra *mFo-DFc* electron density obtained after modeling ritonavir in one orientation indicated a twofold symmetry related inhibitor in the active site, which itself is pseudo-symmetric. The drug, ritonavir binds in the active site cavity in one orientation in WT co-crystal and V82F complexes whereas in the WT soaked and M36I complexes it binds in two orientations. The omit maps in the active site regions for the WT/RTV soaked, WT/RTV co-crystal, V82F/RTV and M36I/RTV structures are shown in Figure 65a-d. The refined occupancies are 0.53 and 0.47 in WT soaked and 0.65 and 0.35 in M36I complexes.





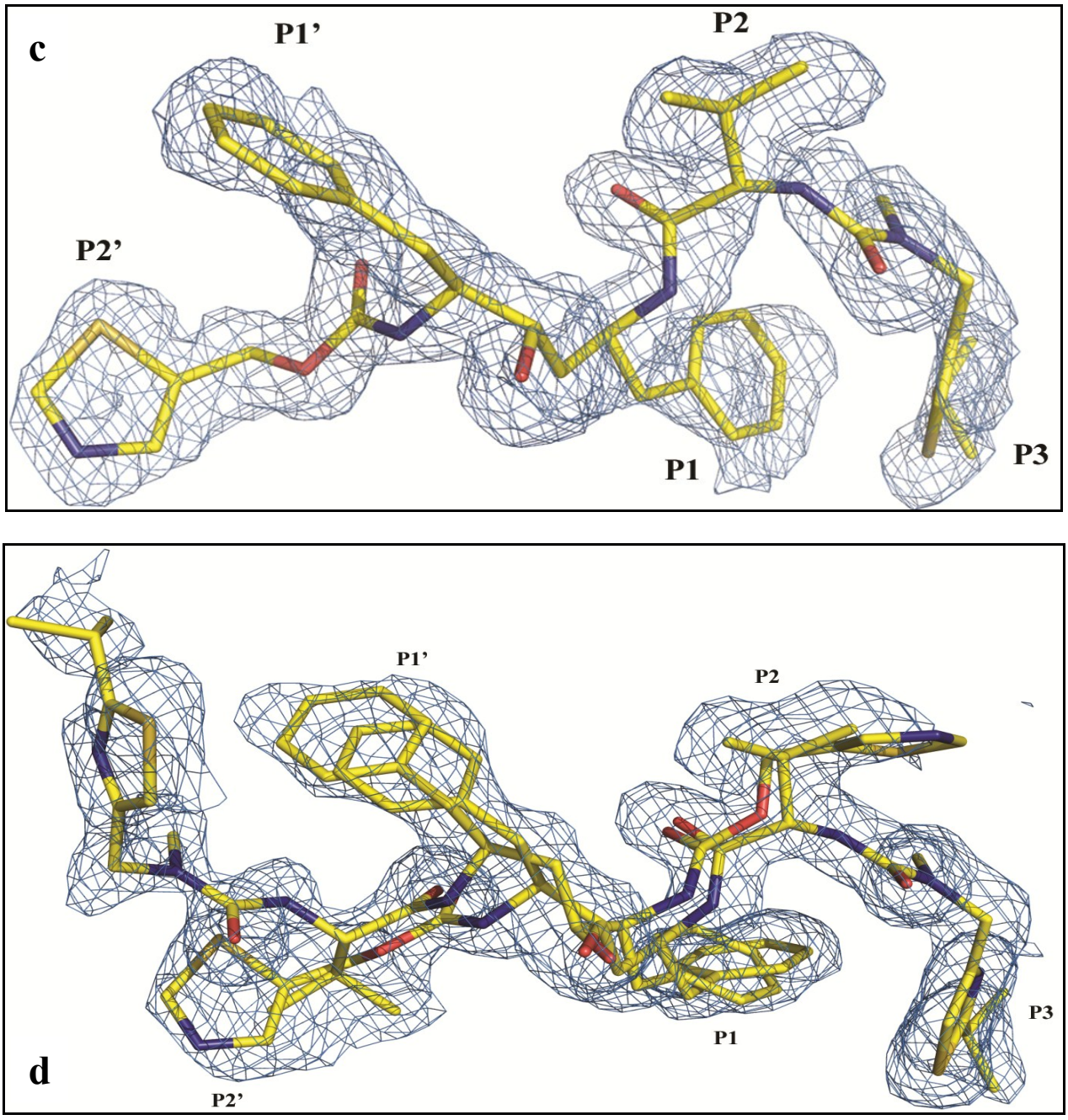


Figure 65: Simulated annealed *mFo-DFc* map, contoured at 2.2 σ, covering the refined coordinates of the inhibitor, ritonavir, bound in the active site cavity. Ritonavir atoms were not included in map calculations. **a)** Grey coloured map corresponds to ritonavir (black sticks) soaked in WT protease crystal. **b)** Red coloured map corresponds to ritonavir (cyan sticks) co-crystallized with WT protease. **c)** Light blue coloured map corresponds to ritonavir (yellow sticks) co-crystallized with V82F protease. **d)** Blue coloured map corresponds to the two orientations of ritonavir (yellow sticks) co-crystallized with M36I protease.

6.3.2 Comparison of the WT/ritonavir complexes obtained by soaking and co-crystallization methods

In the 2.5 Å WT/RTV structure obtained by soaking method, inhibitor is bound in the enzyme active site in two orientations related by two fold axis of symmetry about the protease dimer (Figure 65a). In contrast ritonavir binds in single orientation in the 1.6 Å co-crystallized WT/RTV structure (Figure 65b). The rmsd of the protein Cα superposition is 0.43 Å. The conformation of ritonavir is similar in both the structures. The electron density of the flap water is not clearly visible in the *2mFo-DFc* map

of the soaked crystal structure. At very low map contour level of 0.3 σ, a weak spherical density appears near the flap water position. Two pairs of conserved water molecules present in other HIV-1 protease structures, which hydrogen bond to carbonyls of G27 and G1027, are barely visible. The possible reason may be the low resolution and poor completeness of the data collected on the home source. The higher resolution electron density map obtained from synchrotron data in the co-crystal WT/RTV structure is better and inhibitor is bound in single orientation with full occupancy. The five conserved water molecules showing strong positive peaks are also clearly visible. Hence the co-crystal structure alone will be further used for comparison with the mutant complexes.

6.3.3 Comparison of the V82F/ritonavir and WT/ritonavir complexes

Position and interactions of ritonavir

The relative positions of ritonavir in the WT and V82F complexes are shown in Figure 66. Significant lateral shifts are observed in the positions of P1/P1' benzyl rings and the terminal isopropyl group at P3, while the P2 valine overlaps perfectly in the two structures. As a result, inhibitor-protein van der Waals contacts in the active site are significantly different in the two complexes. The pattern of hydrogen bonding interactions between the drug and active site residues, however, remains similar as shown in Figures 67 and 68 respectively. The polar hydrogen-bonding interactions are mainly in the S1/S1' and S2 binding pockets. The central secondary alcoholic hydroxyl (O41) of ritonavir, is within hydrogen bonding distances from the four carboxyl oxygen atoms of the two catalytic aspartates. The carbonyl oxygens in the P1'/P2 sites interact with flap residues through the flap-water molecule, which is conserved in all inhibitor complexes of HIV-1 protease (Figures 67 and 68). The hydrogen bonds from the flap water to the carbonyl oxygens are of unequal length. In the S2 pocket, there are two NH --- O type direct hydrogen bonds between the backbone atoms of ritonavir and residues forming the floor of the active site cavity, with distances ranging from 2.9 Å to 3.2 Å. There is only one weak hydrogen bond (3.2 Å) at the roof of active site cavity, between G1048 carbonyl oxygen and the nitrogen of the P2 valine. Two water-mediated hydrogen bonds in the S2' pocket link the sulphur atom of the thiazolyl ring to the protein.

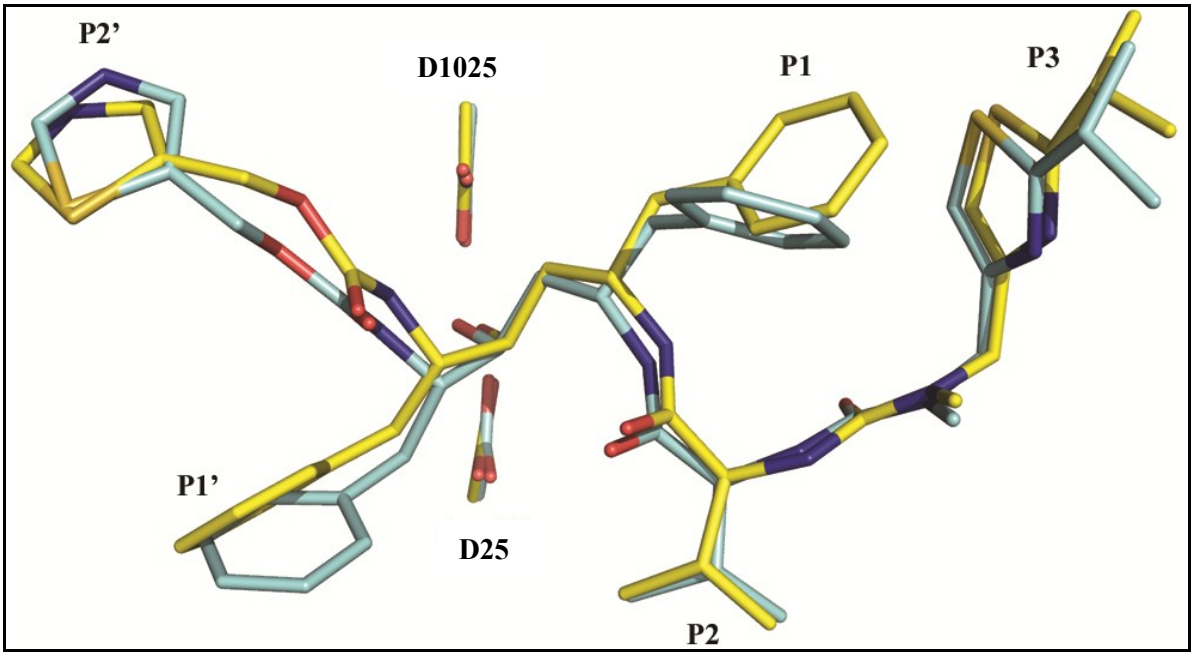


Figure 66: The inhibitor, ritonavir, is shown within the active site of WT (blue sticks) and V82F (yellow sticks) complexes after superposition. Note the difference in orientations of the benzyl residues of ritonavir at the S1/S1' pockets. There is a slight shift perpendicular to the plane of the isopropyl thiazolyl ring in the S3 pocket and slight rotation in thiazolyl ring in the S2' pocket. Note that the catalytic D25 and D1025 residues of the protein are perfectly superposed.

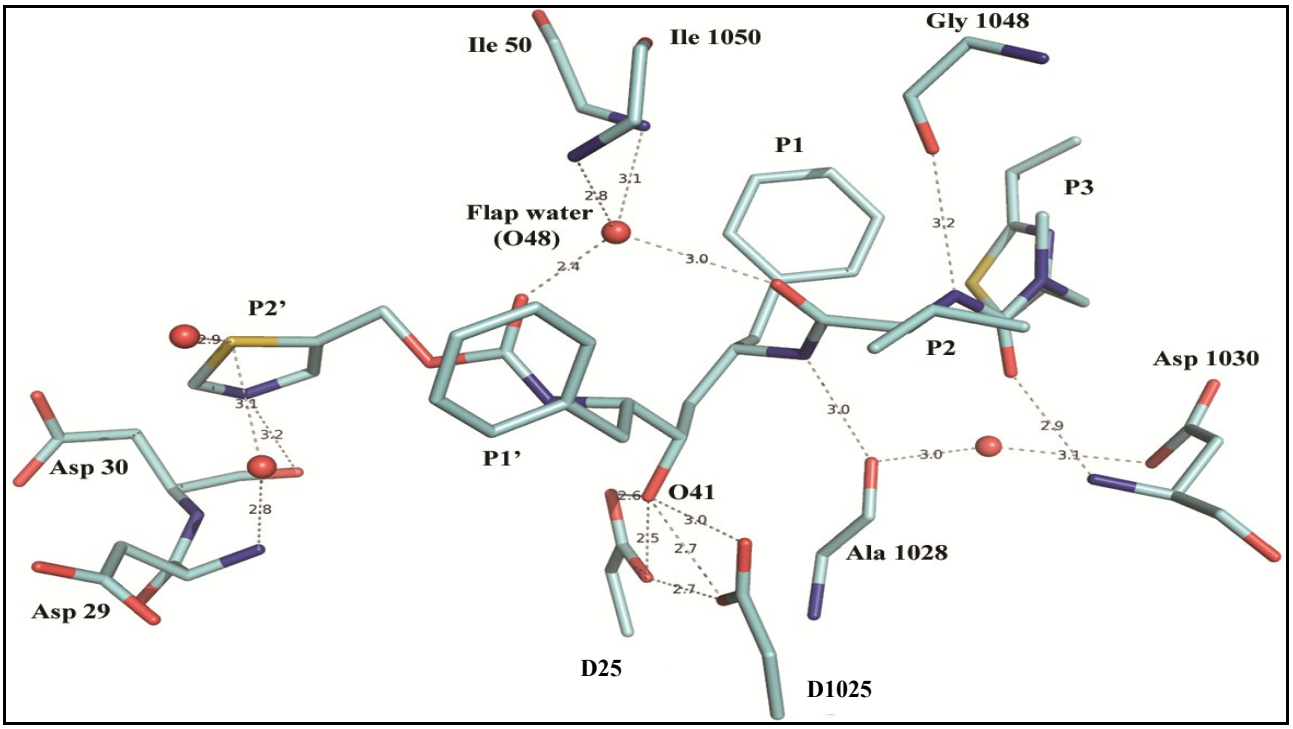


Figure 67: Hydrogen bonding involving ritonavir in the active site of WT protease. Ritonavir central hydroxyl (O41) is placed asymmetrically with respect to the D25 and D1025.

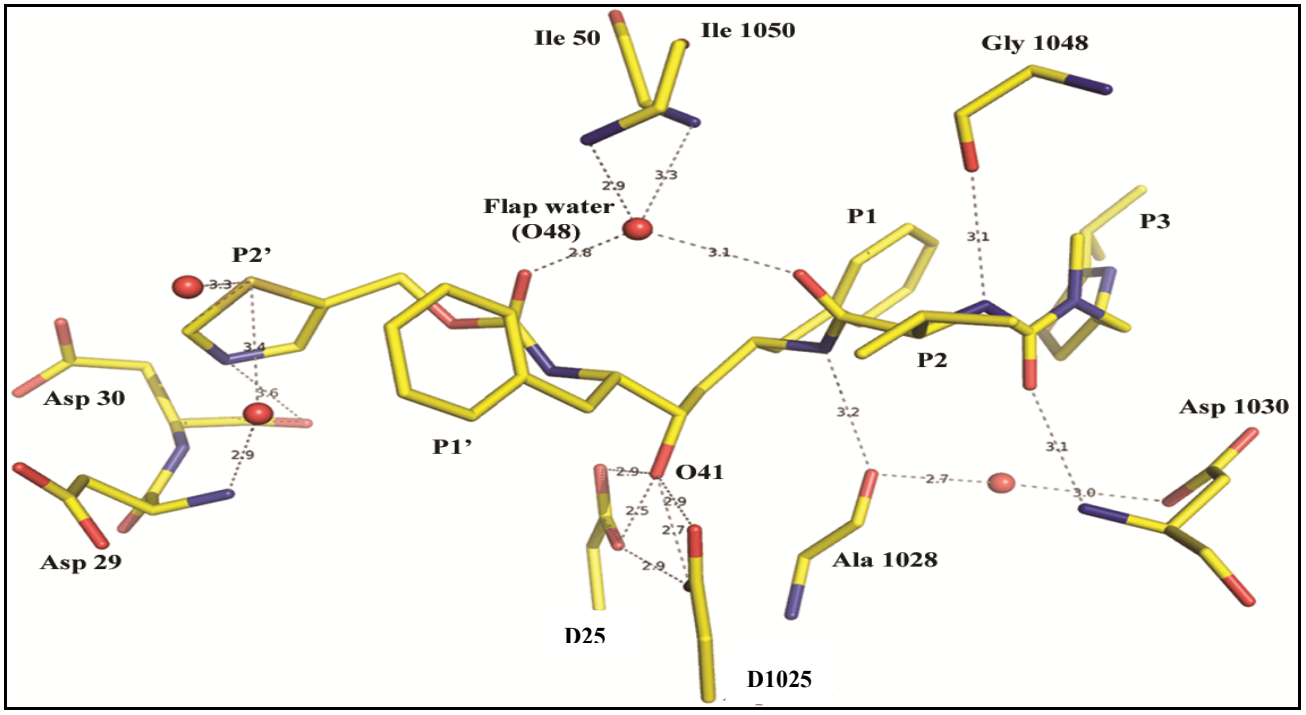


Figure 68: Hydrogen bonding involving ritonavir in the active site of V82F mutant protease. Ritonavir central hydroxyl (O41) is placed symmetrically with respect to the D25 and D1025.

Conformation of the protein

The protein molecules in the two complexes superpose with an rmsd of 0.35Å for 198 Cα atom pairs (Figure 69). Larger differences are in the flexible loop regions and also in the conformations of few residues in the active site. The Cα atom of the mutation residue 82 is displaced by 0.6 Å. While in the WT, the Cγ1 and Cγ2 atoms of V82 point into the active site, in the V82F molecule, the phenyl ring is pointing away from the active site cavity. Another interesting difference is in the conformation of the residues P81/P1081. The Cδ atom of the proline rings in V82F structure is reaching out for non-bonded interactions with P1/P1' benzyl rings, while in the WT the Cδ atoms are too far away for effective van der Waals interactions with the benzyl rings. The *CASTp* analysis shows that the active site cavity is significantly larger in V82F (1322.7 Å^3) than in the WT HIV-1 protease (1225.5 Å^3). The larger active site volume is also reflected in larger distances between residues flanking the active site. For example, the Cα-Cα (V82-V1082) longitudinal distance is 19.56 Å in the WT complex, and 20.24 Å for the V82F complex. The Cα-Cα (I50-P1081) transverse distance is 6.66 Å in the WT complex, while for the V82F complex it is 7.27 Å, indicating that the S2 pocket has expanded in the V82F structure. Interestingly, there is no similar increase in the size of S2' pocket, as the other transverse distance, Cα-Cα (I1050-P81), is similar: 6.83 Å for WT and 6.62 Å for V82F. Thus while S2/S2' pockets are of nearly identical size in the WT complex, in the V82F complex, there is an

asymmetry. The flap conformation is also slightly different in the two complexes. There are six salt bridges between protein residues in the wild type-ritonavir complex as against five in the V82F mutant complex, the salt-bridge E35:R57 missing in the latter.

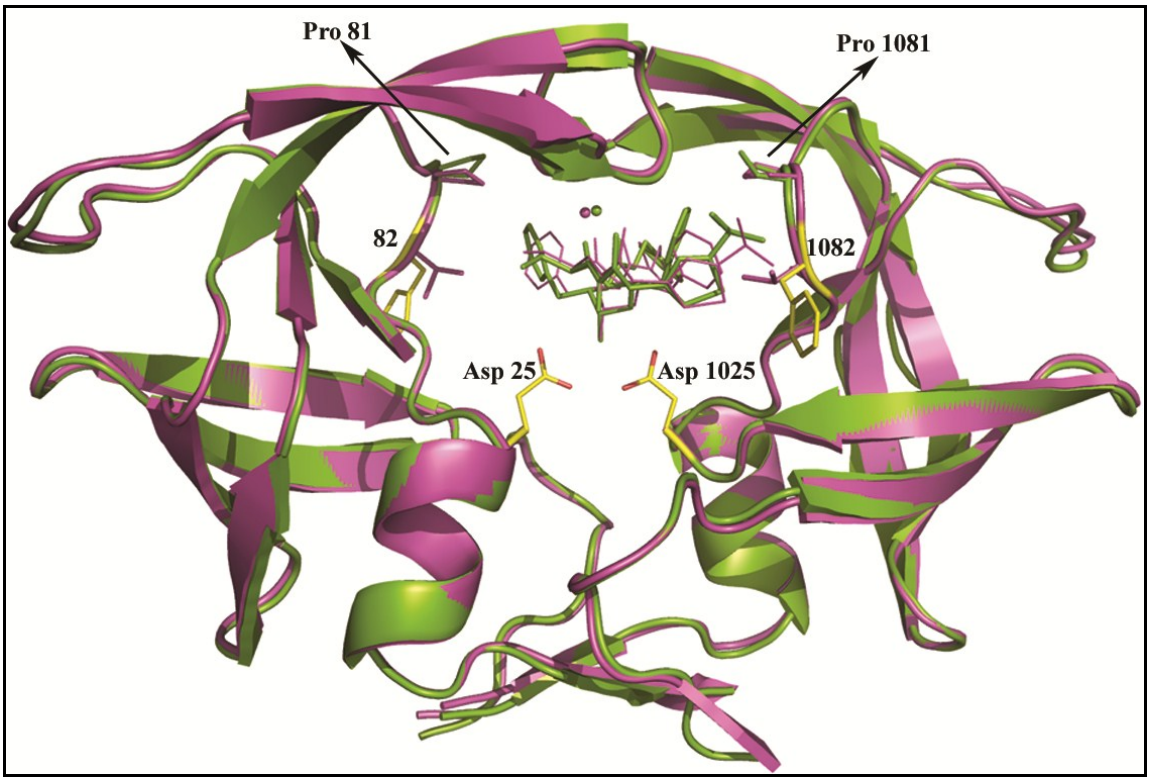


Figure 69: Structural superposition of V82F/ritonavir (magenta) and the WT/ritonavir (green) complexes. Cartoon diagram of the HIV-1 protease showing the F82 side chain (yellow stick) is pointing away from the active site cavity. The V82 (magenta stick) of the WT complex is pointing into the active site cavity. P81 and P1081 also have different conformations in the two complexes.

6.3.4 Comparison of the M36I/ritonavir and WT/ritonavir complexes

Position and interactions of ritonavir

The rmsd between the protein Cα atoms in the WT and M36I complex structures is 0.33Å and the relative positions of ritonavir in the two complexes are shown in Figure 70. Lateral shifts are observed in the positions of the terminal isopropyl group in thiazolyl ring at P3 and thiazolyl ring at P2', while the P2 valine overlaps perfectly in the two structures (Figure 70). The inhibitor-protein van der Waals contacts (using a distance cutoff of 4 Å) in the active site are similar in the two complexes. The pattern of hydrogen bonding interactions between the drug and active site residues in the WT and M36I complexes are shown in Figures 67 and 71 respectively. The central secondary alcoholic hydroxyl (O41) of ritonavir, is within hydrogen bonding distances from the four carboxyl oxygen atoms of the two catalytic aspartates. The carbonyl oxygens in the P1'/P2 sites interact with flap residues through the flap-water molecule, which is conserved in all inhibitor complexes of HIV-1

protease (Figures 67 and 71). The hydrogen bonds from the flap water to the carbonyl oxygens are almost of equal length. In the S2 pocket, there are two NH --- O type direct hydrogen bonds between the backbone atoms of ritonavir and residues forming the floor of the active site cavity, with distances ranging from 2.9 Å to 3.2 Å. There is only one weak hydrogen bond (3.1 Å) at the roof of active site cavity, between G1048 carbonyl oxygen and the nitrogen of the P2 valine. In the primed side of M36I, there is a hydrogen bond (2.9 Å) between the G48 carbonyl oxygen and sulphur of the thiazolyl ring but it is absent in the WT complex due to the lateral shift mentioned earlier (Figures 67 and 71). The nitrogen of the thiazolyl ring forms two hydrogen bond with the OD2 and nitrogen of D29 instead of the water mediated ones observed in WT complex. The central hydroxyl (O41) of ritonavir forms very symmetric hydrogen bonds with the catalytic aspartates in the M36I complex.

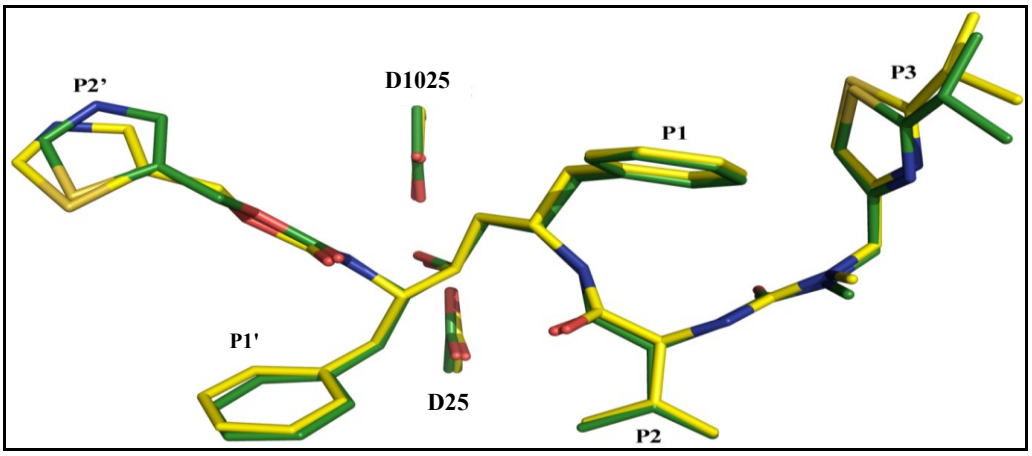


Figure 70: The inhibitor, ritonavir, is shown within the active site of WT (green sticks) and M36I (yellow sticks) complexes after superposition. Note there is very little difference in orientations of the benzyl residues of ritonavir at the S1/S1' pockets. Note that the catalytic D25 and D1025 residues of the protein are perfectly superposed.

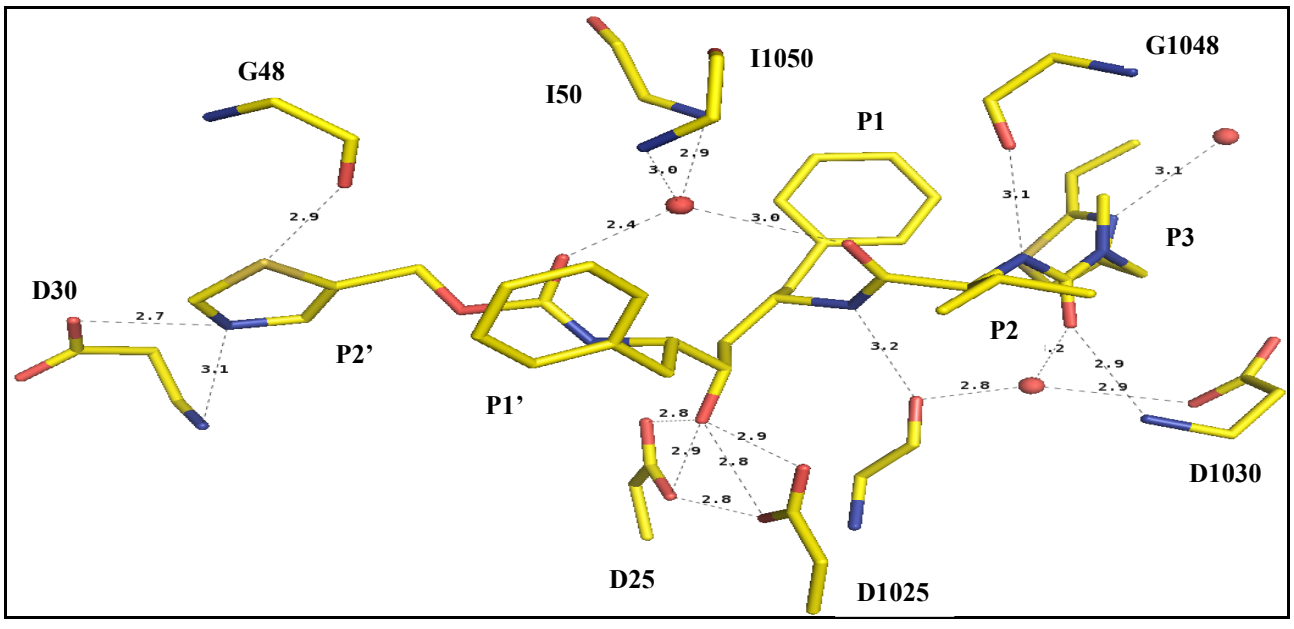


Figure 71: Hydrogen bonding involving ritonavir in the active site of M36I mutant protease. Ritonavir central hydroxyl (O41) is placed symmetrically with respect to the D25 and D1025.

Conformation of the protein

The protein molecules in the two complexes superpose with an rmsd of 0.33 Å for 198 Cα atom pairs (Figure 72). Larger differences are in the flexible loop regions and also in the conformations of I50 and I84 residues facing the active site. The active site cavity region superposed well except in the 80s loop. The Cα atom at the mutation site of residue 36 and 1036 is displaced by 0.72 Å and 0.85 Å respectively. The flap conformations are similar in the two complexes (Figure 72). The concomitant backbone shifts in the 80s loop, especially at P81 is significant. There are six salt bridges present between protein residues in both WT and in M36I mutant complexes. The conformations of K20 and E35 residues are different.

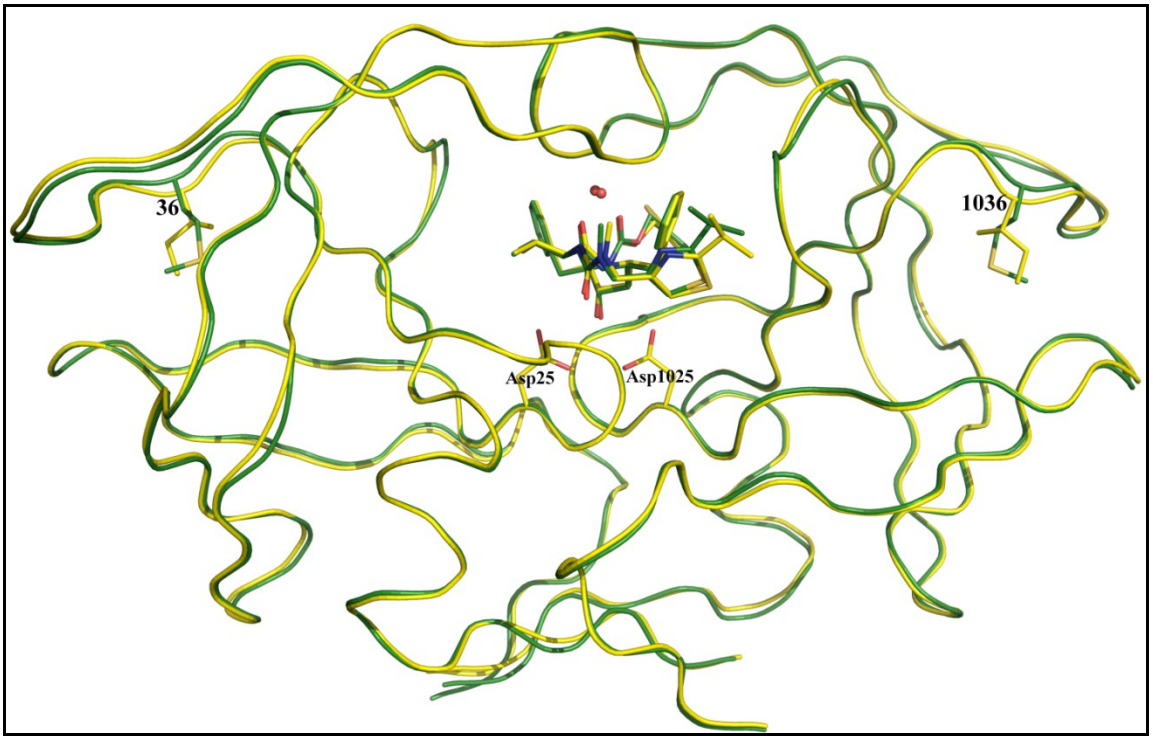


Figure 72. Structural superposition of M36I/RTV (yellow ribbon) and the WT/RTV (green ribbon) complexes. The shifts in the backbone near the 36/1036th and 81/1081st residues may be noticed.

6.3.5 Conformation of ritonavir in the mutant complexes

Ritonavir in the WT and V82F mutant structures has been superposed and its conformation is shown in Figure 66. The P1 phenyl ring in the V82F mutant complex is rotated by more than 60° around the Cα-Cβ bond, when compared to the WT complex. In the altered position, the phenyl ring is in contact with the Cβ group of F82. The phenyl ring is also pointing towards the π-electron cloud of the 5-membered thiazolyl ring in the P2' position. Similar superposition of ritonavir in the WT and M36I mutant structures is shown in Figure 70. There is a slight shift in the plane of the isopropyl thiazolyl

ring in the S3 pocket and slight rotation in thiazolyl ring in the S2' pocket in the M36I and WT complexes.

6.3.6 Flexibility of 82nd phenylalanine

The mutation V82F is elicited as a drug resistance mutation against a variety of clinical inhibitors in addition to ritonavir, and it is also part of the multi-drug-resistant HIV-1 protease [185]. *In vitro*, the mutant protease is insensitive to the C2-symmetric inhibitor DMP323 [566]. Structural data on clinical inhibitor complexes with single mutant V82F HIV-1 protease are needed to dissect out structural effects and cooperative interactions [544] of different mutations in multi-drug resistant enzymes. The present structure is the first report of V82F single mutant complexed with any clinical inhibitor. Since ritonavir has C2-symmetric backbone, we have compared the present structure with other complexes containing C2-symmetric inhibitors. Three-dimensional structure of V82F mutant complexed with DMP323 and XV638, have been determined by X-ray crystallography [566, 567]. The present structure overlaps very well with these two complexes, with rmsd values of 0.35 Å and 0.37 Å respectively over protein Cα atoms. The side chain of the mutation residue, F82, is positioned outside the active site in all the three structures. However, positions of the Cα atoms and orientation of the phenyl rings do not exactly match, and the displacement averaged over all phenyl ring atoms ranges from 4.1 Å (in 2nd subunit 2.7 Å) in DMP323 complex to 3.4 Å (in 2nd subunit 2.9 Å) in XV638 complex. The side chain of F82 takes a completely different position in V82F/I84V complexed to DMP323 as shown in Figure 73. Similarly, in the crystal structure of V82A complexed with another C2-symmetric compound, A77003, the main chain atoms in the 80's loop have repositioned to achieve optimal contacts with phenyl ring of the inhibitor molecule [564]. In the crystal structure of V82F/atazanavir complex, F82 side-chain is oriented inside the active site and is engaged in stacking interactions with P1/P1' benzyl ring from the drug [568]. Interestingly, in the structure of unliganded multi-drug-resistant HIV-1 protease, F82 side chain is oriented outside the active site cavity [569]. These results indicate that the main chain and side chain conformation of F82 is flexible and is influenced by the inhibitor molecule in the active site.

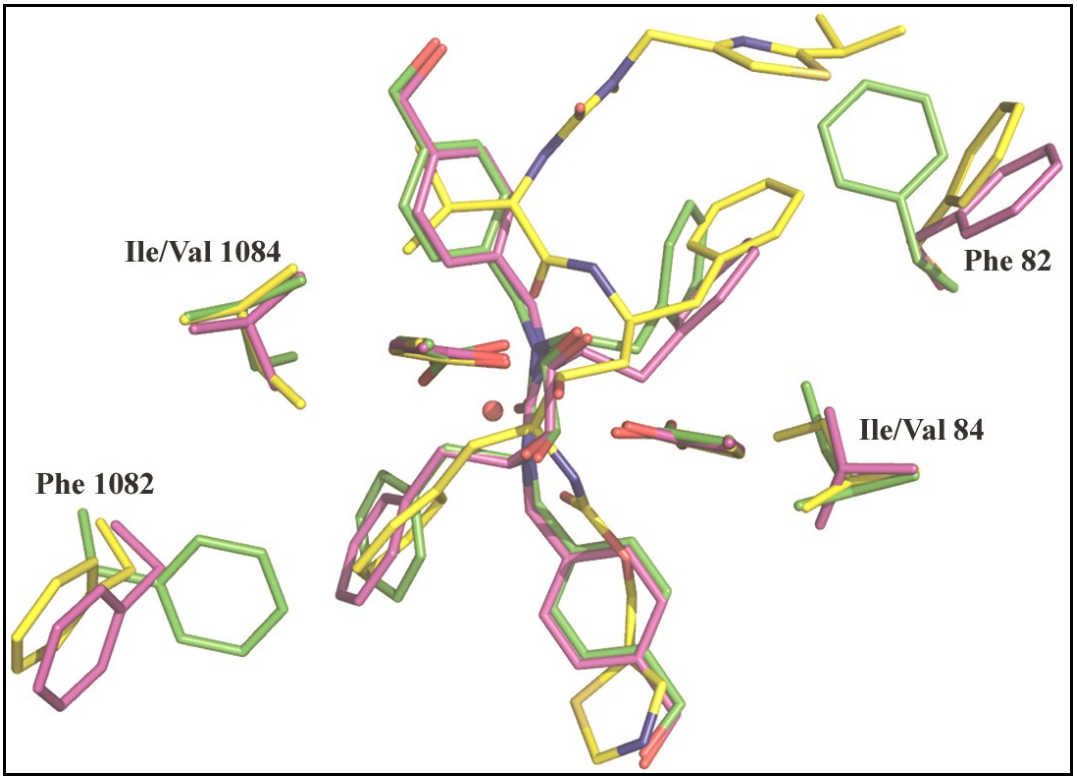


Figure 73: Superposed stick models of V82F/ritonavir (yellow), V82F/DMP323 (magenta) and V82F/I84V/DMP323 (green) complexes. It may be noted that V84 of double mutant is superposed well with I84 of V82F single mutants whereas the F82 exhibits more flexibility.

6.3.7 The mechanism of drug resistance due to the V82F active site mutation

The mechanism of drug resistance due to V82F mutation must be direct since residues 82/1082 occur in the S1'/S1 pockets involved in the binding of ritonavir. It has been reported that ritonavir/HIV-1 protease binding energy is determined mostly by enthalpy, with the major part coming from hydrophobic residues [570]. Aliphatic CH-π interactions, which are known to play key roles in ligand binding [571, 572], are present in the WT complex, between the γ-methyl groups of V82/V1082 and the π-electron cloud of P1'/P1 benzyl rings of the drug (Figure 74, top) [571-573]. In the V82F mutant, corresponding methyl groups are absent because of the change in amino acid type (Figure 74, bottom). Therefore CH-π interactions are lost in the mutant complex. However, the protein and the inhibitor molecules have readjusted their conformations in an effort to make up for this loss by regaining alternate interactions. From the inhibitor side, the P1 benzyl ring has rotated almost by 60^{o} so that CH-π interactions could be made with Cβ hydrogen atoms of F1082. From the protein side, the P81/P1081 residues adopt different conformations to be within van der Waals distances from the P1'/P1 benzyl rings of the drug. Despite these attempts, van der Waals interactions between the inhibitor and the protein in the active site are sub-optimal as may be seen in Figure 74. This fact is also confirmed by the increase of active site cavity by about 100 Å^3 in the V82F complex. Thus there

is an overall reduction in the non-bonded interaction enthalpy, and this, we suggest, is the reason for the ritonavir-resistance of V82F mutant. In support of this hypothesis is the observation that V82F is not a drug-resistant mutation against atazanavir, because in the V82F/atazanavir complex there are good stacking interactions between F82 and a planar ring from the inhibitor molecule [568]. Loss of van der Waals interaction energy is suggested to be the reason for drug-resistance also in V82A-I84V mutant, where again the active site volume has increased by about 30 $Å^3$ compared to the complex with the WT enzyme. The larger increase in the active site cavity in the V82F complex compared to V82A complex implies a greater loss of affinity in the former, and yet V82A is the more prevalent drug-resistant mutation against ritonavir. This observation suggests that factors additional to simple reduction in affinity may play a role in the emergence of drug-resistance mutations. The increase in active site volume is also responsible for the weakening of the interaspartate (D25 OD1 --- D1025 OD1) hydrogen bond (2.90 Å vs 2.68 Å) in V82F mutant in comparison to the WT complex.

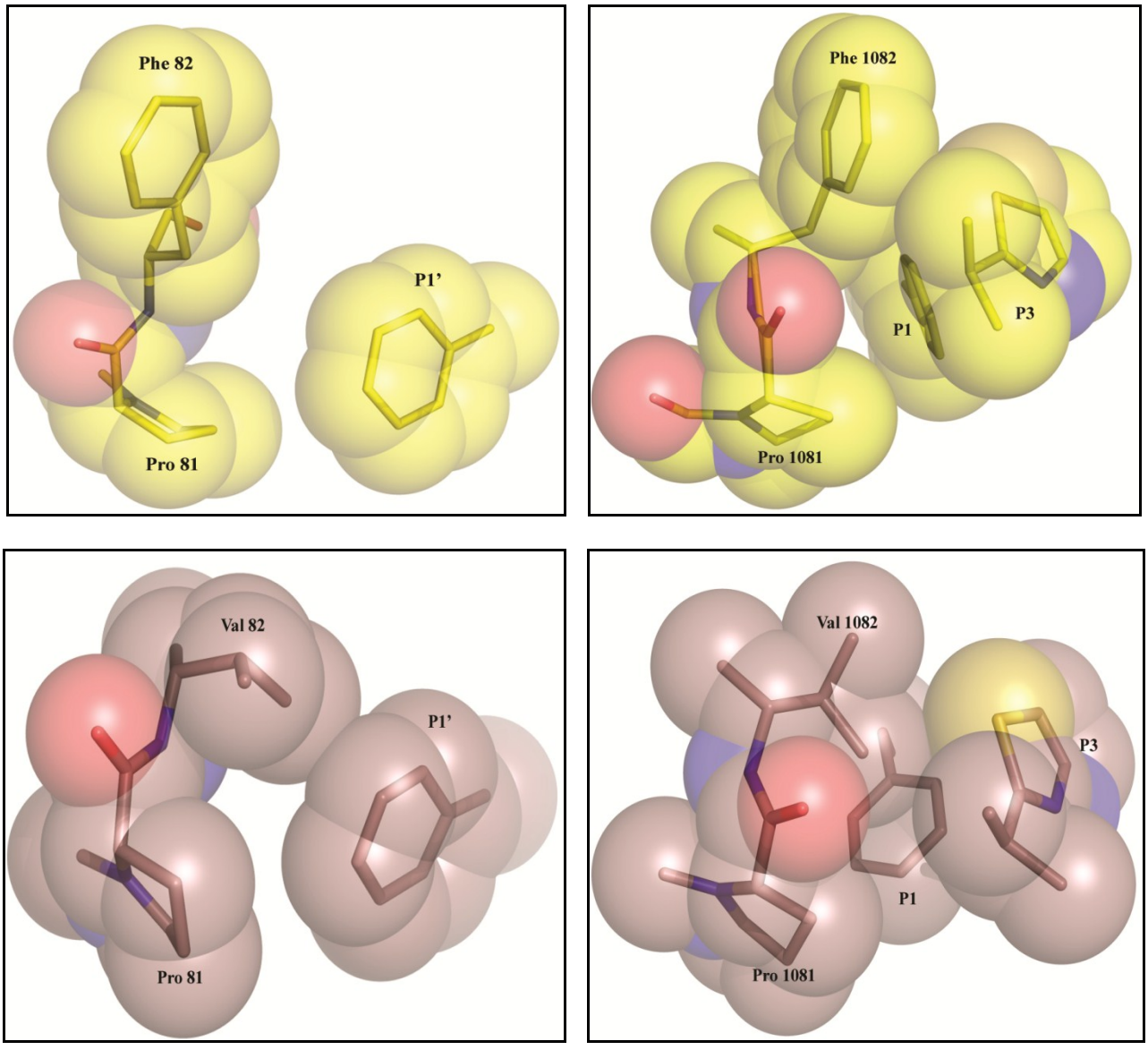


Figure 74: Space filling diagrams of the atoms showing interactions near the 82nd residue in active site cavity of HIV-1 protease. **Top (left and right)**: In the V82F complex, CH-π interactions are disrupted as the F82/F1082 are pointing away from the P1/P1' phenyl rings of ritonavir. **Bottom (left and right)**: In the WT complex, CH-π interactions are present between γ-methyl groups of V82/V1081 residues with the P1/P1' phenyl rings of ritonavir.

6.3.8 Additional mutations associated with V82F

Upon administration of ritonavir in monotherapy or in HAART cocktail, primary mutations appear at position 82, as either a single substitution or as part of a complex mutational pattern [185, 550]. The double mutant V82F-I84V is often found in patients undergoing ritonavir monotherapy or HAART. However, the exact role of each mutation in conferring ritonavir-resistance is not clear. Since three dimensional structure of V82F-I84V/RTV complex is not yet available, the current V82F mutant structure was used as a template to create a molecular model for V82F-I84V double mutant. The side chains at residues 84 and 1084 were replaced with valine, and the most preferred rotamer was selected. Ritonavir was docked into the active-site. Into the generated three-dimensional grid covering the active-site, flaps and 80s' loops of V82F-I84V double mutant, ritonavir was docked using the software package Autodock Vina [574]. The default torsion angles for the rotatable bonds of ritonavir were used in docking. This molecular model was subsequently energy minimized using CNS [43]. When compared with present V82F/RTV complex, only minor variations were found, and that too only in the conformation of the inhibitor. In particular, the backbone near the 84/1084 positions are unchanged (Figure 75), which is perhaps to be expected, since a well defined short helix starts immediately after 84th residue. Inflexibility of the residues 84/1084 may also be inferred from perfect overlap observed at these positions in the three complexes shown in Figure 73.

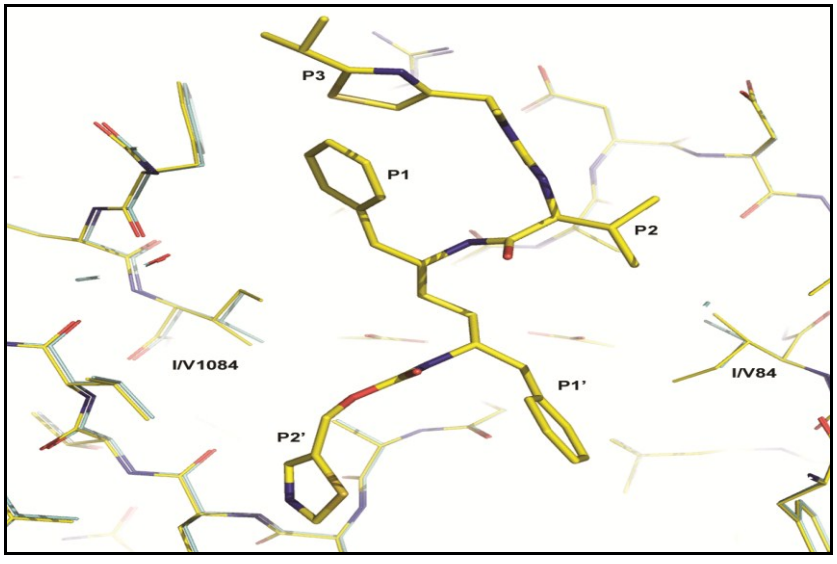


Figure 75: Superposition of docked V82F/I84V/ritonavir complex model (green sticks) with the V82F/ritonavir (yellow sticks) complex. The drug molecule in V82F mutant is not shown for clarity.

The P1/P1' benzyl rings of ritonavir are slightly rotated compared to V82F complex, and are now closer to the methyl groups of V84. The loss of contacts from Cδ atom of I84 is made up by new contacts with methyl groups of V84. Thus from the docking studies it is seen that in the double mutant structure, there is no major change in the interactions between the drug and the enzyme. Therefore, determinant of drug resistance is the V82F mutation. The mutation I84V may enhance catalytic efficiency of the mutant enzyme. This inference is consistent with the observation that I84V mutation emerged *in vivo* late and less frequently [185].

After the appearance of the primary mutation several stepwise secondary mutations appear, most frequently at positions 10, 36, 54 and 71 [575]. These secondary mutations could have a role either in compensation of the catalytic efficiency of the primary mutant enzyme or in further reduction in ritonavir binding. The contribution of different mutations to resistance in a multi-drug-resistant protease has been investigated through micro-calorimetric and enzyme kinetic measurements. These studies have led to discovery of cooperative coupling between distal mutations, with maximum cooperativity effects for mutations in the dimerization region of HIV-1 protease [544]. A similar coupling between residues 82 and 10 is suggested by the present work. The F82 side chain packs itself against the patch of hydrophobic residues L10, L23 and I84 (Figure 73). Compared to the WT complex, in V82F complex, the Cα82-Cα10 separation is increased through shifts in Cα atoms of both 82 and 10. Further, the L10 side chain adopts a different rotameric conformation to relieve steric clash between F82 and Cδ1 atom of WT L10. As a result the contact between F82 and L10 is almost edge-on to Cδ1 and Cδ2 methyl groups of L10 (Figure 76). The secondary mutation, L10I, associated with V82F is located in the dimerization region, and the region around residue 10 of the energy-minimized model for V82F/L10I mutant, built by using V82F complex structure as the template, is shown in Figure 77. It is clear that the contacts between residue 10 and residues L23 and F82 are longer in L10I mutant, leading to enhanced scope for movement, especially of residue L23 (Figure 77). Since L23 is at the floor in the active site, a more flexible L23 is likely to influence both inhibitor and substrate binding. A flexible L23 may easily accommodate changes of conformation of substrates in the S1/S1' pockets, needed for proper substrate binding and/or release, thus leading to enhanced catalysis. Correlation between L10I mutation and conformation of L23 has also been shown recently

from molecular modeling studies [576]. The present work suggests that the mutation L10I may be acting indirectly on substrate and inhibitor binding, by influencing conformation of L23 in the active site.

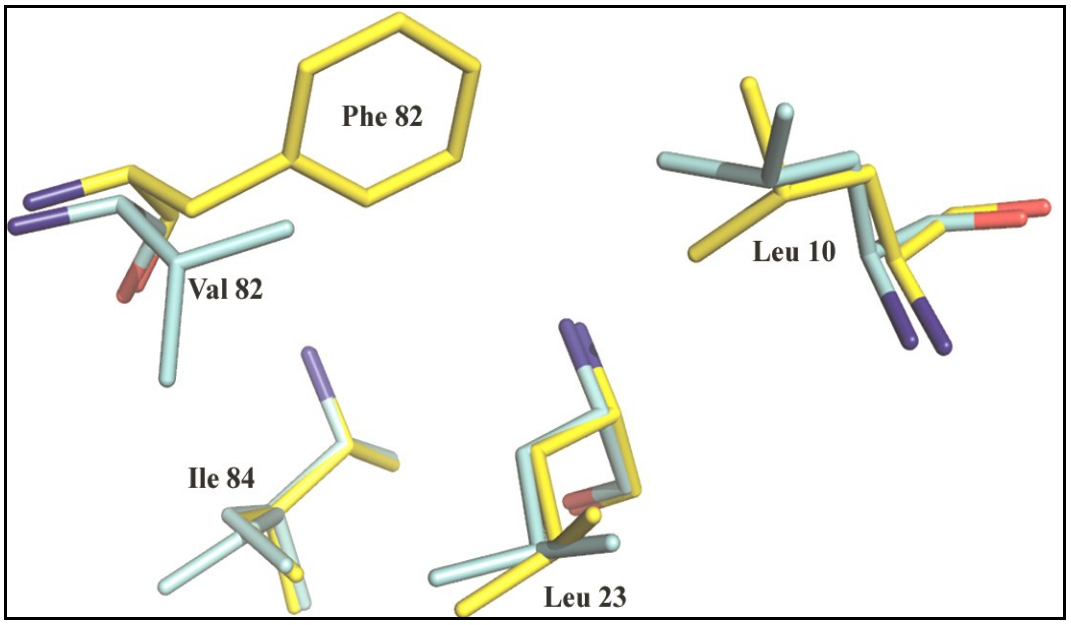


Figure 76: The interactions of the three hydrophobic residues I84, L23 and L10 with the 82nd residue of HIV-1 protease are shown after superposition of the WT (blue sticks) and V82F (yellow sticks) complexes. Note the shifts in the backbone and different rotameric conformations of L10.

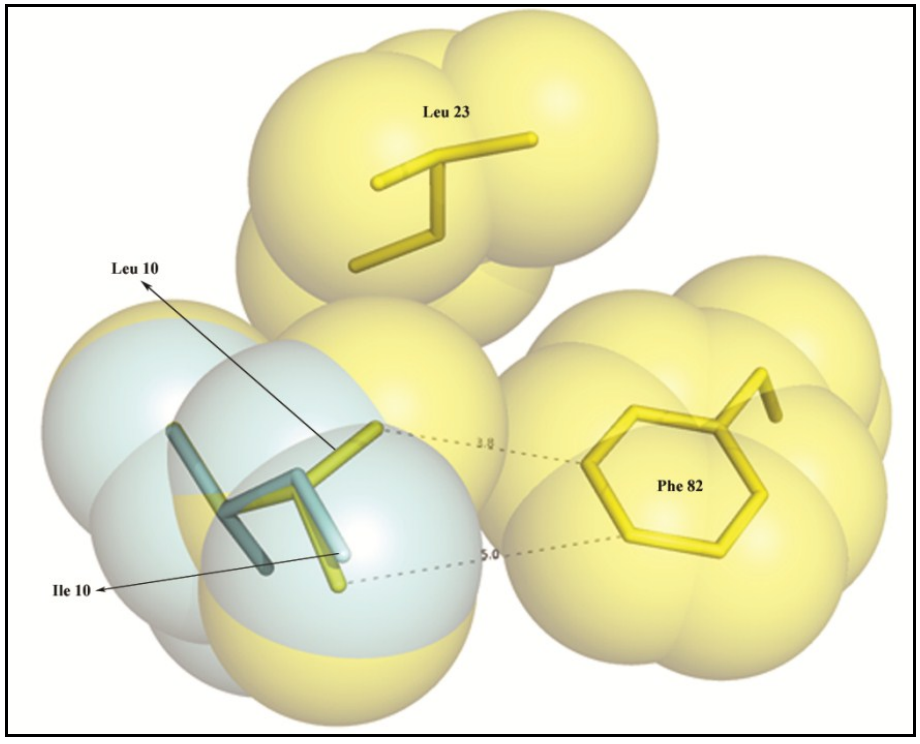


Figure 77: The space filling model of V82F/L10I double mutant showing reduced interactions with F82 and L23 when L10 (green sticks) is mutated to I10 (cyan sticks).

6.3.9 Role of the non-active site M36I mutation

Ritonavir drug resistance acquired through mutation at V82, followed by M36I and other mutations, viz. I54V, A71V and K20R, probably to restore viral fitness to even better than WT values [577]. The

M36I mutation is located in one of the hinge regions of HIV-1 protease, and is a non-active site mutation. In a molecular dynamics study it was proposed that M36I regulates the size of the binding cavity of the unliganded protease [578]. The *CASTp* analysis shows the active site cavity volume of M36I/RTV structure is 1123.5 $Å^3$ as compared to 1322.7 $Å^3$ in V82F/RTV and 1225.5 $Å^3$ in WT/ RTV structures presented earlier. Thus the excess increase in active site volume in V82F/RTV would be compensated in the V82F-M36I/RTV double mutant, indicating a compensatory role of M36I mutation appearing after V82F mutation [575]. In the V82F/RTV structure, E35:R57 salt bridge interactions are absent and only one salt bridge interaction in E1035:R1057 is present. This is because R57/R1057, situated at the end of flaps, shifts outward to interact with W42/W1042. In contrast, in M36I complex, there are two strong salt bridges between E35:R57 and E1035:R1057, just as in WT enzyme (Figure 78). Salt bridges typically contribute energies of 3-4 kcal/mol and their contribution increases (as high as 8 kcal/mol) with increase in a hydrophobic environment. Since the coupling between 82nd residue in the active site and 36th residue in the non-active site is indirect, these electrostatic interactions might be important and are restored in the V82F- M36I/RTV double mutant. These inter-subunit interactions are important for the restoration of the fitness of the dimeric enzyme. The M36 mutated to I36 caused a change in the hydrophobic interactions in the surrounding residues. The changes in conformation at the active site are caused by the alteration of interaction of the 36th residue with L33 and V77 (Figure 78). The maximum change in conformation is seen in the side chain of K20. The altered K20 hydrogen bonds to protein backbone through additional water molecules in the M36I complex. This is an indiation of the next mutation, K20R to appear, where the R20 may form stronger interactions with the nearby protein backbone instead of water mediated ones via K20. The asparagine residue near the I36 residue makes few extra water mediated and protein backbone interactions in comparison to the V82F (Figure 78). The exact reason for enhanced protease dimer stability due to A71V mutation, appearing after mutations in 82nd and 36th residues, is not clear from the current structures.

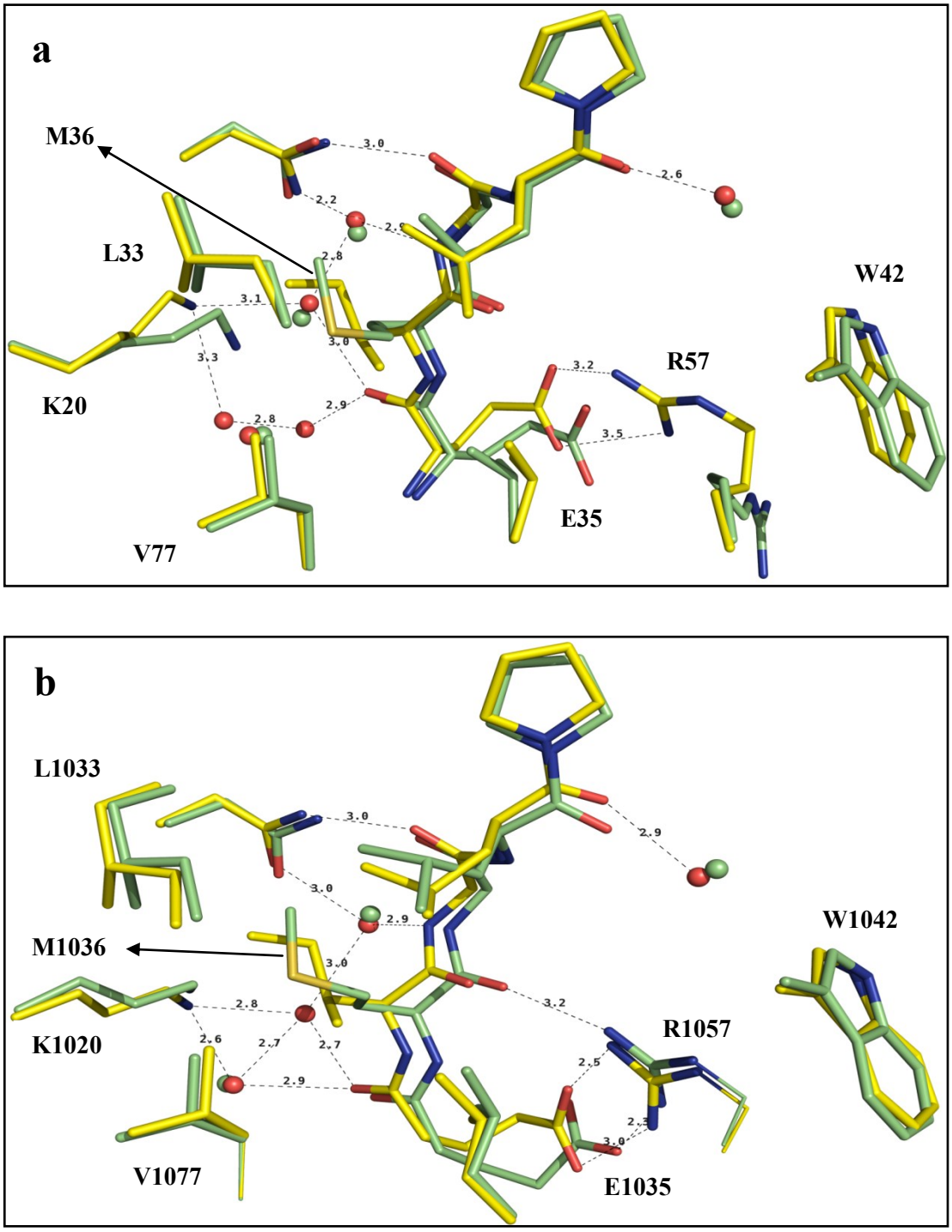


Figure 78: In the M36I (yellow sticks) and V82F (green sticks) complexes the environment around **a)** M36 and **b)** M1036 residues are shown. The hydrogen bonding interactions are shown only for M36I mutant. The water molecules in the V82F complex are shown as green spheres.

6.4 Conclusions

The mutations V82F, I84V, M36I and L10I are all associated with ritonavir-resistance in HIV-1 protease. Three dimensional crystal structures of ritonavir bound to WT, V82F mutant and M36I mutant HIV-1 protease have been determined. Hydrogen bonding interactions between ritonavir and HIV-1 protease are found to be very similar in the three complexes. In the WT complex, methyl groups from V82 in the S1/S1' pocket have hydrophobic and CH-π interactions with benzyl rings of

the drug molecule. In the V82F mutant complex F82 adopts a different rotamer conformation, and as a consequence, CH-π and hydrophobic interactions with the P1/P1' benzyl rings of the drug are lost. Despite conformational adjustments by both the protein and the inhibitor molecule, the active site volume is increased by about 100 $Å^3$ in the mutant complex, suggesting non-optimal van der Waals contacts with ritonavir. These two features are responsible for conferring ritonavir-resistance to this mutant. Molecular models built using the present structure as a template, provide the insight that the secondary mutation L10I acts indirectly by inducing conformational changes in L23. The M36I complex shows that appearaence of M36I mutation helps restore the enzyme fitness. Presence of additional water molecules linking K20 to protein backbone could be an indication of apperance of K20R mutation to further restore enzyme fitness. Ritonavir drug resistance seems to evolve in two stages: an early stage, where active site mutation (V82A/F/T/S) is acquired to cause drug resistance, and a late stage where mutaions, M36I and I54V and later A71V and K20R are acquired for the restoration of fitness and viability of drug resistant mutant enzyme.

References

1. Barre-Sinouussi F, Chermann JC, Rey M, Nugeyre MT, Chamaret S, Gruest J, Dauguet C, Axler-Blin C, Vezinet-Brun F, Rouzioux C, Rozenbaum W and Montagnier L.(1983). Isolation of a zT-lymphotrophic Retrovirus from a Patient at Risk for Acquired Immuno Deficiency Syndrome (AIDS). *Science*, 220, 868-871.
2. Gallo RC, Sahaludin SZ, Popovic M, Shearer GM, Kaplan M, Haynes BF, Palker TJ, Redfield R, Oleske J, Safai B, White G, Foster P and Markham PD. (1984). Frequent detection and isolation of Cytopathic retroviruses (HTLV-III) from patients with AIDS and at Risk for AIDS. *Science*, 224, 500-503.
3. Brun Vezinet. F, Rouzioux C, Barre Sinoussi F, Klatzmann D, Saimot AG, Rozenbaum W, Christol D, Gluckmann JC, Montagnier L and Chermann JC. (1984). Detection of IgG Antibodies to Lymphadenopathy-Associated Virus in patients with AIDS or Lympahadenopathy syndrome. *Lancet*, 1, 1253-1256.
4. Levy JA, Hoffmann AD, Kramer SM, Landis JA, Shimabukuro JM and Oshiro LS. (1984). Isolation of Lymphocytopathic retroviruses from San Francisco patients with AIDS. *Science*, 225, 840-842.
5. Clavel F, Guetard D, Brun Vezinet F, Chameret S, Rey M, SantosFerreira MO, Laurent AG, Dauguet C, Katlama C, Rouzioux C, Klatzmann D, Champalimaud JL and Montagnier L. (1986). Isolation of a new Human Retrovirus from West African Patients with AIDS. *Science*, 233, 343-346.
6. Korber B, Muldoon M, Theiler J, Gao F, Gupta R, Lapedes A, Hahn BH, Wolinsky S and Bhattacharya T. (2000). Timing the ancestor of the HIV-1 pandemic strains. *Science,* 288, 1757–1759.
7. Gao F, Bailes E, Robertson DL, Chen Y, Rodenburg CM, Michael SF, Cummins LB, Arthur LO, Peeters M, Shaw GM, Sharp PM and Hahn BH. (1999). Origin of HIV-1 in the chimpanzee *Pan troglodytes troglodytes*. *Nature*, 397, 436–441.

8. Perrin L, Kaiser L and Yerly S. (2003). Travel and the spread of HIV-1 genetic variants. *The Lancet Infectious Diseases,* 3, 22-27.

9. McCutchan FE. (2000). Understanding the genetic diversity of HIV-1. *AIDS,* 14, S31–S44.

10. Simon F, Mauclere P, Roques P, Loussert-Ajaka I, Muller-Trutwin MC, Saragosti S, Georges-Courbot MC, Barre-Sinoussi F and Brun-Vezinet F. (1998). Identification of a new human immunodeficiency virus type 1 distinct from group M and group O. *Nature Medicine,* 4, 1032–1037.

11. Plantier JC, Leoz M, Dickerson JE, De Oliveira F, Cordonnier F, Lemée V, Damond F, Robertson DL and Simon F. (2009). A new human immunodeficiency virus derived from gorillas. *Nature Medicine*, 8, 871-872.

12. Delgado E, Thomson MM, Villahermosa ML, Sierra M, Ocampo A, Miralles C, Rodriguez-Perez R, Diz-Aren J, Ojea-de Castro R, Losada E, Cuevas MT, Vazquez-de Parga E, Carmona R, Perez-Alvarez L, Medrano L, Cuevas L, Taboada JA and Najera R. (2002). Identification of a newly characterized HIV-1 BG intersubtype circulating recombinant form in Galicia, Spain, which exhibits a pseudotype-like virion structure. *J. Acquir. Immune Defic. Syndr.,* 5, 536–543.

13. Gowda SD, Stein BS, Mohagheghpor N, Benike CJ and Engleman EG. (1989). Evidence that T Cell activation is required for HIV-1 entry in CD4+ Lymphocytes. *J. Immunol.,* 142, 773-780.

14. Zhu P, Winkler H, Chertova E, Taylor KA and Roux KH. (2008). Cryoelectron Tomography of HIV-1 Envelope Spikes: Further Evidence for Tripod-Like Legs. *PLoS Pathogens,* 142 (2), e1000203, 1-9.

15. Muesling MA, Smith DH, Cabradilla CD, Benton CV, Lasky LA and Capon DJ. (1985). Nucleic acid structure and expression of the human AIDS/lymphadenopathy retrovirus. *Nature*, 313, 450-458.

16. Ratner L, Haseline WA, Patarca R, Livak KJ, Starcich B, Josephs SF, Doran ER, Rafalski JA, Whitehorn EA, Baumeister K, Ivanoff L, Petteway SR, Pearson ML, Lautenberger JA, Papas T K, Ghrayeb J, Chang NT, Gallo RC and Wong-Staal F. (1985). Complete nucleotide sequence of the AIDs virus, HTLV-III. *Nature*, 313, 277-284.

17. Mitsuya H and Broder S. (1987). Strategies for antiviral therapy in AIDS. *Nature*, 325, 773-778.

18. Debouk C. (1992). The HIV-1 protease as a therapeutic target for AIDS. *AIDS Res. Hum. Retroviruses,* 8, 153-164.

19. Vondrasek J and Wlodawer A. (2002). HIVdb: a database of the structures of human immunodeficiency virus protease. *Proteins Struct. Funct. Genet.*, 49, 429-431.

20. Leis J, Baltimore D, Bishop JM, Coffin J, Fleissner E, Goff SSP, Oroszlan S, Robinson H, Skalka AM, Temin HM and Vogt V. (1988). Standardized and Simplified Nomenclature for Proteins Common to All Retroviruses. *J. Virol.,* 62 (5), 1808-1809.

21. Dupont S, Sharova N, DeHoratius C, Virbasius CM, Zhu X, Bukrinskaya AG, Stevenson M and Green MR. (1999). A novel nuclear export activity in HIV-1 matrix protein required for viral replication. *Nature*, 402, 681-685.

22. Harrison GP and Lever AM. (1992). The human immunodeficiency virus type 1 packaging signal and major splice donor region have a conserved stable secondary structure. *J. Virol.,* 66, 4144-4153.

23. Turner BG and Summers MF. (1999). Structural Biology of HIV. *J. Mol. Bio.l,* 285, 1-32.

24. Ruben S, Perkins A, Purcell R, Joung K, Sia R, Burghoff R, Haseltine WA and Rosen CA. (1989). Structural and functional characterization of human immunodeficiency virus tat protein. *J. Virol.,* 63, 1-8.

25. Roy S, Delling U, Chen C-H, Rosen CA and Sonenberg N. (1990). A bulge structure in HIV-1 TAR RNA is required for Tat-mediated trans-activation. *Genes Dev.,* 4, 1365-1373.

26. Feinberg MB, Baltimore D and Frankel AD. (1991). The role of Tat in the human immunodeficiency virus life cycle indicates a primary effect on transcriptional elongation. *Proc. Natl. Acad. Sci. USA,* 88, 4045-4049.

27. Daly TJ, Cook KS, Gran GS, Maione TE and Rasche JR. (1989). Specific binding of HIV-1 recombinant Rev protein to the Rev response element in vitro. *Nature*, 342, 816-819.

28. Piguet V, Gu F, Foti M, Demaurex N, Gruenberg J, Carpentier JL and Trono D. (1999). Nef-induced CD4 degradation: a diacidic-based motif in Nef functions as a lysosomal targeting signal through the binding of β-COP in endosomes. *Cell*, 97, 63–73.

29. Mangasarian A, Piguet V, Wang JK, Chen Y and Trono D. (1999). Nef-induced CD4 and major histocompatibility complex class I (MHC-I) down-regulation are governed by distinct determinants:

N-terminal alpha helix and proline repeat of Nef selectively regulate MHC-I trafficking. *J. Virol.,* 73, 1964–1973.

30. Marin M, Rose KM, Kozak SL and Kabat D. (2003). HIV-1 Vif protein binds the editing enzyme APOBEC3G and induces its degradation. *Nature Medicine*, 9, 1398–1403.

31. Bour S, Schubert U and Strebel K. (1995). The human immunodeficiency virus type 1 Vpu protein specifically binds to the cytoplasmic domain of CD4: Implications for the mechanism of degradation. *J. Virol.,* 69, 1510–1520.

32. Van Damme N, Goff D, Katsura C, Jorgenson RL, Mitchell R, Johnson MC, Stephens EB and Guatelli J. (2008). The interferon-induced protein BST-2 restricts HIV-1 release and is downregulated from the cell surface by the viral Vpu protein. *Cell Host Microbe*, 3, 245-252.

33. DeHart JL, Zimmerman ES, Ardon O, Monteiro-Filho CMR, Argañaraz ER and Planelles V. (2007). HIV-1 Vpr activates the G_2 checkpoint through manipulation of the ubiquitin proteasome system. *Virology Journal,* 4, 57.

34. Cunningham A, Donaghy H, Harman A, Kim M and Turville S. (2010). Manipulation of dendritic cell function by viruses. *Current Opinion in Microbiology,* 13 (4), 524–529.

35. Bianchi E, Finotto M, Ingallinella P, Hrin R, Carella, AV, Hou XS, Schleif WA, Miller MD, Geleziunas and Pessi A. (2005). Covalent stabilization of coiled coils of the HIV gp41 N region yields extremely potent and broad inhibitors of viral infection. *Proc. Natl. Acad. Sci. USA*, 102(36), 12903-12908.

36. Deng H, Liu R, Ellmeier W, Choe S, Unutmaz D, Burkhart M, Di-Marzio P, Marmon S, Sutton RE, Hill CM, Davis CB, Peiper SC, Schall TJ, Littman DR and Landau NR. (1996). Identification of a major co-receptor for primary isolates of HIV-1. *Nature*, 381, 661-666.

37. Dragic T, Litwin V, Allaway GP, Martin SR, Huang Y, Nagashima KA, Cayanan C, Maddon PJ, Koup RA, Moore JP and Paxton WA. (1996). HIV-1 entry into the CD4 (+) cells is mediated by the chemokine receptor CC-CKR-5. *Nature*, 381, 667-683.

38. Littman DR. (1998). Chemokine receptors: Keys to AIDS pathogenesis? *Cell*, 93, 677-680.

39. Briggs JAG and Kräusslich H-G. (2011). The Molecular Architecture of HIV. *J. Mol. Biol.*, 410(4), 491-500.

40. Arhel N. (2010). Revisiting HIV-1 uncoating. *Retrovirology*, 7, 96.

41. Beerens N, Groot F and Berkhout B. (2001). Initiation of HIV-1 reverse transcription is regulated by a primer activation signal. *J. Biol. Chem*., 276 (33), 31247-31256.

42. Christ F, Thys W, Rijck JD, Gijsbers R, Albanese A, Arosio D, Emiliani S, Rain J-C, Benarous R, Cereseto A and Debyser Z. (2008). Transportin-SR2 Imports HIV into the Nucleus. *Current Biology*, 18, 1192–1202.

43. Hearps AC and Jans DA. (2006). HIV-1 integrase is capable of targeting DNA to the nucleus via an Importin α/β-dependent mechanism. *Biochem. J.,* 398, 475–484.

44. Huffman KM, Arrigo SJ and Schmidt MG. (1999). HIV-1 Rev promotes the nuclear export of unspliced and singly spliced RNAs in a mammalian cell-free export system. *J. Biomed. Sci*., 6(3), 194-205.

45. Blissenbach M, Grewe B, Hoffmann B, Brandt S and Uberla K. (2010). Nuclear RNA export and packaging functions of HIV-1 Rev revisited. *J. Virol.*, 84, 6598-6604.

46. Parkin NT, Chamorro M and Varmus HE. (1992). Human Immunodeficiency virus type 1 gag-pol frameshifting is dependent on mRNA secondary structure: Demonstration by expression in vivo. *J. Virol.,* 66, 5147-5151.

47. Shehu-Xhilaga M, Kraeusslich HG, Pettit S, Swanstrom R, Lee JY, Marshall JA, Crowe SM and Mak J. (2001). Proteolytic Processing of the P2/Nucleocapsid Cleavage Site Is Critical for Human Immunodeficiency Virus Type 1 RNA Dimer Maturation. *J. Virol.*, 75, 9156–9164.

48. Marco A, Muller B, Glass B, Riches JD, Krausslich H-G,.Briggs JAG. (2010). Structural Analysis of HIV-1 Maturation Using Cryo- Electron Tomography. *PLoS Pathogens*, 6(11), e1001215.

49. Girard MP, Osmanov S, Assossou OM and Kieny MP. (2011). Human immunodeficiency virus (HIV) immunopathogenesis and vaccine development: a review. *Vaccine*, 29(37), 6191-6218.

50. Asmuth DM, Brown EL, DiNubile MJ, Sun X, del Rio C, Harro C, Keefer MC, Kublin JG, Dubey SA, Kierstead L, et. al. (2010). Comparative cell-mediated immunogenicity of DNA/DNA, DNA/adenovirus type 5 (Ad5), or Ad5/Ad5 HIV-1 clade B gag vaccine prime-boost regimens. *J. Infect. Dis.*, 201, 132–141.

51. McElrath MJ, De Rosa SC, Moodie Z, Dubey S, Kierstead L, Janes H, Defawe OD, Carter DK, Hural J, Akondy R., et al. Step Study Protocol Team. (2008). HIV-1 vaccine-induced immunity in the test-of-concept Step Study: A case-cohort analysis. *Lancet,* 372, 1894–1905.

52. Hoxie JA. (2010). Toward antibody based HIV-1 Vaccine. *Annual Review of Medicine*, 61, 135-152.

53. Barouch DH, O'Brien KL, Simmons NL, King SL, Abbink P, Maxfield LF, Sun YH, La Porte A., Rigg AM, Lynch DM, et al. (2010). Mosaic HIV-1 vaccines expand the breadth and depth of cellular immune responses in rhesus monkeys. *Nature Medicine*, 16, 319–323.

54. Mascola JR and Montefiori DC. (2010). The role of antibodies in HIV vaccines. *Annu. Rev. Immunol*., 28, 413–444.

55. Dimitrov DS. (2010). Therapeutic antibodies, vaccines and antibodyomes. *mAbs*, 2 (3), 347–356.

56. Zhou T, Georgiev I, Wu X, Yang ZY, Dai K, Finzi A, Do Kwon Y, Scheid JF, Shi W, Xu L, et al. (2010). Structural basis for broad and potent neutralization of HIV-1 by antibody VRC01. *Science,* 329, 811–817.

57. Garcı´a F, Climent N, Assoumou L et. al. (2011). A therapeutic dendritic cell-based vaccine for HIV-1 infection. *J. Infect. Dis.,* 203, 473–478.

58. Burton DR. (2010). Scaffolding to build a rational vaccine design strategy. *Proc. Natl. Acad. Sci. USA,* 107, 17859–17860.

59. Virgin HW and Walker BD. (2010). Immunology and the elusive AIDS vaccine. *Nature*, 464, 224–231.

60. Van Regenmortel MHV. (2011). Requirements for empirical immunogenicity trials. *Archives of Virology*, DOI: 10.1007/s00705-011-1145-2.

61. Rossi JJ, June CH and Kohn DB. (2007). Genetic therapies against HIV. *Nature Biotechnology*, 25 (12), 1444-1454.

62. Scherer LJ and Rossi JJ (2011). Ex vivo gene therapy for HIV-1 treatment. *Hum. Mol.Genet.,* 20(R1), R100-107.

63. Swan CH, Bühler B, Steinberger P, Tschan MP, Barbas CF and Torbett BE. (2006). T-cell protection and enrichment through lentiviral CCR5 intrabody gene delivery. *Gene Ther.,* 13(20), 1480-1492.

64. Swan CH and Torbett BE. (2006). Can gene delivery close the door to HIV-1 entry after escape? *J. Med. Primatol.,* 35 (4-5), 236-247.

65. Bai J, Rossi J and Akkina R. (2001). Multivalent anti-CCR ribozymes for stem cell-based HIV type 1 gene therapy. *AIDS Res. Hum. Retroviruses,* 17(5), 385-399.

66. Zeller SJ and Kumar P. (2011). RNA-based Gene Therapy for the Treatment and Prevention of HIV: From Bench to Bedside. *Yale J. Biol. Med.,* 84(3), 301-309.

67. Liu YP, Westerink JT, Ter Brake O and Berkhout B. (2011). RNAi-inducing lentiviral vectors for anti-HIV-1 gene therapy. *Methods Mol. Biol.,* 721, 293-311.

68. Wayengera M. (2011). Proviral HIV-genome-wide and pol-gene specific Zinc Finger Nucleases: usability for targeted HIV gene therapy. *Theor. Biol. Med. Model*., 8, 26.

69. Wilen CB, Wang J, Tilton JC, Miller JC, Kim KA, Rebar EJ, Sherrill-Mix SA, Patro SC, Secreto AJ, Jordan AP, Lee G, Kahn J, Aye PP, Bunnell BA, Lackner AA, Hoxie JA, Danet-Desnoyers GA, Bushman FD, Riley JL, Gregory PD, June CH, Holmes MC and Doms RW. (2011). Engineering HIV-resistant human CD4+ T cells with CXCR4-specific zinc-finger nucleases. *PLoS Pathogen*, 7(4), e1002020.

70. Bitton N, Gorochov G, Debre P and Eshhar Z. (1999). Gene therapy approaches to HIV-infection: immunological strategies: use of T bodies and universal receptors to redirect cytolytic T-cells. *Front. Biosci.,* 4, D386-93.

71. Lee CL, Dang J, Joo KI and Wang P. (2011). Engineered lentiviral vectors pseudotyped with a CD4 receptor and a fusogenic protein can target cells expressing HIV-1 envelope proteins. *Virus Res.,* 160 (1-2), 340-350.

72. Holt N, Wang J, Kim K, Friedman G, Wang X, Taupin V, Crooks GM, Kohn DB, Gregory PD, Holmes MC and Cannon PM. (2010). Human hematopoietic stem/progenitor cells modified by zinc-finger nucleases targeted to CCR5 control HIV-1 in vivo. *Nature Biotechnology*, 28(8):839-847.

73. Sarkar I, Hauber I, Hauber J and Buchholz F. (2007). HIV-1 proviral DNA excision using an evolved recombinase. *Science*, 316, 1912-1915.

74. Briones M, Tekeste S, Rusmevichientong A and Cho SA. (2010). Human and rhesus macaque TRIM5α can bind HIV-1 cores and stabilize CA disassembly. *J. Virol*., 84, 5181-5190.

75. Johnson SF and Telesnitsky A. (2010). Retroviral RNA Dimerization and Packaging: The What, How, When, Where, and Why. *PLoS Pathogens*, 6(10), e1001007.

76. Gulick RM, Mellors JW, Havlir D, Eron JJ, Gonzalez C, McMahon D, Richman DD, Valentine FT, Jonas L, Meibohm A, Emini EA and Chodakewitz JA. (1997). Treatment with indinavir, zidovudine, and lamivudine in adults with human immunodeficiency virus infection and prior antiretroviral therapy. *New Eng. J. Med.,* 337, 734–739.

77. Teixeira C, Gomes JRB, Gomes P and Maurel F. (2011). Viral surface glycoproteins, gp120 and gp41, as potential drug targets against HIV-1: Brief overview one quarter of a century past the approval of zidovudine, the first anti-retroviral drug. *European Journal of Medicinal Chemistry*, 46(4), 979-992.

78. Lalezari JP, Eron JJ, Carlson M, Cohen C, DeJesus E, Arduino RC, Gallant JE, Volberding P, Murphy RL, Valentine F, Nelson EL, Sista PR, Dusek A and Kilby JM. (2003). A phase II clinical study of the long-term safety and antiviral activity of enfuvirtide-based antiretroviral therapy. *AIDS*, 17, 691-698.

79. Chun TW, Stuyver L, Mizell SB, Ehler LA, Mican JA, Baseler M, Lloyd AL, Nowak MA and Fauci AS. (1997). Presence of an inducible HIV-1 latent reservoir during highly active antiretroviral therapy. *Proc. Natl. Acad. Sci. USA,* 94, 13193–13197.

80. Finzi D, Hermankova M, Pierson T, Carruth LM, Buck C, Chaisson RE, Quinn TC, Chadwick K, Margolick J, Brookmeyer R, Gallant J, Markowitz M, Ho DD, Richman DD and Siliciano RF. (1997). Identification of a reservoir for HIV-1 in patients on highly active antiretroviral therapy. *Science,* 278, 1295–1300.

81. Wong JK, Hezareh M, Gunthard HF, Havlir DV, Ignacio CC, Spina CA, Richman DD. (1997). Recovery of replication-competent HIV-1 despite prolonged suppression of plasma viremia. *Science,* 278, 1291–1295.

82. Wiegers K, Rutter G, Kottler H, Tessmer U, Hohenberg H and Hans-Georg K. (1998). Sequential steps in Human Immunodeficiency Virus particle maturation revealed by alterations of individual Gag Polyprotein cleavage sites. *J. Virol.,* 72, 2846-2854.

83. Kohl NE, Emini EA, Scheilf WA, Davis LJ, Heimbach JC Dixon RA,Scolnick EM and Sigal IS. (1988). Active Human immunodeficiency virus protease is required for viral infectivity. *Proc. Natl. Acad. Sci. USA,* 85, 4686-4690.

84. Chou KC, Tomasselli AG, Reardon IM, Heinrikson RL. (1996). Predicting human immunodeficiency virus protease cleavage sites in proteins by a discriminant function method. *Proteins*, 24(1), 51-72.

85. Loeb DD, Swanstrom R, Everitt L, Manchester M, Stamper SE and Hutchison CA. (1989). Complete mutagenesis of the HIV-1 protease. *Nature*, 340(6232), 397-400.

86. Seelmeier S, Schimdt 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. Natl. Acad. Sci. USA,* 85, 6612-6616.

87. Mcquade TJ, Tomasseli AG, Liu L, Karacostas V, Moss B Sawyer TK, Heinrickson RL and Tarpley WG. (1990). A synthetic HIV-1 protease inhibitor with antiviral activity arrests HIV-1 like particle maturation. *Science*, 247, 454-456.

88. Cheng YS, Yin FH, Foundling S, Blomstrom D and Kettner CA. (1990). Stability and activity of human immunodeficiency virus protease: comparison of the natural dimer with a homologous, single-chain tethered dimer. *Proc. Natl. Acad. Sci. USA,* 87, 9660–9664.

89. Darke PL, Nutt RF, Brady SF, Garsky VM, Ciccarone TM, Leu CT, Lumma PK, Freidinger RM, Veber DF and Sigal IS. (1988). HIV-1 protease specificity of peptide cleavage is sufficient for processing of *Gag* and *Pol* polyproteins. *Biochem. Biophys. Res. Commun.,* 156, 297–303.

90. Debouck C, Gorniak JG, Strickler JE, Meek TD, Metcalf BW and Rosenberg M. (1987). Human immunodeficiency virus protease expressed in *Escherichia coli* exhibits autoprocessing and specific maturation of the *gag* precursor. *Proc. Natl. Acad. Sci. USA,* 84, 8903–8906.

91. Schneider J and Kent SB. (1988). Enzymatic activity of a synthetic 99 residue protein corresponding to the putative HIV-1 protease. *Cell,* 54, 363–368.

92. Navia MA, Fitzgerald PMD, Mckeever BM, Leu CT, Heimbach JC, Herber WK, Sigal IS, Darke PL and Springer JP. (1989). Three-dimensional structure of aspartic protease from human immunodeficiency virus HIV-1. *Nature*, 337, 615-620.

93. Wlodawer A, Miller M, Jaskolski M, Sathyanarayana BK, Baldwin E, Weber IT, Selk LM, Clawson L, Schneider J and Kent SBH. (1989). Conserved folding in retroviral proteases: crystal structure of syntheic HIV-1 protease. *Science*, 245, 616-621.

94. Lapatto R, Blundell T, Hemmings A, Overington J, Widerspin AM, Wood S, Merson JR, Whittle PJ, Danley DE, Geoghegan KF, Hawrylik SJ, Lee SE, Scheld K and Hobart PM. (1989). X-ray analysis of HIV-1 proteinase at 2.7 Å resolution confirms structural homology among retroviral enzymes. *Nature*, 342, 299-302.

95. Pillai B, Kannan KK and Hosur MV. (2001). 1.9 Å X-ray study shows closed flap conformation in crystals of tethered HIV-1 PR. *Proteins,* 43, 57-64.

96. Kruger K, Grabowski PJ, Zaug AJ, Sands J, Gottschling DE and Cech TR. (1982). Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena. *Cell,* 31 (1), 147–57.

97. Silverman SK. (2009). Deoxyribozymes: selection design and serendipity in the development of DNA catalysts. *Acc. Chem. Res.,* 42(10), 1521-1531.

98. Garrett R and Grisham CM. (2008). *Biochemistry*, Thomson Brooks/cole Ed., pg. 383.

99. Eyring H. (1935). The activated complex in chemical reactions. *J. Chem Phys.,* 3, 107-115.

100. Pauling L. (1948). Chemical achievement and hope for the future. *Am. Sci.,* 36, 50–58.

101. Warshel A, Sharma PK, Kato M, Xiang Y, Liu H and Olsson MH. (2006). Electrostatic basis for enzyme catalysis. *Chem. Rev.,* 106 (8), 3210–35.

102. Mesecar AD, Stoddard BL and Koshland Jr. DE. (1997). Orbital Steering in the catalytic power of enzymes: small structural changes with large catalytic consequences. *Science*, 277, 202-206.

103. Liang ZX and Klinman JP. (2004). Structural basis of hydrogen tunneling in enzymes: progress and puzzles. *Curr. Opin. Struct. Biol.,* 14, 648-655.

104. Gerlt JA, Kreevoy MM, Cleland W and Frey PA. (1997). Understanding enzymic catalysis: the importance of short, strong hydrogen bonds. *Chem. Biol.,* 4, 259-267.

105. Olsson MH, Parson WW and Warshel A. (2006). Dynamical Contributions to Enzyme Catalysis: Critical Tests of A Popular Hypothesis. *Chem. Rev.* 106 (5), 1737–1756.

106. Cornish-Bowden A. (2002). Enthalpy–entropy compensation: a phantom phenomenon. *Journal of Biosciences,* 27 (2), 121–126.

107. Szecsi PB. (1992). The aspartic proteases. *Scand. J. Clin. Lab. In vest. Suppl.* 210, 5–22.

108. Ram CV. (2009). Direct inhibition of renin: a physiological approach to treat hypertension and cardiovascular disease. *Future Cardiol.,* 5 (5), 453–465.

109. Gupta D, Yedidi RS, Varghese S, Kovari LC and Woster PM. (2010). Mechanism-Based Inhibitors of the Aspartyl Protease Plasmepsin II as Potential Antimalarial Agents. *J. Med. Chem.*, 53 (10), 4234–4247.

110. Vashishta A, Ohri SS and Vetvicka V. (2009). Pleiotropic Effects of Cathepsin D. *Endocrine, Metabolic and Immune Disorders - Drug Targets*, 9, 385-391.

111. Selkoe DJ. (2001). Alzheimer's disease: genes, proteins, and therapy. *Physiol. Rev.,* 81(2), 741-766.

112. Miller M, Jaskolski M, Rao JK, Leis J and Wlodawer A. (1989). Crystal structure of a retroviral protease proves relationship to aspartic protease family. *Nature,* 337, 576–579.

113. Loeb DD, Hutchison CA, Edgell MH, Farmerie WG and Swanstrom R. (1989). Mutational analysis of human immunodeficiency virus type 1 protease suggests functional homology with aspartic proteinases. *J. Virol.,* 63, 111–121.

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

115. Pechik V, Gustchina AE, Andreeva NS and Fedorov AA. (1989). Possible role of some groups in the structure and function of HIV-1 protease as revealed by molecular modeling studies. *FEBS Letters,* 247, 118-122.

116. Pearl LH and Taylor WRA. (1987). Structural model for the retroviral proteases. *Nature*, 329, 351-354.

117. Rao JK, Erickson JW and Wlodawer A. (1991). Structural and evolutionary relationships between retroviral and eukaryotic aspartic proteinases. *Biochemistry*, 30, 4663-71.

118. Davies DR. (1990). The structure and function of the aspartic proteinases. *Annu. Rev. Biophys. Chem.,* 19, 189-205.

119. Mulichak AM, Hui JO, Tomasselli AG, Heinrikson RL, Curry KA, Tomich CS, Thaisrivongs S, Sawyer TK and Watenpaugh KD. (1993). The crystallographic structure of the protease from human immunodeficiency virus type 2 with two synthetic peptidic transition state analog inhibitors. *J. Biol. Chem.,* 268, 13103–13109.

120. Tong L, Pav S, Pargellis C, Do F, Lamarre D and Anderson PC. (1993). Crystal structure of human immunodeficiency virus (HIV) type 2 protease in complex with a reduced amide inhibitor and comparison with HIV-1 protease structures. *Proc. Natl. Acad. Sci. USA,* 90, 8387–8391.

121. Rose RB, Rose JR, Salto R, Craik CS and Stroud RM. (1993). Structure of the protease from simian immunodeficiency virus: complex with an irreversible nonpeptide inhibitor. *Biochemistry,* 32, 12498–12507.

122. Zhao B, Winborne E, Minnich MD, Culp JS, Debouck C and Abdel MSS. (1993). Three-dimensional structure of a simian immunodeficiency virus protease/inhibitor complex. Implications for the design of human immunodeficiency virus type 1 and 2 protease inhibitors. *Biochemistry,* 32, 13054–13060.

123. Wlodawer A, Gustchina A, Reshetnikova L, Lubkowski J, Zdanov A, Hui KY, Angleton EL, Farmerie WG, Goodenow MM, Bhatt D, Zhang L and Dunn BM. (1995). Structure of an inhibitor complex of the proteinase from feline immunodeficiency virus. *Nature Struct. Biol.,* 2, 480–488.

124. Gustchina A, Kervinen J, Powell DJ, Zdanov A, Kay J and Wlodawer A. (1996). Structure of equine infectious anemia virus proteinase complexed with an inhibitor. *Protein Sci.,* 5, 1453–1465

125. Wlodawer A and Erickson JW. (1993). Structure-based inhibitors of HIV-1 protease. *Annu. Rev. Biochem.,* 62, 543–585.

126. 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.

127. Rutenber E, Fauman EB, Keenan RJ, Fong S, Furth PS, Ortiz de Montellano PR, Meng E, Kuntz ID, DeCamp DL and Salto R. (1993). Structure of a non-peptide inhibitor complexed with HIV-1 protease. Developing a cycle of structure-based drug design. *J. Biol. Chem.,* 268, 15343–15346.

128. Lam PYS, Jadhav PK, Eyermann CJ, Hodge CN, Ru Y, Bacheler LT, Meek JL, Otto MJ, Rayner MM, Wong YN, Chang CH, Weber PC, Jackson DA, Sharpe TR and Erickson-Viitanen S. (1994). Rational design of potent, bio-available, non-peptide cyclic ureas as HIV-1 protease inhibitors. *Science,* 263, 380-384.

129. Tomasselli AG, Howe WJ, Sawyer TK, Wlodawer A and Heinrikson RL. (1991). The complexities of AIDS: an assessment of the HIV-1 protease as a therapeutic target. *Chimica Oggi,* 9, 6–27.

130. Jeyabalan MP, Nalivaika E and Schiffer CA. (2002). Substrate Shape determines Specificity of Recognition for HIV-1 Protease: Analysis of Crystal Structures of Six Substrate complexes. *Structure*, 10, 369-381.

131. Gustchina A, Sansom C, Prevost M, Richelle J, Wodak SY, Wlodawer A and Weber IT. (1994). Energy calculations and analysis of HIV-1 protease-inhibitor crystal structure. *Protein Engg.,* 7, 309–317.

132. Miller M, Schneider J, Sathyanarayana BK, Toth MV, Marshall GR, Clawson L, Selk L, Kent SBH and Wlodawer A. (1989). Structure of complex of synthetic HIV-1 protease with a substrate-based inhibitor at 2.3 Å resolution. *Science,* 246, 1149-1152.

133. Schechter I and Berger A. (1967). On the size of the active site in proteases. *Biochem. Biophys. Res. Commun.,* 27, 157–162.

134. Tözsér J, Weber IT, Gustchina A, Blaha I, Copeland TD, Louis JM and Oroszlan S. (1992). Kinetic and modeling studies of S3-S3' subsites of HIV-1 proteinases. *Biochemistry,* 31, 4793–4800.

135. Dunn BM, Gustchina A, Wlodawer A and Kay J. (1994). Subsite preferences of retroviral proteinases. *Methods Enzymol.,* 241, 254–278.

136. Kervinen J, Lubkowski J, Zdanov A, Bhatt D, Dunn BM, Hui KY, Powell DJ, Kay J, Wlodawer A and Gustchina A. (1998). Toward a universal inhibitor of retroviral proteases: comparative analysis of the interactions of LP-130 complexed with proteases from HIV-1, FIV, and EIAV. *Protein Sci.,* 7, 2314-2323.

137. Wlodawer A, Gustchina A, Reshetnikova L, Lubkowski J, Zdanov A, Hui KY, Angleton EL, Farmerie WG, Goodenow MM, Bhatt D, Zhang L and Dunn BM. (1995). Structure of an inhibitor complex of the proteinase from feline immunodeficiency virus. *Nature Struct. Biol*., 2, 480–488.

138. Gustchina A, Kervinen J, Powell DJ, Zdanov A, Kay J and Wlodawer A. (1996). Structure of equine infectious anemia virus proteinase complexed with an inhibitor. *Protein Sci.,* 5, 1453–1465.

139. Tözsér J, Gustchina A, Weber IT, Blaha I, Wondrak EM and Oroszlan S. (1991). Studies on the role of the S4 substrate binding site of HIV-1 proteinases. *FEBS Lett.,* 279, 356-360.

140. Vondrasek J and Wlodawer A. (2002). HIVdb: a database of the structures of human immunodeficiency virus protease. *Proteins,* 49(4), 429-431.

141. Suguna K, Padlan EA, Smith CW, Carlson WD and Davies DR. (1987). Binding of a reduced peptide inhibitor to the aspartic proteinase from Rhizopus chinensis: implications for a mechanism of action. *Proc. Natl. Acad. Sci. USA,* 84, 7009-7013.

142. Pearl LH. (1987). The catalytic mechanism of aspartic proteinases. *FEBS Lett,* 214, 8-12.

143. Hyland LJ, Tomaszek TA and Meek TD. (1991). HIV-1 protease. Use of pH rate studies and solvent kinetic isotope effects to elucidate details of the chemical mechanism. *Biochemistry*, 30, 8454-8463.

144. Hyland LJ, Tomaszek TA, Roberts GD, Carr SA, Magaard VW, Bryan HL, Fakhoury SA, Moore ML, Minnich MD, Culp JS, DesJarlias RL and Meek TD. (1991). Human immunodeficiency virus-1 protease. Initial velocity studies and kinetic characterization of reaction intermediates by ^{18}O isotope exchange. *Biochemistry*, 30, 8441 – 8453.

145. Jaskolski M, Tomasselli AG, Sawyer TK, Staples DG, Heinrickson RL, Schneider J, Kent SBH and Wlodawer A. (1991). Structure at 2.5-A resolution of chemically synthesized human immunodeficiency type 1 protease complexed with a hydroxyethylene-based inhibitor. *Biochemistry*, 30, 1600 – 1609; Dunn BM and Fink AL. (1984), *Biochemistry*, 23, 5241.

146. James MNG, Sielecki AR, Hayakawa K and Gelb MH. (1992). Crystallographic analysis of transition state mimics bound to penicillopepsin: difluorostatine- and difluorostatone-containing peptides. *Biochemistry*, 31, 3872-3886.

147. Rodriguez EJ, Angeles TS and Meek TD. (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.

148. Silva AM, Cachau RE, Sham HL and Erickson JW. (1996). Inhibition and catalytic mechanism of HIV-1 aspartic protease. *J. Mol. Biol.,* 255, 321-346.

149. Rose RB, Craik CS, Douglas NL and Stroud RM. (1996). Three-dimensional structures of HIV-1 and SIV protease product complexes. *Biochemistry,* 35, 12933-12944.

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

151. Dunn BM. (2002). Structure and mechanism of the pepsin-like family of aspartic peptidases. *Chem. Rev.*, 102, 4431–4458.

152. Brik A and Wong CH. (2003). HIV-1 protease: mechanism and drug discovery. *Org. Biomol. Chem.*, 1, 5-14.

153. Kovalevsky AV, Chumanevich AA, Liu F, Louis JM and Weber IT. (2007). Caught in the act: the 1.5 Å resolution crystal structures of the HIV-1 protease and the I54V mutant reveal a tetrahedral reaction intermediate. *Biochemistry*, 46, 14854–14864.

154. Torbeev VY, Mandal K, Terechko VA and Kent SBH. (2008). Crystal structure of chemically synthesized HIV-1 protease and a ketomethylene isostere inhibitor based on the p2/NC cleavage site. *Bioorg. Med. Chem. Lett.,* 18, 4554-4557.

155. Adachi M, Ohhara T, Kazuo Kurihara, Taro Tamada, Eijiro Honjo, Nobuo Okazaki, Shigeki Arai, Yoshinari Shoyama, Kaname Kimura, Hiroyoshi Matsumura, Shigeru Ugiyama, Hiroaki Adachi, Kazufumi Takano, Yusuke Mori, Koushi Hidaka, Tooru Kimura, Yoshio Hayashi, Yoshiaki Kiso and Ryota Kuroki. (2009). Structure of HIV-1 protease in complex with potent inhibitor KNI-272 determined by high-resolution X-ray and neutron crystallography. *Proc. Natl. Acad. Sci. USA,* 106(12), 4641–4646.

156. Piana S, Sebastiani D, Carloni P and Parinello M. (2001). Ab Initio Molecular Dynamics-Based Assignment of the Protonation State of Pepstatin A/HIV-1 Protease Cleavage Site. *J. Am. Chem. Soc.,* 123, 8730 – 8737.

157. Piana S and Carloni P. (2001). Conformational flexibility of the catalytic dyad in HIV-1 protease: an ab initio study on the free enzyme. *Proteins: Struct. Funct. Genet.*, 39, 26-36.

158. Piana S, Carloni P and Parrinello M. (2002). Role of conformational fluctuations in the enzymatic reaction of HIV-1 protease. *J. Mol. Biol.,* 319, 567-83.

159. Venturini A, Lopez-Ortiz F, Alvarez JM and Gonzalez J. (1998). Theoretical proposal of a catalytic mechanism for the HIV-1 protease involving an enzyme bound tetrahedral intermediate. *J. Am. Chem. Soc.,* 120, 1110 – 1111.

160. Okimoto N, Tsukui T, Hata M, Hoshino T and Tsuda M. (1999). Hydrolysis Mechanism of the Phenylalanine−Proline Peptide Bond Specific to HIV-1 Protease: Investigation by the ab initio Molecular Orbital Method. *J. Am. Chem. Soc.,* 121, 7349 – 7354.

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

162. Bjelic S and Aqvist J. (2006). Catalysis and linear free energy relationships in aspartic proteases. *Biochemistry*, 45, 7709 – 7723.

163. Liu H, Mu¨ller-Plathe F and van Gunsteren WF. (1996). A combined quantum/ classical molecular dynamics study of the catalytic mechanism of HIV protease. *J. Mol. Biol.*, 261, 454–469.

164. 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.

165. Trylska J, Grochowski P and McCammon JA. (2004). The role of hydrogen bonding in the enzymatic reaction catalyzed by HIV-1 protease. *Protein Sci.,* 13, 513–528; Coates L, Tuan H-T, Tomanicek S, Kovalevsky A, Mustyakimov M, Erskine P and Cooper J. (2008). *J. Am. Chem. Soc.,* 130, 7235.

166. Fruton JS. (1976). Mechanism of Catalytic Action of Pepsin and Related Acid Proteinases. *Advances in Enzymology and related areas of molecular biology*, 44, 1-36.

167. Fruton JS. (1987). Aspartyl proteinases. In *Hydrolytic Enzymes* (Neuberger A and Brocklehurst K eds), pp. 1–37, Elsevier, Amsterdam.

168. Cleland WW, Frey PA and Gerlt JA. (1998). The Low-Barrier Hydrogen Bonds in Enzymatic Catalysis. *J. Biol. Chem.,* 273, 25529–25532.

169. Frey PA, Whitt SA and Tobin JB. (1994). Low-barrier hydrogen bonding in molecular complexes analogous to histidine and aspartate in the catalytic triad of serine proteases. *Science,* 264, 1927–1930.

170. Yamaguchi S, Kamikuboa H, Kuriharab K, Kurokib R, Niimurac N, Shimizud N, Yamazakia Y and Kataokaa M. (2009). Low-barrier hydrogen bond in photoactive yellow protein. *Proc. Natl. Acad. Sci. USA,* 106(2), 440-444.

171. Tözsér J, Bláha I, Copeland TD, Wondrak EM and Oroszlan S. (1991). Comparison of the HIV-1 and HIV-2 proteinases using oligopeptide substrates representing cleavage sites in Gag and Gag-Pol polyproteins. *FEBS Lett*., 281, 77-80.

172. Ho DD, Neumann AU, Perelson AS, Chen W, Leonard JM and Markowitz M. (1995). Rapid turnover of plasma virions and CD3 lymphocytes in HIV-1 infection. *Nature,* 373, 123-126.

173. Wei X, Ghosh SK, Taylor ME, Johnson VA, Emini EA, Deutsch P, et.al. *(*1995). Viral dynamics in HIV-1 infection. *Nature*, 373, 117-122.

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

175. Erickson JW. (1995). The not-so-great escape. *Nature Struct. Biol.,* 2(7), 523–529.

176. Richman DD. (1996). Antiretroviral drug resistance: Mechanisms, pathogenesis, clinical significance. *Adv. Exp. Med. Biol.,* 394, 383–395.

177. Anderson J, Schiffer C, Lee S-K and Swanstrom R. (2009). Viral protease inhibitors. *Handb. Exp. Pharmacol.,* 189, 85-110.

178. Pokorná J, Machala L, Rezácová P and Konvalinka J. (2009). Current and Novel Inhibitors of HIV Protease. *Viruses*, 1, 1209-1239.

179. Wensing AM, van Maarseveen NM and Nijhuis M. (2010). Fifteen years of HIV Protease Inhibitors: raising the barrier to resistance. *Antiviral Res.,* 85, 59-74.

180. Katzenstein TL, Petersen AB, Storgaard M, Obel N, Jensen-Fangel S, Nielsen C and Jørgensen LB. (2010). Phylogeny and resistance profiles of HIV-1 POL sequences from rectal biopsies and blood. *J. Med. Virol.,* 82(7), 1103-1109.

181. Dam E, Quercia R, Glass B, Descamps D, Launay O, Duval X, Kräusslich H-G, Hance AJ and Clavel F. (2009). Gag Mutations Strongly Contribute to HIV-1 Resistance to Protease Inhibitors in Highly Drug-Experienced Patients besides Compensating for Fitness Loss. *PLoS Pathogens*, 5(3), e1000345.

182. Bennett DE, Camacho RJ, Otelea D, Kuritzkes DR, Fleury H, Kiuchi M, Heneine W, et al. (2009). Drug Resistance Mutations for Surveillance of Transmitted HIV-1 Drug-Resistance: 2009 Update. *PLoS One*, 4, e4724.

183. Johnson VA, Brun-Vézinet F, Clotet B, Günthard HF, Kuritzkes DR, Pillay D, Schapiro JM and Richman DD. (2010). Update of the Drug Resistance Mutations in HIV-1: December 2010. *Top. HIV Med.,* 18, 156-163.

184. Condra JH, Schleif WA, Blahy OM, Gabryelski LJ, Graham DJ, Quintero JC, Rhodes A, Robbins HL, Roth E, Shivaprakash M, Titus D, Yang T, Teppler H, Squires KE, Deutsch PJ and

Emini EA. (1995). *In vivo* emergence of HIV-1 variants resistant to multiple protease inhibitors. *Nature,* 374, 569–571.

185. Molla A, Korneyeva M, Gao Q, Vasavanonda S, Schipper PJ, Mo HM, Markowitz M, Chernyavskiy T, Niu P, Lyons N, Hsu A, Granneman GR, Ho DD, Boucher CA, Leonard JM, Norbeck DW and Kempf DJ. (1996). Ordered accumulation of mutations in HIV protease confers resistance to ritonavir. *Nature Medicine,* 2(7), 760 –766.

186. Ives KJ, Jacobsen H, Galpin SA, Garaev MM, Dorrell L, Mous J, Bragman K and Weber JN. (1997). Emergence of resistant variants of HIV *in vivo* during monotherapy with the proteinase inhibitor saquinavir. *J. Antimicrob. Chemother.,* 39(6), 771–779.

187. Sardana VV, Schlabach A., Graham P, Bush BL, Condra JH, Culberson JC, Gotlib L, Graham DJ, Kohl NE, LaFemina RL, Schneider CL, Wolanski BS, Wolfgang JA and Emini EA. (1994). Human immunodeficiency virus type 1 protease inhibitors: Evaluation of resistance engendered by amino acid substitutions in the enzyme's substrate binding site. *Biochemistry,* 33, 2004 –2010.

188. Gulnik SV, Suvorov LI, Liu B, Yu B, Anderson B, Mitsuya H and Erickson JW. (1995). Kinetic characterization and cross-resistance patterns of HIV-1 protease mutants selected under drug pressure. *Biochemistry,* 34(29), 9282–9287.

189. Rosin CD, Belew RK, Walker WL, Morris GM, Olson AJ and Goodsell DS. (1999). Coevolution and subsite decomposition for the design of resistance-evading HIV-1 protease inhibitors. *J. Mol. Biol.,* 287, 77–92.

190. Ridky TW, Kikonyogo A, Leis J, Gulnik S, Copeland T, Erickson J,Wlodawer A, Kurinov I, Harrison RW and Weber IT. (1998). Drug-resistant HIV-1 proteases identify enzyme residues important for substrate selection and catalytic rate. *Biochemistry,* 37(39), 13835– 13845.

191. Schock HB, Garsky VM and Kuo LC. (1996). Mutational anatomy of an HIV-1 protease variant conferring cross-resistance to protease inhibitors in clinical trials. *J. Biol. Chem.,* 271, 31957-31963.

192. Zhang YM, Imamichi H, Imamichi T, Lane HC, Falloon J, Vasudevachar, MB and Salzman NP. (1997). Drug resistance during indinavir therapy is caused by mutations in the protease gene and in its Gag substrate cleavage sites. *J. Virol.,* 71(9), 6662– 6670.

193. Doyon L, Croteau G, Thibeault D, Poulin F, Pilote L and Lamarre D. (1996). Second locus involved in human immunodeficiency virus type 1 resistance to protease inhibitors. *J. Virol.,* 70(6), 3763–3769.

194. Lamarre D, Croteau G, Pilote L, Rousseau P and Doyon L. (1994). Molecular characterization of HIV-1 variants resistant to specific vital protease inhibitors. *Proceedings of the Third International Workshop on HIV Drug Resistance.*

195. Larder BA, Bloor S, Hertogs K, Van den Eynde C and Pauwels R. (1998). HIV-1 Gag cleavage site changes are associated with specific protease mutations in plasma HIV-1 RNA but are not always retained in replication-competent, protease inhibitor-resistance recombinant viruses. Abstract 23 from the *2nd International Workshop on HIV Drug Resistance Treatment Strategies, Lake Maggiore, Italy.*

196. Mammano F, de la Carriere LC, Petit C and Clavel F. (1998). Multiple impacts of HIV resistance to protease inhibitors on Gag and reverse transcriptase function. Abstract 47 from the *2nd International Workshop on HIV Drug Resistance Treatment Strategies, Lake Maggiore, Italy.*

197. Nijhuis M, van Maarseveen NM, Lastere S, Schipper P, Coakley E, et al. (2007). A Novel Substrate-Based HIV-1 Protease Inhibitor Drug Resistance Mechanism. *PLoS Med.,* 4(1), e36.

198. Broglia R, 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(1), 60-66.

199. James RW. *The Optical Principles of the Diffraction of X-rays*. Bell, 1948.

200. Friedrich W, Knipping P and von Laue M. (1912). Interferenz-Erscheinungen bei Rontgenstrahlen. *Sitzungsberichte der Mathematisch-Physikalischen Classe der Königlich-Bayerischen Akademie der Wissenschaften zu München*., 303-322.

201. Bragg WL. (1912). The Specular Reflexion of X-rays. *Nature*, 90 (2250), 410.

202. Hauptman H. (1997). Phasing methods for protein crystallography. *Curr. Opin. Struct. Biol.,* 7 (5), 672.

203. Stout GH and Jensen LH. (1989). *X-Ray Structure Determination: A Practical Guide,* 2nd Ed. John Wiley and Sons Inc.

204. Helmut W. (2003). *Synchrotron Radiation*. Springer-Verlag, Berlin.

205. Ramaseshan S, Venkatesan K and Mani NV. (1957). The use of anomalous scattering for the determination of crystal structures- $KMnO_4$. *Proc. Ind. Acad. Sci.*, 46A, 95-111.

206. Moffat K. (2001). Time-Resolved Biochemical Crystallography: A Mechanistic Perspective. *Chem. Rev.*, 101(6), 1569-1582.

207. Riekel C. (2004). Recent developments in microdiffraction on protein crystals. *J. Syn. Rad.*, 11, 4–6.

208. Wallace BA and Janes RW. (2010). Synchrotron radiation circular dichroism (SRCD) spectroscopy: an enhanced method for examining protein conformations and protein interactions. *Biochem. Soc. Trans.,* 38(4), 861-73.

209. Ayvazyan V. et al. (2002). First operation of a free-electron laser generating GW power radiation at 32 nm wavelength. *The European Physical Journal D*, 20(1), 149-156.

210. Emma P. et al. (2010). First lasing and operation of an Å-wavelength free-electron laser. *Nature Photonics,* 4, 641–647.

211. Neutze R, Wout R, van der Spoel D, Weckert E and Hajdu J. (2000). Potential for biomolecular imaging with femtosecond X-ray pulses. *Nature*, 406, 752–757.

212. Jordan P. et al. (2001). Three-dimensional structure of cyanobacterial photosystem I at 2.5Å resolution. *Nature*, 411, 909–917.

213. Chapman HN. et.al. (2011). Femtosecond X-ray protein nanocrystallography. *Nature*, 470, 73-78.

214. Huang Z and Ruth RD. (1998). Laser-Electron Storage Ring. *Phys. Rev. Lett.,* 80, 976-979.

215. Abendroth J. et. al. (2010). X-Ray Structure determination of the Glycine Cleavage System Protein H of Mycobacterium tuberculosis using An Inverse Compton Synchrotron X-Ray Source. *J. Struct. Funct. Genomics,* 11(1), 91–100.

216. Miyahara J, Takahashi K, Amemiya Y, Kamiya N and Satow Y. (1986). A new type of X-ray area detector utilizing laser stimulated luminescence. *Nucl. Instrum. Methods Phys. Res. A,* 246, 572-578.

217. Gruner SM, Tate MW and Eikenberry EF. (2002). Charge-coupled device area x-ray detectors. *Rev. Sci. Instr.,* 73 (8), 2815-2842.

218. Amemiya Y. (1997). X-ray storage-phosphor imaging-plate detectors:high-sensitivity X-ray area detector. *Methods in Enzymology*, 276, 233-245.

219. Westbrook EM and Naday I. (1997). Charge-coupled device-based area detectors. *Methods in Enzymology*, 276, 244-268.

220. Eikenberry EF. et. al. (1991). X-ray detectors: comparison of film, storage phosphors and CCD detectors. *Photoelectronic Image Devices*, E.L. Morgan, ed., 273-280.

221. Henrich et. al. (2009). PILATUS: A single photon counting pixel detector for X-ray applications. *Nuc. Instr. and Meth. A,* 607, 247-249.

222. Kraft et. al.. (2009). Performance of single-photon-counting PILATUS detector modules. *J. Synchrotron Rad.,* 16, 368-375.

223. Bergamaschi A, Dinapoli R, Henrich B, Johnson I, Mozzanica A, Shi X and Schmitt B. (2011). Beyond single photon counting X-ray detectors. *Nuc. Instr. and Meth. A,* 628 (1), 238-241.

224. Blakeley MP. (2009). Neutron macromolecular crystallography. *Crystallography Reviews,* 15(3), 157-218.

225. Korszun ZR. (1997). Neutron Macromolecular Crystallography. *Methods in Enzymology*, 276, 218-232.

226. Niimura N and Podjarny A. (2011). *Neutron Protein Crystallography* in IUCr Monographs on Crystallography- 25, Oxford University Press Inc., pp. 124.

227. Shu F, Ramakrishnan V and Schoenborn BP. (2000). Enhanced visibility of hydrogen atoms by neutron crystallography on fully deuterated myoglobin. *Proc. Natl. Acad. Sci. USA,* 97(8), 3872-3877.

228. Wlodawer A and Hendrickson WA. (1982). A Procedure for Joint Refinement of Macromolecular Structures with X-ray and Neutron Diffraction Data from Single Crystals. *Acta Cryst*., A38, 239-247.

229. Dykstra MJ and Reuss LE. (2003). *Biological electron microscopy: theory, techniques, and troubleshooting*. 2nd ed.

230. Henderson R. (1995). The potential and limitations of neutrons, electrons and X-rays for atomic resolution microscopy of unstained biological molecules. *Q. Rev. Biophys*., 28, 171–193.

231. Rule GS and Hitchens TK. (2006). *Fundamentals of Protein NMR Spectroscopy*. Focus on Structural Biology, Vol. 5, XXVI, 530 p.

232. Dyson HJ and Wright PE. (2005). Intrinsically unstructured proteins and their functions. *Nature Rev. Mol. Cell Biol*., 6, 197–208.

233. Bertini I, Luchinat C, Parigi G, Ravera E and Reif B and Turano P. (2005). Solid-state NMR of proteins sedimented by ultracentrifugation. *Proc. Natl. Acad. Sci. U S A.,* 108(26), 10396-10399.

234. Koch MH, Vachette P and Svergun DI. (2003). Small-angle scattering: a view on the properties, structures and structural changes of biological macromolecules in solution. *Q. Rev. Biophys.,* 36 (2), 147–227.

235. Bernadó P and Blackledge M. (2010). Structural biology: Proteins in dynamic equilibrium. *Nature*, 468 (7327), 1046–1048.

236. Putnam CD, Hammel M, Hura GL and Tainer JA. (2007). X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Q. Rev. Biophys.,* 40 (3), 191–285.

237. Rambo RP and Tainer JA. (2010). Bridging the solution divide: comprehensive structural analyses of dynamic RNA, DNA, and protein assemblies by small angle X-ray scattering. *Curr. Opin. Struct. Biol.,* 20(1), 128–137.

238. Fujii K, Young MT and Harris KDM. (2011). Exploiting powder X-ray diffraction for direct structure determination in structural biology: The P2X4 receptor trafficking motif YEQGL. *J. Struct. Biol.,* 174(3), 461–467.

239. Von Dreele RB. (2003). Protein crystal structure analysis from high-resolution X-ray powder-diffraction data. *Methods in Enzymology*, 368, 254-67.

240. Margiolaki I and Wright JP. (2008). Powder crystallography on macromolecules. *Acta Cryst.,* A64 (1), 169-80.

241. Otwinowski Z and Minor W. (1997). Processing of X-ray diffraction data collected in oscillation mode. In: *Methods in Enzymology,* 276*,* (Carter CW, Jr. and Sweet RM, eds.), Academic Press, San Diego, CA, pp. 307–326.

242. Schutt C and Winkler FK. (1977). The Rotation method in crystallography. Arndt UW and Wonnacott. (Eds.). North Holland, Amsterdam.

243. Buerger MJ. (1964). The Precession method in X-ray crystallography. John Wiley and Sons.

244. Moffat K, Szebenyi D and Bilderback D. (1984). X-ray Laue Diffraction from Protein Crystals. *Science,* 223(4643), 1423-1425.

245. Darwin CG. (1914). The Theory of X-ray Reflection. *Phil. Mag.,* 27, 315-333; Darwin CG. (1914). The Theory of X-ray Reflection. Part II. *Phil. Mag.*, 27, 675-690.

246. Blundell TL and Johnson LN. (1976). *Protein Crystallography.* Equation 9.1. New York: Academic Press.

247. Drenth J. (1999). *Principles of Protein X-ray Crystallography*. Equation 4.31. Berlin: Springer-Verlag.

248. Holton JM and Frankel KA. (2010). The minimum crystal size needed for a complete diffraction data set. *Acta Cryst*., D66, 393–408.

249. Kabsch W. (1988). XDS. *Acta Cryst.,* D66, 125–132.

250. Kabsch W. (1988). Evaluation of single-crystal X-ray diffraction data from a Position-Sensitive Detector. *J. Appl. Cryst.,* 21, 916–924.

251. Kabsch W. (1988). Automatic Indexing of Rotation Diffraction Patterns. *J. Appl. Cryst.,* 21, 67-81.

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

253. Kabsch W. (2010). Integration, scaling, space-group assignment and post-refinement. *Acta Cryst.,* D66, 133–144.

254. Kabsch W. (1993). "Recent Extension of the Data- Processing Program XDS". Data Collection and Processing in Proceedings of the CCP4 Study Weekend, 29-30 January, 1993; Compiled by Sawyer L, Isaac N and Bailey S. pp. 63-70.

255. Leslie AG. **(**2006). The integration of macromolecular diffraction data. *Acta Cryst.,* D62(1), 48-57.

256. Battye TG, Kontogiannis L, Johnson O, Powell HR and Leslie AG. (2011). iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Cryst.,* D67(4), 271-281.

257. Powell HR. (1999). The Rossmann Fourier autoindexing algorithm in *MOSFLM. Acta Cryst.,* D55, 1690-1695

258. Steller I, Bolotovsky R and Rossmann MG. (1997). An algorithm for automatic indexing of oscillation images using Fourier analysis. *J. Appl. Cryst.,* 30, 1036-1040.

259. Rossmann MG and van Beek CG. (1999). Data processing. *Acta Cryst.,* D55, 1631-1653.

260. Rossmann MG. (1979). Processing oscillation diffraction data for very large unit cells with an automatic convolution technique and profile fitting. *J. Appl. Cryst.,* 12, 225-238.

261. Steller I, Bolotovsky R and Rossmann MG. (1998). The use of partial reflections for scaling and averaging Xray area-detector data. *J. Appl. Cryst.,* 31, 708-717.

262. Messerschmidt A and Pflugrath JW. (1987). Crystal Orientation and X-ray Pattern Prediction Routines for Area-Detector Diffraction Systems. *J. Appl. Cryst.,* 20, 306-315.

263. Pflugrath JW. (1999). The finer things in X-ray diffraction data collection. *Acta Cryst*., D55, 1718-1725.

264. Bricogne G. (1986). *Proceedings of the EEC Cooperative Workshop on Position-Sensitive Detector Software (Phase III)*, p. 28. Paris: LURE.Kim S. (1989). Auto-Indexing Oscillation Photographs. *J. Appl. Cryst.,* 22, 53-60.

265. Higashi TJ. (1990). Auto-Indexing of Oscillation Images. *J. Appl. Cryst.,* 23, 253-257.

266. Leslie A. (1993). "Autoindexing of rotation diffraction images and parameter refinement, Data Collection and Processing" in Proceedings of the CCP4 Study Weekend, 29-30 January, 1993; Compiled by Sawyer L, Isaac N and Bailey S. pp. 44-51.

267. Vriend G and Rossmann MG. (1987). Determination of the orientation of a randomly placed crystal from a single oscillation photograph. *J. Appl. Cryst.,* 20, 338-343.

268. Kabsch W. (1988). Evaluation of single-crystal X-ray diffraction data from a position-sensitive detector. *J. Appl. Cryst*., 21, 916-924.

269. Howard A. (1986). *Proceedings of the EEC Cooperative Workshop on Position-Sensitive Detector Software (Phases I and II)*, pp. 89–94 Paris: LURE.

270. Leslie AGW. (1987). "Profile Fitting" in Proceedings of the Daresbury Study Weekend at Daresbury Laboratory, 23-24 January, 1987; Compiled by Helliwell JR, Machin PA and Papiz MZ, pp. 39-50.

271. Rossmann MG and Erickson JW. (1983). Oscillation photography of radiation-sensitive crystals using a synchrotron source. *J. Appl. Cryst.,* 16, 629-636.

272. Diamond R. (1974). Profile Analysis in Single Crystal Diffractometry. *Acta Cryst.*, A25, 43-55.

273. Ford G. (1974). Intensity Determination by Profile Fitting Applied to Precessing Photograph. *J. Appl. Cryst.,* 7, 555-564.

274. Evans P. (2011). An introduction to data reduction: space-group determination, scaling and intensity statistics. *Acta Cryst*., D67(4), 282-292.

275. Evans P. (2006). Scaling and assessment of data quality. *Acta Cryst*., D62(1), 72-82.

276. Otwinowski Z, Borek D, Majewski W and Minor W. (2003). Multiparametric scaling of diffraction intensities. *Acta Cryst.,* A59(3), 228-234.

277. Rossmann MG. (1985). Determining the intensity of Bragg reflections from oscillation photographs. *Methods in Enzymology*, 114, 237-280.

278. Harrison SC, Winkler FK, Schutt CE, Durbin RM. (1985). Diffraction methods for biological macromolecules. Oscillation method with large unit cells. *Methods in Enzymology*, 114, 211-237.

279. Rossmann MG, Leslie AGW, Abdel- Meguid SS and Tsukihara T. (1979). Processing and postrefinement of oscillation camera data. *J. Appl. Cryst.,* 12, 570-581.

280. Evans PR. (1987). "Postrefinement of Oscillation Camera Data" in Proceedings of the Daresbury Study Weekend at Daresbury Laboratory, 23-24 January, 1987; Compiled by Helliwell JR, Machin PA and Papiz MZ., pp; 58-66.

281. Greenhough T. (1987). "Partials and Partiality" in Proceedings of the Daresbury Study Weekend at Daresbury Laboratory, 23-24 January, 1987; Compiled by Helliwell JR, Machin PA and Papiz MZ, pp. 51-57.

282. *International Tables for Crystallography* (1989). Vol. *A*, pp. 738–749. Dordrecht: Kluwer Academic Publishers.

283. Diederichs K and Karplus PA. (1997). Improved R-factors for diffraction data analysis in macromolecular crystallography. *Nature Struct. Biol.,* 4(4), 269-275.

284. Weiss MS. (2001). Global indicators of X-ray data quality. *J. Appl. Cryst.,* 34, 130–135.

285. Rossmann MG and Arnold E. (2001). *International Tables for Crystallography*, Vol. *F*, edited by M. G. Rossmann & E. Arnold, pp. 263–292. Dordrecht: Kluwer Academic Publishers.

286. Rossmann MG and Arnold E. (2006). Patterson and molecular-replacement techniques. *International Tables for Crystallography,* Vol. B, ch. 2.3, pp. 235-263.

287. Rossmann MG. (1990). The molecular replacement method. *Acta Cryst*., A46 (2), 73-82.

288. Hoppe W. (1957). Strukturanalyse über drei-und zweidimensionale Schnitte aus doppelten Pattersonfunktionen. *Acta Cryst.,* 10, 751.

289. Rossmann MG and Blow DM. (1962). The detection of sub-units within the crystallographic asymmetric unit. *Acta Cryst.,* 15, 24-31.

290. Brunger AT. (1990). Extension of molecular replacement: a new search strategy based on Patterson correlation refinement. *Acta Cryst.,* A46, 46-57.

291. Vagin A and Teplyakov A. (2000). An approach to multi-copy search in molecular replacement. *Acta Cryst.,* D56, 1622-1624.

292. Read R. (2001). Pushing the boundaries of molecular replacement with maximum likelihood. *Acta Cryst.,* D57, 1373-1382.

293. Chen YW. (2001). Solution solution: using NMR models for molecular replacement. *Acta Cryst.,* D57(10), 1457-1461.

294. Glykos NM and Kokkinidis M. (2001). Multidimensional molecular replacement. *Acta Cryst.,* D57, 1462-1473.

295. Kissinger CR, Gehlhaar DK, Smith BA and Bouzida D. (2001). Molecular replacement by evolutionary search. *Acta Cryst.,* D57, 1475-1479.

296. Chang G and Lewis M. (1997). Molecular Replacement Using Genetic Algorithms. *Acta Cryst.,* D53, 279-289.

297. Tickle IJ and Driessen HPC. (1996). Molecular replacement using known structural information. *Methods Mol. Biol.,* 56, 173-203.

298. Patterson AL. (1934). A Fourier series method for the determination of the components of interatomic distances in crystals. *Phys. Rev.,* 46, 372-376.

299. Patterson AL. (1935). A direct method for the determination of the components of interatomic distances in crystals. *Z. Kristallogr.,* 90, 517–542.

300. Crowther RA. (1972). The molecular replacement method. A collection of papers on the use of non-crystallographic symmetry (Ed, M.G., R.) Gordon and Breach, Science Publishers, Inc., New York, pp. 173-178.

301. Harada Y, Lifschitz A, Berthou J and Jolles P. (1981). A translation function combining packing and diffraction information: an application to lysozyme (high-temperature form). *Acta Cryst.,* A37, 398-406.

302. Navaza J. (1994). AMoRe: an automated package for molecular replacement. *Acta Cryst.,* A50, 157-163.

303. Navaza J and Saludjian P. (1997). AMoRe: an automated molecular replacement program package. *Methods in Enzymology,* 276, 581-594.

304. Grosse-Kunstleve RW and Adams PD. (2001). Patterson correlation methods: a review of molecular replacement with CNS. *Acta Cryst.,* D57, 1390-1396.

305. Vagin A and Teplyakov A. (1997). *MOLREP*: an automated program for molecular replacement. *J. Appl. Cryst.,* 30, 1022-1025.

306. Green DW, Ingram VM and Perutz MF. (1954). The structure of haemoglobin IV. Sign determination by the isomorphous replacement method. *Proc. Roy. Soc. (London),* A225, 287-307.

307. Perutz MF. (1956). Isomorphous replacement and phase detrmination in noncentrosymmetric space groups. *Acta Cryst.,* 9, 867-873.

308. Blow DM and Rossmann MG. (1961). The single isomorphous replacement method. *Acta Cryst*., 14, 1195-1202.

309. Harker D. (1956). The determination of the phases of the structure factors of noncentrosymmetric crystals by the method of double iomorphous replacement. *Acta Cryst*., 9, 1-9.

310. Hendrickson WA. (1991). Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. *Science,* 254**,** 5l-58.

311. Karle J. (1980). Some developments in anomalous dispersion for the structural investigation of macromolecular systems in biology. *Int. J. Quant. Chem.: Quant. Biol. Symp*., 7, 357-367.

312. Hendrickson WA, Smith JL and Sheriff S. (1985). Direct phase determination based on anomalous diffraction measurements. *Methods in Enzymology*, 115, 41-55.

313. Guss JM, Merritt EA, Phizackerley RP, Hedman B, Murata M, Hodgson KO and Freeman HC. (1989). Phase determination by multiple-wavelength X-ray diffraction: crystal structure of a basic blue copper protein from cucumbers. *Science,* 241 (4867), 806–811.

314. Smith JK. (1997). Multiwavelength anomalous diffraction in macromolecular crystallography in: Wilson, K.S., Davies, G., Ashton, A.W. and Bailey, S. (Ed.) Recent Advances in Phasing, Proceedings of the CCP4 Study Weekend, Daresbury Laboratory.

315. Pahler A, Smith JL and Hendrickson WA. (1990). A Probability Representation for Phase Information from Multiwavelength Anomalous Dispersion. *Acta Cryst.,* A46, 537–540.

316. Terwilliger TC. (1994). MAD Phasing: Bayesian Estimates of FA. *Acta Cryst.,* D50**,** 11–16.

317. de la Fortelle E and Bricogne G. (1997). Maximum-Likelihood Heavy-Atom Parameter Refinement for Multiple Isomorphous Replacement and Multiwavelength Anomalous Diffraction Methods. *Methods in Enzymology*, 276, 472–494.

318. Hendrickson WA and Ogata CM. (1997). Phase Determination from Multiwavelength Anomalous Diffraction Measurements. *Methods in Enzymology*, 276, 494–523.

319. Hendrickson WA, Horton JR and LeMaster DM. (1990). Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determinarion of three-dimensional strucrure. *EMBO J.,* 9, 1665-1672.

320. Turkenberg JP and Dodson EJ. (1996). Modern developments in molecular replacement. *Curr. Opin. Struct. Biol.,* 6, 604-610.

321. Navaza J and Vernoslova E. (1995). On the fast translation functions for molecular replacement. *Acta Cryst.,* A51, 445-449.

322. Tong L. (1996). Combined molecular replacement. *Acta Cryst.,* A52, 782–784.

323. Fisher RA. (1922). On the Mathematical Foundations of Theoretical Statistics. *Philosophical Transactions of the Royal Society,* A, 222, 309-368.

324. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC and Read RJ. (2007). Phaser Crystallographic Software. *J. Appl. Cryst*., 40, 658-674.

325. Sim GA. (1959). The distribution of phase angles for structures containing heavy atoms. *Acta Cryst.,* 12, 813-815.

326. Jaroszewski L, Li Z, Cai X-H, Weber C and Godzik A. (2011). FFAS server: novel features and applications. *Nucl. Acids Res.,* 39, W38-W44.

327. Jamrog DC, Zhang Y and Phillips GN. (2003). *SOMoRe*: a multi-dimensional search and optimization approach to molecular replacement. *Acta Cryst.,* D59, 304–314.

328. Doublie S. (2007). Production of selenomethionyl proteins in prokaryotic and eukaryotic expression systems. *Methods in Mol. Biol.,* 363, 91-108.

329. Wang B-C. (1985). Resolution of phase ambiguity in macromolecular crystallography. *Methods in Enzymology,* 115, 90-112.

330. Wang JW, Chen JR, Gu YX, Zheng CD, Jiang F, Fan HF, Terwilliger TC and Hao Q. (2004). SAD phasing by combination of direct methods with the SOLVE/RESOLVE procedure. *Acta Cryst.,* D60, 1244-1253.

331. McCoy AJ, Storoni LC and Read RJ. (2004). Simple algorithm for a maximum-likelihood SAD function. *Acta Cryst.,* D60,1220-1228.

332. Goulet A, Vestergaard G, Felisberto-Rodrigues C, Campanacci V, Garrett RA, Cambillau C and Ortiz-Lombardía M. (2010). Getting the best out of long-wavelength X-rays: de novo chlorine/sulfur SAD phasing of a structural protein from ATV. *Acta Cryst.,* D66, 304-308.

333. Sarma GN and Karplus PA. (2006). In-house sulfur SAD phasing: a case study of the effects of data quality and resolution cutoffs. *Acta Cryst.,* D62, 707-716.

334. Sayre D. (1952). The squaring method: a new method for phase determination. *Acta Cryst*., 5, 60-65.

335. Karle J and Hauptman H. (1956). Structure invariants and semiinvariants for no-centrosymmetric space groups. *Acta Cryst.,* 9, 635-651.

336. Karle J and Karle IL. (1966). The symbolic addition procedure for phase determination for centrosymmetric and non-centrosymmetric crystals. *Acta Cryst.,* 21, 849-859.

337. Karle J. (1989). Direct methods in protein crystallography. A*cta Cryst.,* A45, 765-781.

338. Detitta GT, Weeks CM, Thuman P, Miller R and Hauptman HA. (1994). Structure solution by minimal-function phase refinement and fourier filtering. I. theoretical basis. *Acta Cryst.*, A50, 203-210.

339. Weeks CM, Detitta GT, Miller R and Hauptman HA. (1993). Applications of the minimal principle to peptide structures. *Acta Cryst.,* D49, 179-181.

340. Miller R, Gallo SM, Khalak HG and Weeks CM. (1994). SnB: Crystal structure determination via Shake-and-Bake. *J. Appl. Cryst.,* 27, 613-621.

341. Smith GD, Nagar B, Rini JM, Hauptman HA and Blessing RH. (1998). The use of Snb to determine an anomalous scattering substructure. *Acta Cryst., D*54, 799–804.

342. Frazao C, Sieker L, Sheldrick GM, Lamzin V, LeGall J and Carrondo MA. (1999). Ab initio structure solution of a dimeric cytochro, 346me c3 from Desulfovibrio gigas containing disulfide bridges. *J. Biol. Inorg. Chem.,* 4, 162-165.

343. Mathiesen RH and Mo F. (1998). Application of known triplet phases in the crystallographic study of bovine pancreatic trypsin inhibitor. II: study at 2.0 Å resolution. *Acta Cryst.,* D54, 237-42.

344. Xu H, Hauptman HA and Weeks CM. (2002). Sine-enhanced Shake-and-Bake: the theoretical basis for applications to Se-atom substructures. *Acta Cryst.,* D58, 90-96.

345. Sheldrick GM. (1998). In: S. Fortier, Editor, *Direct Methods for Solving Macromolecular Structures*, Kluwer Academic Publishers, Dordrecht, The Netherlands, pp. 401-411

346. Dall'Antonia F, Baker PJ and Schneider TR. (2003). Optimization of selenium substructures as obtained from SHELXD. *Acta Cryst.,* D59, 1987-1994.

347. Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, et al. (1998). Crystallography and NMR System: a new software suite for macromolecular structure determination. *Acta Cryst.,* D54, 905–921.

348. Otwinowski Z. (1991). In Wolf, W., Evans, P.R. and Leslie, A.G.W. (ed.) *Isomorphous replacement and anomalous scattering*, Daresbury Laboratory, Warrington: pp. 80.

349. de la Fortelle E and Bricogne G. (1997). Maximum likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multivavelength anomalous diffraction methods. In Carter, C.W. and Sweet, R.M. (ed.). *Methods in Enzymology,* 276, Academic Press, Orlando, Fl: pp. 472-494.

350. Terwilliger TC and Berendzen J. (1999). Automated MAD and MIR structure solution. *Acta Cryst.,* D55, 849-861.

351. Grosse-Kunstleve RW and Adams PD. (2003). Substructure search procedures for macromolecular structures. *Acta Cryst.,* D59**,** 1966–1973.

352. Burla MC, Giacovazzo C and Polidori G. (2011). From a random to the correct structure: the VLD algorithm. *J. Appl. Cryst*., 43, 825-836.

353. Burla MC, Carrozzini B, Cascarano GL, Giacovazzo C and Polidori G. (2011). Phasing medium-size structures and proteins by the VLD algorithm. *J. Appl. Cryst.,* 193-199.

354. Burla MC, Caliandro R, Giacovazzo C and Polidori G. (2010). The difference electron density: a probabilistic reformulation. *Acta Cryst.,* A66, 347-361.

355. Burla MC, Carrozzini B, Cascarano GL, Giacovazzo C and Polidori G. (2011). About the hybrid Fourier syntheses: a probabilistic approach. *Acta Cryst.,* A67, 447-455.

356. Giacovazzo C and Mazzone A. (2011). Variance of electron-density maps in space group P1. *Acta Cryst.,* A67, 210-218.

357. Giacovazzo C and Mazzone A and Comunale G. (2011). Estimation of the variance in any point of an electron-density map for any space group. *Acta Cryst.,* A67, 368-382.

358. Burla MC, Caliandro R, Camalli M, Carrozzini B, Cascarano GL, De Caro L, Giacovazzo C and Polidori G, Siliqi D and Spagna R. (2007). *J. Appl. Cryst.,* 40*,* 609-613.

359. Cowtan KD and Zhang KY. (1999). Density modification for macromolecular phase improvement. *Prog. Biophys. Mol. Biol.,* 72, 245-70.

360. Terwilliger TC. (2000). Maximum likelihood density modification. *Acta Cryst*., D56, 965-972.

361. Cowtan K. (2010). Recent developments in classicl density modification. *Acta Cryst.,* D66, 470–478.

362. Kleywegt GJ. (1996). Use of Non-crystallographic symmetry in protein structure refinement. *Acta Cryst.,* D52**,** 842-857.

363. Leslie AGW. (1987). A reciprocal-space method for calculating a molecular envelope using the algorithm of B. C. Wang. *Acta Cryst.,* A43, 134-136.

364. Matthews B. (1968). Solvent content of protein crystal. *J. Mol. Biol.,* 33, 491-497.

365. Zhang KYJ and Main P. (1990). Histogram matching as a new density modification technique for phase refinement and extension of protein molecules. *Acta Cryst.,* A46, 41-46.

366. Jones TA, Zou JY, Cowan SW and Kjeldgaard M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Cryst.,* A47, 110-119.

367. Emsley P and Cowtan K. (2004). Coot: model-building tools for molecular graphics. *Acta Cryst.,* D60, 2126-2132.

368. Read RJ. (1986). Improved Fourier coefficients for maps using phases from partial structures with error. *Acta Cryst.,* A42, 140-149.

369. Bhat TN. (1988). Calculation of an omit map. *J. Appl. Cryst*., 21, 279 – 281.

370. Brunger AT and Rice LM. (1997). Crystallographic refinement by simulated annealing: methods and applications. *Methods in Enzymology*, 277, 266-267.

371. Read RJ. (1997). Model phases: probabilities and bias. *Methods in Enzymology*, 277, 110-128.

372. McCoy AJ. (2004). Liking likelihood. *Acta Cryst.,* D60, 2169-2183.

373. Brooks B, Bruccoleri R, Olafson B, States D, Swaminathan S and Karplus M. (1983). CHARMM: A Program for Macromolecular Energy, Minimization, and Molecular Dynamics Calculations. *J. Comp. Chem.,* 4, 187-217.

374. Brunger AT, Kuriyan J and Karplus M. (1987). Crystallographic R factor refinement by molecular dynamics. *Science,* 235, 458–460.

375. Adams PD, Grosse-Kunstleve RW, Hung L-W, Ioerger TR, McCoy AJ, Moriarty NW, Read RJ, Sacchettini JC, Sauter NK and Terwilliger TC. (2002). PHENIX: building new software for automated crystallographic structure determination. *Acta Cryst*., D58, 1948-1954.

376. Collaborative Computational Project, Number 4. (1994). The CCP4 suite: programs for protein crystallography. *Acta Cryst*., D50, 760-763.

377. Tronrud DE. (1997). The TNT refinement package. *Methods in Enzymology*, 277, 306–319.

378. Sheldrick GM and Schneider TR. (1997). SHELXL: high-resolution refinement. *Methods in Enzymology,* 277, 319–344.

379. Engh RA and Huber R. (1991). Accurate bond and angle parameters for X-ray protein structure refinement. *Acta Cryst.*, A47, 392–400.

380. Parkinson G, Vojtechovsky J, Clowney L, Brunger AT and Berman HM. (1996). New parameters for the refinement of nucleic acid-containing structures. *Acta Cryst.,* D52, 57–64.

381. Hendrickson WA and Konnert JH. (1980). Incorporation of stereochemical information into crystallographic refinement. *Computing in Crystallography,* Diamond R, Ramaseshan S and Venkatesan K, eds. (Bangalore, India: Indian Academy of Science), 13.01–13.25.

382. Hendrickson WA. (1985). Stereochemically restrained refinement of macromolecular structures. *Methods in Enzymology,* 115, 252–270.

383. Sussman JL, Holbrook SR, Church GM and Kim S-H. (1977). A Structure-Factor Least-Squares Refinement Procedure for Macromolecular Structure using Constrained and Restrained Parameters. *Acta Cryst.,* A33, 800-804.

384. Bricogne G. (1997). Bayesian statistical viewpoint on structure determination: basic concepts and example. *Methods in Enzymology*, 276, 361-423.

385. Murshudov GN, Vagin AA and Dodson EJ. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Cryst.,* D53, 240–255.

386. Silva AM and Rossmann MG. (1985). The refinement of southern bean mosaic virus in reciprocal space. *Acta Cryst.,* B41, 147-157.

387. Pannu NS and Read RJ. (1996). Improved structure refinement through maximum likelihood. *Acta Cryst.,* A52, 659-668.

388. Pannu NS, Murshudov GN, Dodson EJ and Read RJ. (1998). Incorporation of prior phase information strengthens maximum likelihood structural refinement. *Acta Cryst.,* D54, 1285-94.

389. Adams PD, Pannu NS, Read RJ and Brünger AT. (1999). Extending the limits of molecular replacement through combined simulated annealing and maximum likelihood refinement. *Acta Cryst.,* D55, 181-90.

390. Kleywegt GJ and Brünger AT. (1996). Cross-validation in crystallography: practice and applications. *Structure*, 4, 897-904.

391. Adams PD, Pannu NS, Read RJ and Brünger AT. (1997). Cross-validated maximum likelihood enhances crystallographic simulated annealing refinement. *Proc. Natl. Acad. Sci. USA*, 94, 5018-5023.

392. Brünger AT. (1997). Free R value: cross-validation in crystallography. *Methods in Enzymology*, 277, 366-396.

393. Brünger AT. (1992). The free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature*, 355, 472-474.

394. Afonine PV, Grosse-Kunstleve RW, Urzhumtsev A and Adams PD. (2009). Automatic multiple-zone rigid-body refinement with a large convergence radius. *J. Appl. Cryst.,* 42, 607-615.

395. Kirkpatrick S, Gelatt CD and Vecchi MP Jr. (1983). Optimization by simulated annealing. *Science*, 220, 671-680.

396. Laarhoven PJM and Aarts EHL. (Eds) (1987). *Simulated Annealing: Theory and Applications*. Dordrecht: Reidel Publishing Company.

397. Brünger AT, Krukowski A and Erickson J. (1990). Slow-cooling protocols for crystallographic refinement by simulated annealing. *Acta Cryst.,* A46, 585-593.

398. Brünger AT. (1988). Crystallographic refinement by simulated annealing: application to a 2.8 Å resolution structure of aspartate aminotransferase. *J. Mol. Biol.*, 203, 803-816.

399. Kuriyan J, Brünger AT, Karplus M and Hendrickson WA. (1989). X-ray refinement of protein structures by simulated annealing: test of the method on myohemerythrin. *Acta Cryst.,* A45, 396-409.

400. Brünger AT, Adams PD and Rice LM. (1998). Recent developments for the efficient crystallographic refinement of macromolecular structures. *Curr. Opin. Struct. Biol.,* 8, 606-611.

401. Fujinaga M, Gros P and van Gunsteren WF. (1989). Testing the method of crystallographic refinement using molecular dynamics. *J. Appl. Cryst.,* 22, 1-8.

402. Rice LM and Brünger AT. (1994). Torsion angle dynamics: reduced variable conformational sampling enhances crystallographic structure refinement. *Proteins*, 19, 277-290.

403. Afonine PV, Grosse-Kunstleve RW, Urzhumtsev A and Adams PD. (2005). A robust bulk-solvent correction and anisotropic scaling procedure. *Acta Cryst.,* D61, 850–855.

404. Tronrud DE. (2004). Introduction to macromolecular refinement. *Acta Cryst*., D60, 2156-2168.

405. Rice LM, Shamoo Y and Brünger AT. (1998). Phase Improvement by Multi-Start Simulated Annealing Refinement and Structure-Factor Averaging. *J. Appl. Cryst.,* 31, 798-805.

406. Perrakis A, Sixma TK, Wilson KS and Lamzin VS. (1997). wARP: improvement and extension of crystallographic phases by weighted averaging of multiple-refined dummy atomic model. *Acta Cryst.,* D53, 448-455.

407. Schomaker V and Trueblood KN. (1968). On the rigid-body motion of molecules in crystals. *Acta Cryst.*, B24, 63.

408. Howlin B, Butler SA, Moss DS, Harris GW and Driessen HPC. (1993). TLSANL: TLS parameter analysis program for segmented anisotropic refinement of macromolecular structures. *J. Appl. Cryst*., 26, 622-624.

409. Winn MD, Isupov MN and Murshudov GN. (2001). Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Cryst.*, D57, 122-133.

410. Luzzati V. (1952). Traitement statistique des erreur dans la determination des structures crystallines. *Acta Cryst.,* 5, 802-810.

411. Laskowski RA, MacArthur MW, Moss DS and Thornton JM. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.,* 26, 283-291.

412. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN and Bourne PE. (2000). The Protein Data Bank. *Nucleic Acids Res.,* 28, 235-242.

413. Taylor R. (2002). Life science applications of the Cambridge Structural Database. *Acta Cryst.,* D58, 879-888.

414. Ramachandran GN and Sasisekharan V. (1968). Conformation of polypeptides and proteins. *Adv. Protein Chem.,* 23, 283-438.

415. DeLano WL. (2002). The PyMOL Molecular Graphics System, DeLano Scientific, San Carlos, CA.

416. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC and Ferrin TE. (2004). UCSF Chimera - a visualization system for exploratory research and analysis. *J. Comput. Chem.,* 25(13), 1605-1612.

417. Levitt DG. (2001). A new software routine that automates the fitting of protein X-ray crystallographic electron-density maps. *Acta Cryst.,* D57, 1013-1019.

418. Terwilliger TC. (2003). Automated main-chain model building by template matching and iterative fragment extension. *Acta Cryst.,* D59, 38-44.

419. Morris RJ, Perrakis A and Lamzin VS. (2002). *ARP/wARP*'s model-building algorithms. I. The main chain. *Acta Cryst.*, D58, 968-975.

420. Lamzin VS and Wilson K. (1997). Automated refinement for protein crystallography. *Methods in Enzymology*. (Carter C and Sweet B. eds.), 277, 269-305.

421. Cruickshank DWJ. (1999). Remarks about protein structure precision. *Acta Cryst.* D55, 583-601.

422. Stroud RM. and Fauman EB. (1995). Significance of structural changes in proteins: expected errors in refined protein structures. *Protein Science*, 4, 2392-2404.

423. Kumar M, Kannan KK, Hosur MV, Bhavesh NS, Chatterjee A, Mittal R and Hosur RV. (2002). *Biochem. Biophys. Res. Comm.*, 294, 395-401.

424. Pillai B, Bhat SV, Kannan KK and Hosur MV. (2004). Rapid screening for HIV-1 protease inhibitor leads through X-ray diffraction. *Acta Cryst*., D60, 594-596.

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

426. Tomaszek TA Jr, Magaard VW, Bryan HG, Moore ML and Meek TD. (1990). Chromophoric peptide substrates for the spectrophotometric assay of HIV-1 protease. *Biochem. Biophys. Res. Comm.*, 168(1), 274-280.

427. McPherson A. (2004). Introduction to protein crystallization. *Methods,* 34, 254–265.

428. Benvenuti M and Mangani S. (2007). Crystallization of soluble proteins in vapor diffusion for x-ray crystallography. *Nature Protocols,* 2, 1633 – 1651.

429. Asherie N. (2004). Protein crystallization and phase diagrams. *Methods,* 34, 266– 272.

430. McPherson A. (2004). Protein crystallization in the structural genomics era. *J. Struct Funct. Genomics,* 5, 3–12.

431. Vera L, Czarny B, Georgiadis D, Dive V and Stura EA. (2011). Practical Use of Glycerol in Protein Crystallization. *Cryst. Growth Des.,* 11, 2755–2762.

432. Pflugrath JW. (2004). Macromolecular cryocrystallography--methods for cooling and mounting protein crystals at cryogenic temperatures. *Methods*, 34(3), 415-423.

433. Eizert H, Bagossi P, Sperka T, Miklossy G, Bander P, Boross P, Weber IT and Tozer J. (2008). Amino acid preferences of retroviral proteases for amino terminal positions in a type-1 cleavage site. *J. Virol.,* 82, 10111-10117.

434. Massover WH. (2007). Radiation damage to protein specimens from electron beam imaging and diffraction: a mini- review of anti-damage approaches, with special reference to synchrotron X-ray crystallography. *J. Synchrotron Rad*., 14, 116–127.

435. Garman E and Owen RL. (2007). Cryocrystallography of macromolecules: practice and optimization. *Methods Mol. Biol.,* 364, 1-18.

436. Owen RL, Rudino-Pinera E and Garman EF. (2006). Experimental determination of the radiation dose limit for cryocooled protein crystals. *Proc. Natl Acad. Sci. USA*, 103, 4912–4917.

437. Watenpaugh KD. (1990). Macromolecular Crystallography at Cryogenic Temperatures. *Curr. Opin. Struct. Biol.,* 1, 1012-1015.

438. Hope H. (1988). Cryo-crystallography of biological macromolecules; a generally applicable method. *Acta Cryst.,* B44, 22-26.

439. Roth M, Carpentier P, Kaikati O, Joly J, Charrault P, Pirocchi M, Kahn R, Fanchon E, Jacquamet L, Borel F, Bertoni A, Israel-Gouy P and Ferrer J-L. (2002). FIP: a highly automated beamline for multiwavelength anomalous diffraction experiments. *Acta Cryst.,* D58, 805-814.

440. Brünger AT, Adams PD, Clore GM, DeLano WL, Gross P, Grosse-Kunstleve RW, J-S Jiang, Pannu NS, Read RJ, Rice LM and Simonson T. (2001). *International Tables for Crystallography*, *Crystallography of Biological Macromolecules*; Dordrecht: Kluwer Academic Publishers, The Netherlands; Volume F, pp. 710-716.

441. Schüttelkopf AW and van Aalten DMF. (2004). PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Acta Cryst*., D60, 1355-1363.

442. Krissinel E and Henrick K. (2004). Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Cryst.,* D60, 2256-2268.

443. Cleland WW. (2000). Low-barrier hydrogen bonds and enzymatic catalysis. *Arch. Biochem. Biophys.,* 382(1), 1-5.

444. Katz BA, Spencer JR, Elrod K, Luong C, Mackman RL, Rice M, Sprengeler PA, Allen D and Janc J. (2002) Contribution of multicentered short hydrogen bond arrays to potency of active site-directed serine protease inhibitors. *J. Am. Chem. Soc.,* 124, 11657-11668.

445. Hooft RWW, Sander C and Vriend G. (1996). Positioning hydrogen atoms by optimizing hydrogen-bond networks in protein structures. *Proteins: Structure, Function, and Genetics,* 26, 363-376.

446. Li Y, Roy A and Zhang Y. (2009). HAAD: A quick algorithm for accurate prediction of hydrogen atoms in protein structures. *PLoS ONE,* 4, e6701.

447. Prashar V and Hosur MV. (2004). 1.8Å X-ray structure of C95M/C1095F double mutant of tethered HIV-1 protease dimer complexed with acetyl-pepstatin. *Biochem. Biophys. Res. Commun*., 323(4), 1229-1235.

448. Harris TK and Turner GJ. (2002). Structural basis of perturbed pKa values of catalytic groups in enzyme active sites. *IUBMB Life*, 53, 85-98.

449. Trylska J, Bała P, Geller M and Grochulski P. (2002). Molecular dynamics simulations of the first steps of the reaction catalyzed by HIV-1 protease. *Biophysical Journal*, 83, 794–807.

450. Smith R, Brereton I, M, Chai RY and Kent SBH. (1996). Ionization states of the catalytic residues in HIV-1 protease. *Nature Str. Biol.*, 3(11), 946-950.

451. Wang YX, Freedberg DI, Yamazaki T, Wingfield PT, Stahl SJ, Kaufman JD, Kiso Y and Torchia DA. (1996). Solution NMR evidence that the HIV-1 protease catalytic aspartyl groups have different ionization states in the complex formed with the asymmetric drug KNI-272. *Biochemistry,* 35, 9945–9950.

452. Ichikawa M. (1979). The C-O vs O-H length correlation in hydrogen-bonded carboxyl groups. *Acta Cryst.,* B35, 1300-1301.

453. Perrin CL and Nielson JB. (1997). "Strong" hydrogen bonds in chemistry and biology. *Annu. Rev. Phys. Chem*., 48, 511-544.

454. Jeffrey GA. *An Introduction to Hydrogen Bonding.* 1997. Oxford University Press.

455. Fuhrmann CN, Daugherty MD and Agard DA. (2006). Subangstrom crystallography reveals that short ionic hydrogen bonds, and not a His-Asp low-barrier hydrogen bond, stabilize the transition state in serine protease catalysis. *J. Am. Chem. Soc.,*128(28), 9086-9102.

456. Guthrie JP. (1996). Short strong hydrogen bonds: can they explain enzymatic catalysis? *Chem, Biol.,* 3, 163–170.

457. Langkilde A, Kristensen SM, Leggio LL, Mølgaard A, Jensen JH, Houk AR, Navarro J-C, Kauppinen PS and Larsen S. (2008). Short strong hydrogen bonds in proteins: a case study of rhamnogalacturonan acetylesterase. *Acta Cryst.,* D64, 851-863.

458. Stranz GR, Gruber K, Steinkellner G, Schwab H and Kratky C. (2004). Observation of a Short, Strong Hydrogen Bond in the Active Site of Hydroxynitrile Lyase from *Hevea brasiliensis* explains a Large pK_a Shift of the Catalytic Base Induced by the Reaction Intermediate. *J. Biol. Chem*., 279, 3699-3707.

459. Carnevale V, Raugei S, Piana S and Carloni P. (2008). On the nature of the reaction intermediate in the HIV-1 protease: a quantum chemical study. *Comp. Phy. Comm.,* 179, 120-123.

460. Das A, Prashar V, Mahale S, Serre L, Ferrer J-L and Hosur MV. (2006) .Crystal structure of HIV-1 *in situ* product complex and observation of a low-barrier hydrogen bond between catalytic aspartates. *Proc. Natl. Acad. Sci. USA*., 103(49), 18464-18469.

461. Manikem MW and Fink AL. (1977). Reactivity and cryoenzymology of enzymes in the crystalline state. *Annu. Rev. Biophys. Bioeng*., 6, 301-343.

462. Piana S, Bucher D, Carloni P and Rothlisberger U. (2004). Reaction Mechanism of HIV-1 Protease by Hybrid Car-Parrinello MD/Classical MD Simulations. *J. Phys. Chem.,* 108, 11139–11149.

463. Jacques F. (1983). On the stability of tetrahedral intermediates within the active sites of serine and cysteine proteases. *Eur. J. Biochem.*, 135, 339-341.

464. Hajdu J, Machin PA, Campbell, JW, Greenhough TJ, Clifton IJ, Zurek S, Gover S, Johnson LN and Elder M. (1987). Millisecond X-ray diffraction and the first electron density map from Laue photographs of a protein crystal. *Nature*, 329, 178-181.

465. Singer PT, Smalas A, Carty RP, Mangel WF and Sweet RM. (1993). The hydrolytic water molecule in trypsin, revealed by time-resolved Laue crystallography. *Science*, 259 (5095), 669-673.

466. Yu N, Hayik SA, Wang B, Liao N, Reynolds CH Jr. and Merz KM. (2006). Assigning the protonation states of the key aspartates in β-Secretase using QM/MM X-ray structure refinement. *J. Chem. Theory Comput.*, 2(4), 1057 – 1069.

467. Louis JM, Weber IT, Tozser J, Clore GM. and Gronenborn AM. (2000). HIV-1 protease: maturation, enzyme specificity, and drug resistance. *Adv. Pharmacol.,* 49, 111–146.

468. Louis JM, Ishima R, Torchia DA and Weber IT. (2007). HIV-1 protease: structure, dynamics and inhibition. *Adv. Pharmacol.,* 55, 261–298.

469. Ishima R, Torchia DA, Lynch SM, Gronenborn AM and Louis JM. (2003). Solution structure of the mature HIV-1 protease monomer: insight into the tertiary fold and stability of a precursor. *J. Biol. Chem.,* 278, 43311–43319.

470. Miller M, et. al. (1989). Conserved folding in retroviral proteases: crystal structure of a synthetic HIV-1 protease. *Science*, 246, 1149–1152.

471. Louis JM, Nashed NT, Parris KD, Kimmel AR and Jerina DM. (1994). Kinetics and mechanism of autoprocessing of human immunodeficiency virus type 1 protease from an analog of the Gag-Pol polyprotein. *Proc. Natl Acad. Sci. USA*, 91, 7970–7974.

472. Co E, et. al. (1994). Proteolytic processing mechanisms of a mini precursor of the aspartic protease of human immunodeficiency virus type 1. *Biochemistry,* 33, 1248–1254.

473. Louis JM, Wondrak EM, Kimmel AR, Wingfield PT and Nashed NT. (1999). Proteolytic processing of HIV-1 protease precursor, kinetics and mechanism. *J. Biol. Chem.,* 274, 23437–23442.

474. Louis JM, Clore GM and Gronenborn AM. (1999). Autoprocessing of HIV-1 protease is tightly coupled to protein folding. *Nature Struct. Biol.,* 6, 868–874.

475. Pettit SC, Everitt LE, Choudhury S, Dunn BM and Kaplan AH. (2004). Initial cleavage of the human immunodeficiency virus type 1 GagPol precursor by its activated protease occurs by an intramolecular mechanism. *J. Virol.,* 78, 8477–8485.

476. Tessmer U and Krausslich H-G. (1998). Cleavage of human immunodeficiency virus type 1 proteinase from the N-terminally adjacent p6* protein is essential for efficient Gag polyprotein processing and viral infectivity. *J. Virol.,* 72, 3459–3463.

477. Ludwig C, Leiherer A and Wagner G. (2008). Importance of protease cleavage sites within and flanking human immunodeficiency virus type 1 transframe protein p6* for spatiotemporal regulation of protease activation. *J. Virol.,* 82, 4573–4584.

478. Wondrak EM, Nashed NT, Haber MT, Jerina DM and Louis JM. (1996). A transient precursor of the HIV-1 protease: isolation, characterization and kinetics of maturation. *J. Biol. Chem*., 271, 4477–4481.

479. Cherry E, et. al. (1998). Characterization of human immunodeficiency virus type-1 (HIV-1) particles that express protease-reverse transcriptase fusion proteins. *J. Mol. Biol*., 284, 43–56.

480. Bihani S, Das A, Prashar V, Ferrer J-L and Hosur MV. (2009). X-ray structure of HIV-1 protease in situ product complex. *Proteins: Structure, Function, and Bioinformatics*, 74 (3), 594-602.

481. Kozal MJ, Shah N, Shen N, Yang R, Fucini R, Merigan TC, Richman DD, Morris D, Hubbell E, Chee M and Gingeras TR. (1996). Extensive polymorphisms observed in HIV-1 clade B protease gene using high-density oligonucleotide arrays. *Nature Medicine*, 2, 753-759.

482. Stewart DE, Sarkar A and Wampler JE. (1990). Occurrence and role of *cis* peptide bonds in protein structures. *J. Mol. Biol*., 214, 253–260.

483. MacArthur MW and Thornton JM. (1991). Influence of proline residues on protein conformation. *J. Mol. Biol.*, 218: 397–412.

484. Weiss MS., Jabs A and Hilgenfeld R. (1998). Peptide bonds revisited. *Nature Struct. Biol.,* 5, 676.

485. Radzicka A, Pedersen L and Wolfenden R. (1988). Influences of solvent water on protein folding: free energies of solvation of cis and trans peptides are nearly identical. *Biochemistry*, 27, 4538-41.

486. Drakenberg T and Forsén S. (1971). The barrier to internal rotation in monosubstituted amides. *J. Chem. Soc. D: Chem. Comm.*, 1404-1405.

487. Grathwohl C and Wüthrich K. (1976). Nmr studies of the molecular conformations in the linear oligopeptides H-(L-Ala)$_n$-L-Pro-OH. *Biopolymers*, 15, 2043-2057.

488. Herzberg O and Moul J. (1991). Analysis of the steric strain in the polypeptide backbone of protein molecules. *Proteins: Struct. Fund. Genet.,* 11, 223–229.

489. Stoddard BL and Pietrokovski S. (1998). Breaking up is hard to do. *Nature Struct. Biol.,* 5, 3-5.

490. Weiss MS and Hilgenfeld R. (1999). A Method to Detect Nonproline cis Peptide Bonds in Proteins. Biopolymers, 50: 536-44.

491. Jabs A, Weiss MS and Hilgenfeld R. (1999). Nonproline cis peptide bonds in proteins. *J. Mol. Biol*. 286, 291–304.

492. Bouckaert J, Dewallef Y, Poortmans F, Wyns L and Loris R. (2000). The structural features of concanavalin A governing non-proline peptide isomerization. *J. Biol. Chem*., 275, 19778–19787.

493. Heroux A, White EL, Ross LJ, Davis RL and Borhani DW. (1999). Crystal structures of the Toxoplasma gondii hypoxanthine-guanine phosphoribosyltransferase- GMP and -IMP complexes: comparison of purine binding interactions with the XMP complex. *Biochemistry*, 38, 14485–14494.

494. Heroux A, White EL, Ross LJ, Davis RL and Borhani DW. (1999). Crystal structure of Toxoplasma gondii hypoxanthine-guanine phosphoribosyltransferase with XMP, pyrophosphate, and two Mg (2C) ions bound: insights into the catalytic mechanism. *Biochemistry*, 38, 14495–14506.

495. Ludwig ML, Pattridge KA, Metzger AL, Dixon MM, Eren M, Feng Y and Swenson RP. (1997). Control of oxidation-reduction potentials in flavodoxin from Clostridium beijerinckii: the role of conformation changes. *Biochemistry*, 36, 1259–1280.

496. Deacon A, Gleichmann T, Kalb (Gilboa) AJ, Price H, Raftery J, Bradbrook G, Yariv J and Helliwell JR. (1997). The structure of concanavalin A and its bound solvent determined with small-molecule accuracy at 0.94Å resolution. *J. Chem. Soc., Faraday Trans.,* 93, 4305-4312

497. Leesong M., Henderson BS, Gillig JR, Schwab JM and Smith JL. (1996) Structure of a dehydratase-isomerase from the bacterial pathway for biosynthesis of unsaturated fatty acids: two catalytic activities in one active site. Structure, 4:253-64.

498. Trylska J, Tozzini V, Chang CA and McCammon JA. (2007). HIV-1 Protease Substrate Binding and Product Release Pathways Explored with Coarse-Grained Molecular Dynamics. *Biophysical Journal*, 92, 4179-4187

499. Tozer J, Blaha I and Oroszlan S. (1991). Comparison of the HIV-1 and HIV-2 proteinases using oligopeptide substrates representing cleavage sites in Gag and Gag-Pol polyproteins. *FEBS Letters*, 281, 77-80.

500. Nalam MN, Ali A, Altman MD, Reddy GS, Chellappan S, Kairys V, Ozen A, Cao H, Gilson MK, Tidor B, Rana TM and Schiffer CA. (2010). Evaluating the substrate-envelope hypothesis:

structural analysis of novel HIV-1 protease inhibitors designed to be robust against drug resistance. *J. Virol.*, 84(10), 5368-5378.

501. Niimura N and Podjarny A. (2011). *Neutron Protein Crystallography* in IUCr Monographs on Crystallography- 25, Oxford University Press Inc., pp. 124.

502. Schmitt E, Moulinier L, Fujiwara S, Imanaka T, Thierry J-C and Moras D. (1998). Crystal structure of aspartyl-tRNA synthetase from *Pyrococcus kodakaraensis* KOD: archaeon specificity and catalytic mechanism of adenylate formation. *The EMBO Journal*, 17, 5227 – 5237.

503. Koen HG, Verschueren FS, Rozeboom HJ, Kalk KH and Dijkstra BW. (1993). Crystallographic analysis of the catalytic mechanism of haloalkane dehalogenase. *Nature*, 363, 693-698.

504. Liepinsh E and Otting G. (1996). Proton exchange rates from amino acid side chains—implications for image contrast. *Magnetic Resonance in Medicine*, 35, 30-42.

505. Zheng J, Balendu SA, Chingkuang T, McKenna R and Silverman DN. (2008). Role of Hydrophilic Residues in Proton Transfer during Catalysis by Human Carbonic Anhydrase II. *Biochemistry*, 47, 12028–12036

506. Maupin CM. (2009). Elucidation of the proton transport mechanism in human carbonic anhydrase II. *J. Am. Chem. Soc.,* 131, 7598-7608.

507. Blakeley MP, Kalb (Gilboa) AJ, Helliwell JR and Myles DAA. (2004). The 15-K neutron structure of saccharide-free concanavalin A. *Proc. Natl. Acad. Sci. USA*, 101, 16405–16410.

508. Blakeley MP, Ruiz F, Cachau R, Hazemann I, Meilleur F, Mitschler A, Ginell S, Afonine P, Ventura ON., Cousido-Siah A, Haertlein M, Joachimiak A, Myles DAA and Podjarny AD. (2008). *Proc. Natl. Acad. Sci. USA.,* 105(6), 1844-1848.

509. Shu F, Ramakrishnan V and Schoenborn BP. (2000). Enhanced visibility of hydrogen atoms by neutron crystallography on fully deuterated myoglobin. *Proc. Natl. Acad. Sci. USA,* 97(8), 3872-3877.

510. Stura EA. (1999). *Crystallization of nucleic acids and Proteins;* edited by A. Ducruix and R. Geige. Oxford University Press, pp. 177-207.

511. Bergfors T. (2003). Seeds to crystals. *J. Struct. Biol*., 142, 66-76.

512. Stura EA and Wilson IA. (1991). Applications of the streak seeding technique in protein crystallization. *J. Cryst. Growth*, 110, 270-282.

513. Luft JR and DeTitta GT. (1999). A Method to Produce Microseed Stock for Use in the Crystallization of Biological. Macromolecules. *Acta Cryst.,* D55, 988-993.

514. Scheiner S and Čuma M. (1996). Relative Stability of Hydrogen and Deuterium Bonds. *J. Am. Chem. Soc.*, 118 (6), 1511–1521.

515. Shimizu N, Sugiyama S, Maruyama M, Yoshikawa HY, Takahashi Y, Adachi H, Takano K, Murakami S, Inoue T, Matsumura H and Mori Y. (2009). Growth of Large Protein Crystals by Top-Seeded Solution Growth Together with the Floating and Solution-Stirring Technique. *Cryst. Growth Des.*, 9 (12), 5227–5232.

516. Sears VF. (1992). Neutron scattering lengths and cross sections. *Neutron News*, 3 (3), 26-37.

517. Maeda M, Fujiwara S, Yonezawa Y and Nimura N. (2004). Crystallization of a large single crystal of cubic insulin for neutron protein crystallography. *J. Synchr. Rad*., 11, 41-44.

518. Bon C, Lehmann MS and Wilkinson C. (1999). Quasi-Laue neutron diffraction study of the water arrangement in crystals of triclinic hen egg-white lysozyme. *Acta Cryst*., D55, 978-987.

519. Teixeira SCM and Zaccai G. (2008). *Chemical Phys*., 345, 133-151.

520. Wlodawer A and Hendrickson WA. (1982). A Procedure for Joint Refinement of Macromolecular Structures with X-ray and Neutron Diffraction Data from Single Crystals. *Acta Cryst.,* A38, 239-247.

521. Fisher SZ. (2009). Preliminary joint neutron and X-ray crystallographic study of human carbonic anhydrase II. *Acta Cryst*., F65, 495-498.

522. Teixeira SCM, Allibon J, Mason S, Forsyth T, Demé B and Johnson M. (2011). Low resolution neutron crystallography of biological macromolecules. *Acta Cryst*., A67, C188-C189.

523. Madhusudhan MS and Vishveshwara S. (2001). Deducing hydration sites of a protein from molecular dynamics simulations. *J. Biomol. Struct. Dyn.,* 19, 105–114.

524. Mahalingam B, Louis JM, Hung J, Harrision RW and Weber IT. (2001). Structural implications of drug-resistant mutants of HIV-1 protease: high-resolution crystal structures of the mutant protease/substrate analogue complexes. *Proteins*, 43, 455–464.

525. Macrae CF, Bruno IJ, Chisholm JA, Edgington PR, McCabe P, Pidcock EL, Rodriguez M, Taylor R, van de Streek J and Wood PA. (2008). Mercury CSD 2.0 - New Features for the Visualization and Investigation of Crystal Structures. *J. Appl. Cryst.*, 41, 466-470.

526. Gilli P, Pretto L, Bertolasi V and Gilli G. (2009). Predicting Hydrogen-Bond Strengths from Acid−Base Molecular Properties. The pK_a Slide Rule: Toward the Solution of a Long-Lasting Problem. *Acc. Chem. Res*., 42, 33-44.

527. Gilli P, Bertolasi V, Ferritti V and Gilli G. (1994). Evidence for resonance-assisted hydrogen bonding. Covalent nature of the strong homonuclear hydrogen bond. Study of the O-H--O system by crystal structure correlation methods. *J. Am. Chem. Soc*., 116, 909-915.

528. Sham HL, Kempf DJ, Molla A, Marsh KC, Kumar GN, Chen CM, Kati W, Stewart K, Lal R, Hsu A, Betebenner DA, Korneyeva M, Vasavanonda S, McDonald E, Saldivar A, Wideburg N, Chen X, Niu P, Park O, Jayanti V, Grabowski B, Granneman GR, Sun E, Japour AJ, Leonard JM, Plattner JJ and Norbeck DW. (1998). ABT-378, a Highly Potent Inhibitor of the Human Immunodeficiency Virus Protease. *Antimicrob. Agents Chemother.*, 42, 3218-3224.

529. Kaldor SW, Hammond M, Dressman BA, Fritz JE, Crowell TA and Hermann RA. (1994). New dipeptide isosteres useful for the inhibition of HIV-1 protease. *Bioorg. Med. Chem. Lett*., 4, 1385-1390.

530. Coulson CA and Longuet-Higgins HC. (1947). The Electronic Structure of Conjugated Systems. I. General Theory. *Proc. Roy. Soc.* (London) Al9l, 39; Al92, 16.

531. Fukui, K. *Theory of Orientation and Stereoselection;* Springer-Verlag: Berlin, 1975.

532. Fujimoto H and Fukui K. *Chemical Reactivity and Reaction Paths;* Klopman, *G.,* Ed.; Wiley-Interscience: New York, 1974; pp 23-54.

533. Fukui K. (1982). Role of Frontier Orbitals in Chemical Reactions. *Science*, 218, 747-754.

534. Buergi H-B. (1973). Geometrical Reaction Coordinates. II. Nucleophilic Addition to a Carbonyl Group. *J. Am. Chem. Soc*., 95, 5065.

535. Buergi H-B. (1974). Stereochemistry of Reaction Paths at Carbonyl Centers. *Tetrahedron*, 30, 1563

536. Buergi H-B and Dunitz JD. (1983). From Crystal Statics to Chemical Dynamics. *Acc. Chem. Res*., 16, 153.

537. Buergi H-B. (1975). Stereochemistry of Reaction Paths as Determined from Crystal Structure Data. A Relationship between Structure and Energy. *Angew. Chem. Int. Ed. Engl*., 14, 460.

538. Huff JR. (1991). HIV protease: a novel chemotherapeutic target for AIDS. *J Med Chem*., 34, 2305–2314.

539. Jacks T, Power MD, Masiarz FR, Luciw PA, Barr PJ and Varmus HE. (1998). Characterization of ribosomal frameshifting in HIV-1 gag-pol expression. *Nature*, 331, 280–283.

540. Jacobsen H, Brun-Vezinet F, Duncan I, Hanggi M, Ott M, Vella S, Weber J and Mous J. (1994). Genotypic characterization of HIV-1 from patients after prolonged treatment with increased resistance to a C_2-symmetric protease inhibitor. *J. Virol*., 68, 2016–2020.

541. Ohta Y and Shinkai I. **(**1997). New Drugs-reports of New Drugs Recently Approved by the FDA. *Bioorganic & Medicinal Chemistry*, 5(3), 461-462.

542. Markowitz M, Saag M, Powderly WG, Hurley AM, Hsu A, Valdes JM, Henry D, Sattler F, La Marca A, Leonard JM and Ho DD. (1995). A preliminary study of ritonavir, an inhibitor of HIV-1 protease, to treat HIV-1 infection. *N. Engl. J. Med*., 333, 1534–1539.

543. Markowitz M, Mo H, Kempf DJ, Norbeck DW, Bhat TN, Erickson JW and Ho DD. (1995). Selection and analysis of human immunodeficiency virus type 1 variants with increased resistance to ABT-538, a novel protease inhibitor. *J. Virol*., 69, 701–706.

544. Ohtaka H, Schon A and Freire E. (2003). Multidrug Resistance to HIV-1 Protease Inhibition Requires Cooperative Coupling between Distal Mutations. *Biochemistry*, 42, 13659 – 13666.

545. Todd MJ, Luque I, Vela´zquez-Campoy A and Freire E. (2000). Thermodynamic basis of resistance to HIV-1 protease inhibition: calorimetric analysis of the V82F/I84V active site resistant mutant. *Biochemistry*, 39, 11876-11883.

546. Ho DD, Toyoshima T, Mo H, Kempf DJ, Norbeck DW, Chen CM, Wideburg NE, Burt SK, Erickson JW and Singh MK. (1994). Characterization of human immunodeficiency virus type 1 variants with increased resistance to a C2-symmetric protease inhibitor. *J. Virol*., 68 (3), 2016–2020.

547. Hosur MV, Bhat TN, Kempf DJ, Baldwin ET, Liu B, Gulnik S, Wideburg NE, Norbeck DW, Appelt K and Erickson JW. (1994). Influence of stereochemistry on activity and binding modes for C2 symmetry-based-inhibitors of HIV-1 protease. *J. Am. Chem. Soc*., 116, 847–855.

548. Kempf DJ, Marsh KC, Denissen J, McDonald E, Vasavanonda S, Flentge C, Green BG, Fino L, Park C, Kong X, Wideburg NE, Saldivar A, Ruiz L, Kati WM, Sham HL, Robins T, Stewart KD, Plattner JJ, Leonard J and Norbeck DW. (1995). *Proc. Natl. Acad. Sci. USA*, 92(7), 2484-2488.

549. Boden D and Markowitz M. (1998). Resistance to Human Immunodeficiency Virus Type 1 Protease Inhibitors. *Antimicrob. Agents Chemother*., 42, 2775-2783.

550. Johnson VA, Brun-Vézinet F, Clotet B, Kuritzkes DR, Pillay D, Schapiro JM and Richman DD. (2006). Update of the drug resistance mutations in HIV-1: Fall 2006. Top. HIV Med., 14, 125–130.

551. Clavel F and Hance AJ. (2004). HIV Drug Resistance. *N. Engl. J. Med*., 350, 1023-35.

552. Nair AC, Bonin I, Tossi A, Welsh WJ and Miertus S. (2002). Computational studies of the resistance patterns of mutant HIV-1 aspartic proteases towards ABT-538 (ritonavir) and design of new derivatives. *J. Mol. Graphics Modelling*, 21, 171-179.

553. Clemente JC, Coman, RM, Thiaville MM, Janka LK, Jeung JA, Nukoolkarn S, Govindaswamy L, McKenna MA-, McKenna R, Leelamanit W, Goodenow MM and Dunn BM. (2006). Analysis of HIV-1 CRF_01 A/E protease inhibitor resistance: Structural determinants for maintaining sensitivity and developing resistance to Atazanavir. *Biochemistry*, 45, 5468-5477.

554. Shafer RW and Schapiro JM. (2008). HIV-1 drug resistance mutations: an updated framework for the second decade of HAART. *AIDS Rev*iew, 10, 67–84.

555. Kutilek VD, Sheeter DA, Elder JH and Torbett BE. (2003). Is resistance futile? *Curr. Drug Targets. Infect. Disord*., 3, 295–309.

556. Yanchunas J. Jr., Langley DR, Tao L, Rose RE, Friborg J, Colonno RJ and Doyle ML. (2005). Molecular basis for increased susceptibility of isolates with atazanavir resistance-conferring substitution I50L to other protease inhibitors. *Antimicrob. Agents Chemother*., 49, 3825–3832.

557. Baxter JD, Schapiro JM, Boucher CA, Kohlbrenner VM, Hall DB, Scherer JR and Mayers DL. (2006). Genotypic changes in human immunodeficiency virus type 1 protease associated with reduced susceptibility and virologic response to the protease inhibitor tipranavir. *J. Virol*., 80, 10794-10801.

558. Scherer JR and Mayers DL. (2007). 11th Eur. HIV AIDS Conf. / EACS, October 24 - 27, 2007, Madrid, Spain. P3.4/07.

559. Arora SK, Gupta S, Toor JS and Singla A. (2008). Drug resistance-associated genotypic alterations in the pol gene of HIV type 1 isolates in ART-naive individuals in North India. *AIDS Res. Hum. Retroviruses*, 24(2), 125-130.

560. He´Le`Ne CF, CoˆTe´ Z, Brumme L and Harrigan PR. (2001). Human Immunodeficiency Virus Type 1 Protease Cleavage Site Mutations Associated with Protease Inhibitor Cross-Resistance Selected by Indinavir, Ritonavir, and/or Saquinavir. *J. Virol*., 75, 589–594.

561. Africa HN, Aracil BN, A´ Lvarez A, Barros C and Soriano V. (2001). Prevalence of Human Immunodeficiency Virus Type 1 (HIV-1) Non-B Subtypes in Foreigners Living in Madrid, Spain, and Comparison of the Performances of the AMPLICOR HIV-1 MONITOR Version 1.0 and the New Automated Version 1.5. *Journal of Clinical Microbiology*, 39, 1850–1854.

562. Hornak V, Okur A, Rizzo RC and Simmerling C. (2006). HIV-1 protease flaps spontaneously open and reclose in molecular dynamics simulations. *Proc. Natl. Acad. Sci. USA*, 103, 915–920.

563. Perryman AL, Lin J-H and McCammon JA. (2004). HIV-1 protease molecular dynamics of a wild-type and of the V82F/I84V mutant: Possible contributions to drug resistance and a potential new target site for drugs. *Protein Science*, 13, 1108–1123.

564. Baldwin ET, Bhat TN, Liu B, Pattabiraman N and Erickson JW. (1995). Structural basis of drug resistance for the V82A mutant of HIV-1 proteinase. *Nature Struct. Biol*., 2, 244-249.

565. Dundas J, Ouyang Z, Tseng J, Binkowski A, Turpaz Y and Liang J. (2006). CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. *Nucl. Acids Res*., 34, W116-W118.

566. Ala PJ, Huston EE, Klabe RM, McCabe DD, Duke JL, Rizzo CJ, Korant BD, DeLoskey RJ, Lam PYS, Hodge C.N., et.al. (1997). Molecular basis of HIV-1 protease drug resistance: Structural analysis of mutant proteases complexed with cyclic urea inhibitors. *Biochemistry*, 36, 1573–1580.

567. Ala PJ, Huston EE, Klabe RM, Jadhav PK, Lam PY and Chang CH. (1998). Counteracting HIV-1 protease drug resistance: structural analysis of mutant proteases complexed with XV638 and SD146, cyclic urea amides with broad specificities. *Biochemistry*, 37, 15042-15049.

568. Klei HE, Kish K, Lin, Pin-Fang M, Guo G, Friborg J, Rose RE, Zang Y, Goldfarb V, Langley DR, Wittekind M and Sheriff S. (2007). X-ray crystal structures of Human Immunodeficiency Virus Type 1 protease mutants complexed to Atazanavir. *J. Virol*., 81(17), 9525-9535.

569. Yedidi RS, Proteasa G, Martinez JL, Vickrey JF, Martin PD, Wawrzak Z, et.al. (2011). Contribution of the 80s loop of HIV-1 protease to the multidrug-resistance mechanism: crystallographic study of MDR769 HIV-1 protease variants. *Acta Cryst*., D67, 524-532.

570. Velazquez-Campoy A, Vega S and Freire E. (2002). Amplification of the effects of drug resistance mutations by background polymorphisms in HIV-1 protease from African subtypes. *Biochemistry*, 41, 8613-8619.

571. Weiss MS, Jabs A, Sühnel J and Hilgenfeld R. (2001). C-H...pi-interactions in proteins. *J. Mol. Biol.*, 307, 357-377.

572. Brandl M, Nishio M., Hirota M and Umezawa Y. (1998). The CH/π Interaction: Evidence, Nature and Consequences. Wiley-VCH, New York.

573. Takagi T, Tanaka A, Matsuo S, Maezaki H, Tani M, Fujiwara H and Sasaki Y. (1987). Computational studies on CH/π interactions. *J. Chem. Soc.: Perkin*, 2, 1015.

574. Trott O and Olson AJ. (2010). AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comb. Chem*., 31, 455-46.

575. HIV-1 Database, Stanford; http:// hivdb.stanford.edu.

576. Flor-Parra F, Pérez-Pulido AJ, Pachón J and Pérez-Romero P. (2011). The HIV type 1 protease L10I minor mutation decreases replication capacity and confers resistance to protease inhibitors. *AIDS Res. Hum. Retroviruses,* 27, 65-70.

577. Baharuddin A, Fischer M, van Maarseveen N, Nijhuis M, Boucher CAB, Hilfenfeld R and Mesters JR. (2006). Crystal Structures Of Related Hiv-1 Proteinase Variants: Relating Three-Dimensional Structure And Viral Fitness. *Antivir Therapy*, 11, S71.

578. Ode H, Matsuyama S, Hata M, Neya S, Kakizawa J, Sugiura W and Hoshino T. (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.

Keywords:

HIV-1 protease, enzyme, catalysis, mechanism, drug resistance, substrate, RT-RH, TFR-PR, tetrahedral intermediate, *Michaelis* complex, hydrogen bond, molecular replacement, LBHB, SIHB, V82A, M36I, ritonavir, crystal structure.

Abbreviations:

AIDS - Acquired Immuno-Deficiency Syndrome
HIV-1: Human Immunodeficiency Virus Type 1
SIV-1: Simian Immunodeficiency Virus Type 1
CD4: Cluster of Differentiation
CRF: Circulatory Recombinant Form
gag : Group Specific Antigen
pol : Polyprotein
RT: Reverse Transcriptase
RH: RNAse Hybrid
TFR: Trans FRame
$p6^{pol}$: 6 kDa protein flanking *gag* and *pol* region
PR: Protease
PDB: Protein Data Bank
RCSB: Research Collaboratory for Structural Bioinformatics
NMR: Nuclear Magnetic Resonance
TD : Tethered Dimer
TS: Transition State
NAC: Near Attack Conformation
TI : Tetrahedral Intermediate
LBHB: Low Barrier Hydrogen Bond
SIHB: Short Ionic Hydrogen Bond
SA : Simulated Annealing
ML: Maximum Likelihood
LLG : Log Lokelihood Gain
EWG: Electron withdrawing group