CHAPTER 1

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

1.1 History of Human Immunodeficiency Virus and AIDS 27
1.2 The Human Immunodeficiency Virus 28
  1.2.1 HIV-1 life cycle 29
  1.2.2 Challenges in vaccine development against AIDS 31
  1.2.3 Possibility of Gene Therapy 31
1.3 Targets in AIDS therapy 32
1.4 Role of HIV-1 protease in the viral life cycle 33
1.5 Structure of HIV-1 protease 34
1.6 Enzymes 37
1.7 HIV-1 protease is an aspartyl protease 37
1.8 Inhibitor and substrate complexes of HIV-1 protease 37
1.9 HIV-1 protease catalysis – the mechanistic proposals 39
1.10 Rates of cleavage of substrates by HIV-1 protease 42
1.11 Drug-resistance in HIV-1 protease 43
1.12 Aims of present study on HIV-1 protease 46
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, 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

![Figure 1: Geographical distribution (left) and clades (right) of HIV-1 genetic forms.](image-url)
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].
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.](image)

The *gag*, *pol* and *env* gene products are initially expressed as precursors Pr55\(^{gag}\), Pr160\(^{gag-pol}\) and gp160 respectively. The Pr160\(^{gag-pol}\) 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. There 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 nuclease 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.

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

### 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 Pr55gag and Pr160gag-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 $\alpha$-helix and two approximate antiparallel $\beta$-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 $\beta$-sheet involving N- and C-terminal $\beta$-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 $a\text{-}d$, $a'\text{-}d'$, $h$ and $q$. For each monomer, the N-terminal $\beta$-strand $a$ (residues 1-4) continues, through a turn, into $\beta$-strand $b$ (residues 9-15) which after a small turn is followed by $\beta$-strand $c$ (residues 18-23) and then the active-site $\beta$-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 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 accomplished 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 (= ΔG^‡_uncat - ΔG^‡_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 non-scissile 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].](image-url)
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 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.

<table>
<thead>
<tr>
<th>No.</th>
<th>Peptide sequences</th>
<th>Cleavage domain</th>
<th>Km (mM)</th>
<th>kcat (sec⁻¹)</th>
<th>kcat/Km (mM⁻¹. sec⁻¹)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>VSQNY*PIVQ</td>
<td>MA-CA</td>
<td>0.150</td>
<td>6.80</td>
<td>45.3</td>
</tr>
<tr>
<td>2</td>
<td>KARVL*AEAMS</td>
<td>CA-p2</td>
<td>0.010</td>
<td>0.09</td>
<td>90.0</td>
</tr>
<tr>
<td>3</td>
<td>TATIM*MQRGN</td>
<td>p2-NC</td>
<td>0.050</td>
<td>3.70</td>
<td>74.0</td>
</tr>
<tr>
<td>4</td>
<td>TERQAN*FLGKI</td>
<td>NC-p1</td>
<td>not detected</td>
<td>not detected</td>
<td>not detected</td>
</tr>
<tr>
<td>5</td>
<td>RPQNF*LQSRP</td>
<td>p1-p6</td>
<td>0.530</td>
<td>0.30</td>
<td>0.6</td>
</tr>
</tbody>
</table>

Cleavage sites in gag:

<table>
<thead>
<tr>
<th>No.</th>
<th>Peptide sequences</th>
<th>Cleavage domain</th>
<th>Km (mM)</th>
<th>kcat (sec⁻¹)</th>
<th>kcat/Km (mM⁻¹. sec⁻¹)</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>VSFNF*PQITL</td>
<td>TFRp6ʰ⁶¹-PR</td>
<td>&lt; 0.010</td>
<td>0.06</td>
<td>6.9</td>
</tr>
<tr>
<td>7</td>
<td>CTLNF*PISP</td>
<td>PR-RT</td>
<td>0.070</td>
<td>1.50</td>
<td>24.1</td>
</tr>
<tr>
<td>8</td>
<td>AETF*YVDGAA</td>
<td>RT-RH</td>
<td>0.040</td>
<td>0.40</td>
<td>10.0</td>
</tr>
<tr>
<td>9</td>
<td>IRKIL*FLDG</td>
<td>RH-IN</td>
<td>0.006</td>
<td>1.20</td>
<td>202.0</td>
</tr>
</tbody>
</table>

Cleavage sites in pol:

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 (approximately $3 \times 10^{-5}$ 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 1: 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.