

Characterization of a Non-Active Site Mutation of HIV Protease

A Major Qualifying Project:

Submitted to the Faculty

of the

WORCESTER POLYTECHNIC INSTITUTE

in partial fulfillment of the requirements for the

Degree of Bachelor of Science

By

Angelica Wzorek

Date: April 29, 2009

Approved

Professor Destin Heilman, Advisor

Abstract

HIV-1 protease is a significant target for the development of inhibitory drugs for the treatment of HIV-1. Inhibitors that have been designed to the shape of the active site have been shown to be potent inhibitors of protease activity. In order to expose the active site to the substrate, HIV protease displays significant movement in the flap region of the protease due to hydrophobic sliding. It has been proposed that by establishing a disulfide bridge in this region of the protease the conformational flexibility of the flap region to expose its active site will be restricted. To this end, a cysteine and alanine mutant has been created to test the feasibility of disulfide bridge formation in the flap regions. This study establishes that the single alanine and single cysteine mutations in the flap region do not affect the folding or binding ability of the protease significantly.

Acknowledgements

I would to thank the Schiffer lab for providing me with the opportunity to work with them by sponsoring my project. I would also like to thank Pr. Destin Heilman for his guidance and patience in advising my project.

Table of Contents

Abstract	2
Acknowledgements	3
Table of Figures	5
Table of Tables	6
1.0 Introduction.....	7
1.1 HIV History and Population Overview	7
1.2 HIV Genome	8
1.3 HIV Life-Cycle	9
1.4 Targeting HIV Treatment.....	11
1.5 HIV Protease Substrate Envelope Hypothesis: Structure Driven Design	14
1.6 Structure and Functions of the HIV Protease.....	15
1.7 Hydrophobic Sliding in HIV Protease	16
2.0 Methods.....	18
2.1 Expression of HIV Protease	18
2.2 Purification	18
2.3 Crystallization	19
2.4 Crystal Harvesting and Diffraction	19
2.5 Diffraction Data Processing	19
3.0 Results.....	20
4.0 Discussion	24
5.0 Figures.....	27
6.0 Tables	33

Table of Figures

Figure 1 HIV Clades	27
Figure 2 Prevalence of HIV in 2007.....	27
Figure 3 HIV Life Cycle Diagram	28
Figure 4 HIV Protease Schematic.....	28
Figure 5 Genomic Organization of HIV-1	29
Figure 6 Induction and Purification Gels.....	29
Figure 7 HIV Protease Crystals.	29
Figure 8 Obtained HIV Protease Structure	30
Figure 9 HIV Protease Comparision.....	30
Figure 10 Overlay of Residues 35-42	31
Figure 11 Cysteines in the Disulfide Bridge.....	32

Table of Tables

Table 1 Crystallographic Data Statistics..... 33
Table 2 Mutations of the Protease. 33

1.0 Introduction

1.1 HIV History and Population Overview

Human immunodeficiency virus (HIV) claimed the lives of two million people in 2007, with another thirty-three million estimated to be infected by the World Health Organization¹. The United States Center for Disease Control estimated that there were over a million people infected in the United States at the end of 2006².

AIDS (Acquired Immunodeficiency Syndrome) was recognized for the first time in the late 1970s, when patients with severe immunologic dysfunctions began to emerge; by 1983 scientists at the Pasteur Institute had isolated the HIV virus from patients with AIDS³. However, it was not until 1984 that Robert Gallo demonstrated that AIDS was caused by a retrovirus⁴. In 1986 the retrovirus was named HIV and the distinction between HIV-1 and HIV-2 had been made³. The difference between these two viruses arises solely from their species of origin, HIV-1 originated from the chimpanzee simian immunodeficiency virus (SIV), while HIV-2 originated from the corresponding virus in the feral sooty mangabey monkey³. This is substantiated by the similarities in the genomes and phylum of the two viruses, along with abundance of the virus in the monkeys and possible routes of transmission⁵. The realization that the virus had crossed over from monkeys in the region of Africa led to the sequencing of twenty different simian immunodeficiency viruses and the discovery that the virus is non-pathogenic in its natural host³. The earliest documented case of HIV-1 was identified by a blood sample taken in 1959 from a man in the Democratic Republic of Congo. This evidence along with genetic analysis has led scientists to propose that the virus entered the human population sometime around 1930 and that there have been several crossover events into the human population⁵.

Phylogenetic analyses of HIV-1 have led to the identification of distinct viral groups; these groups are named M (Main group), O (Outlier group), and N (the non-O and non-M group)⁵. The M- group virus is the most prevalent globally and contains eleven clades, or clusters of genetic subtypes of the virus named A through K along with their circulating recombinants. Together they consist of 95% of the world's cases of HIV (see Figure 1)³. Clade B virus is the most widely researched virus because of its prevalence in the United States and Europe where the primary research on HIV is occurring. Interestingly, the more prevalent subtypes in Africa and Asia, while affecting more people are less researched.⁶ The wide diversity of the virus contributes to recombination between subtypes when someone is coinfecting with two or more subtypes of the virus resulting in circulating recombinant forms of HIV, or CRFs. Scientists have been studying these CRFs to track the spread of the virus.

HIV is a retrovirus which is classified as a lentivirus. This virus is characterized by its long incubation period before infection becomes evident. Also, it is unique in that it has a cylindrical or conical core. Most lentiviruses have genomes that express *gag*, *pro*, *pol* genes; however, HIV expresses many more proteins than the average lentivirus which contributes to its virulence³.

1.2 HIV Genome

The RNA genome of HIV (see figure 5) is different from that of other retroviruses that contain only three genes (*gag*, *pol*, and *env*). HIV, in addition, to these encodes fifteen proteins and contains the genes *nef*, *vif*, *vpu*, *vpr*, *tat*, *rev*, along with an open reading frame near the 3' long terminal repeat that codes for a super antigen³. The *gag*, *pol*, and *env* genes encode for the viral packing proteins. The *nef* gene encodes for the nef protein, this protein plays a role in down regulating the cell surface molecules CD4, MHC-I, and MHC II, along with the *env* and *vpu*

proteins³. It has also been observed to enhance viral infectivity in comparison to the *nef*-deleted virus³. The *vif* gene has been observed to increase the viral infectivity factor by aiding virus assembly; this is indicated by experiments with viruses with a *vif* mutated virus which results in altered core assembly³. The *vpu* gene is unique to the HIV-1 virus, it enhances viral release and the degradation of CD4 cells.³ The *vpr* gene has been observed to enhance the transcription of the genome by encoding the nuclear localization sequence and arresting the host cell in the G₂ phase of the cell cycle which is important because the lentivirus replicates in nondividing cells³. The *tat* and *rev* genes encode for HIV regulatory proteins where *tat* stimulates transcriptional processivity by assembling transcription complexes and promoting the phosphorylation of RNA pol II and *rev* promotes the nuclear export of intron-containing RNA³.

The HIV is a pseudodiploid, meaning that there are two ssRNA strands in the virion, and only one will be used as the functioning provirus³. When cells are coinfecting with subtypes of HIV, the RNA that are packaged into the virion can recombine to form a new CRF. This is packaged with the help of the non-coding region of the Gag-polyprotein which interacts with four RNA stem loops to create the core encapsidation signal, then the RNA associates and is then packaged with the Gag-polyprotein which will autocatalyze to form the necessary replication and assembly enzymes integrase, protease, reverse transcriptase³.

1.3 HIV Life-Cycle

The HIV life cycle consists of the binding and fusion of the virus with the cell membrane and reverse transcription of the ssRNA. This continues to the integration of the newly synthesized DNA into the host DNA, transcription and translation of the DNA to RNA then to protein, the assembly of the proteins, and budding of the new virus (See Figure 3).

The binding and fusion of the virus to the host cell is mediated by the Env glycoproteins located on the surface of the virus. Glycoprotein120 (gp120) binds to CD4 cells with high affinity on the cell surface, these cells are immunologically important for their interactions with T-cell lymphocytic MHC-II receptors^{3,7}. The HIV gp120 and gp41 bind to the CD4 receptor with high affinity encouraging the fusion of the cell and viral surfaces³. The fusion of the two does not occur unless the chemokine coreceptors CCR5 and the CXCR4 bind to the viral protein³. In some individuals a mutation in the inherited alleles of these coreceptors has led to resistance against the HIV-1 virus³.

The direct fusion of the two surfaces leads to the introduction of viral subparticles in the cytoplasm of the host cell; this event initiates the reverse transcriptase of HIV. The HIV reverse transcriptase does not operate with high fidelity leading to five to ten base pair errors in the HIV genome for every round of replication⁸. After the RNA has been processed to dsDNA, it is transported to the nucleus for integration into the host chromosome by integrase. During this process the new DNA is associated with a pre-integration complex (PIC); this complex is transported to the nucleus with the DNA by the Vpr protein contained in the PIC³. HIV-1 is unique among retroviruses because of the selectivity with which it integrates into the actively expressed regions of the chromosome. This selectivity has been attributed to PIC and is likely one of the many factors involved in the level of virulence of HIV-1³. Contributing to the promotion of the replication of the HIV DNA are the HIVs long terminal repeats that include eukaryotic enhancers and promoters for RNA pol II.

After the mRNA has been assembled into proteins by ribosomes the resulting Pol and Gag-Pol polyprotein must be processed by the HIV-1 protease. The resulting proteins are matrix, nucleocapsid and, capsid proteins, and p6; these proteins assemble to make the structural

components of the virus³. The proteins receive the encapsidation signal from the interaction of the four RNA stem loops and the Gag polyprotein.

Once the virus is encapsulated by the protein p6, which promotes the incorporation of accessory proteins into the virus and with the endosomal sorting proteins in the cell promotes the release of the virus from the plasma membrane³. When the virion is released from the cell its new surface membrane is that of the host cell plasma membrane. This is where the virion uptakes many of the cellular surface proteins along with the Env proteins that were expressed in the ER and incorporated into the plasma membrane prior to the budding process³. The mature virus contains a conical core with ssRNA and Gag-Pol proteins ready to infect another cell.

HIV *vpu* protein associates with CD4 molecules in the endoplasmic reticulum and promotes the degradation of the molecule allowing for the transport of Env to the cell surface for viral packaging⁷. In addition the *nef* and *vpu* proteins down regulate the surface expression of MHC proteins, which would normally present viral epitopes for recognition by cytotoxic T-cells⁷. By degrading the CD4 receptor and suppressing MHC expression the virus can essentially hide in the body while slowly destroying the host immune system. The depletion of CD4 cells in the body eventually leads to Acquired Immunodeficiency Syndrome or AIDS.

1.4 Targeting HIV Treatment

The main targets for HIV treatments occur at major events in the HIV life cycle. The list of available inhibitors include: fusion inhibitors, reverse transcriptase nucleoside and reverse transcriptase non-nucleoside inhibitors, protease inhibitors, and integrase inhibitors. In addition there are various pioneering efforts being made to develop a vaccine against HIV; however,

these have been met with limited or no success, leaving the option of preventing the disease untapped.

The high rate of replication and error prone reverse transcriptase of the HIV provide a pathway for the creation of mutant forms of the virus and its components. However, it should be noted that in untreated patients the wild-type protease is the most prevalent. In response to treatment over 50% of patients in the U.S. receiving anti-retroviral therapy express resistance to at least one of the available anti-retroviral drugs⁹. This also brings concerns that the transmission of drug resistant HIV will become prevalent. The variability of the virus has led to a comprehensive course of treatment incorporating the use of multiple anti-retroviral drugs, mainly a mixture of reverse transcriptase and protease inhibitors, nicknamed HAART or highly active anti-retroviral therapy⁹. HAART therapy focuses on targeting the reverse transcriptase and the protease of the virus, creating a situation where the virus is forced to mutate at multiple sites to become resistant⁹. The success of the therapy is limited due to the fact that mutations will slowly arise and the treatment will need to be altered or the patient will experience treatment failures⁹.

Nucleoside and nucleotide analog reverse transcriptase inhibitors were the first antiretroviral drugs approved by the Food and Drug Administration for the treatment of HIV in 1987¹⁰. The drug is structurally similar to the nucleic acid and so competes with the natural nucleic acid as a substrate¹⁰. The inhibition is achieved by replacement of the 3' hydroxyl group of the analog, making the phosphodiester linkage required for elongation of the strand impossible¹⁰. The non-nucleoside analog reverse transcriptase inhibitors directly bind to the reverse transcriptase and are non-competitive inhibitors of the enzyme¹⁰. They work to block the DNA polymerase activity by binding to the hydrophobic domain next to the catalytic site of the enzyme¹⁰⁹.

As mentioned before, the glycoproteins 41 and 120 play a role in the fusion of the viral and cellular membrane. Upon binding to the CD4 and CCR5 two peptide motifs are formed called heptad repeats, these two motifs unite to form a six-helix bundle which pulls the virus into the cell¹⁰. The only FDA approved fusion inhibitor is called enfuvirtide and interrupts the formation of this bundle of heptads to prevent the fusion of the viral and cellular membranes¹⁰. This has provided drug experienced patients with some success in treatment. However, due to the intensive process to make the peptide drug is quite expensive compounded by the fact that it is administered twice a day and that most patients will experience a rash at the injection site, the cost-effectiveness of the drug remains a large factor in its use. Treatments with protease and transcription inhibitors are preferable to fusion inhibitors not only because of their cost, but their relative effectiveness. To date, the treatments for HIV that focus on targeting HIV protease have not only proven to be successful but there is a wide selection available.

The role of HIV protease in the maturation of the human immunodeficiency virus verifies it as a target for inhibitors with the aim of arresting the development of the virus. There are currently nine protease inhibitors available on the market that have been approved by the United States Food and Drug Administration¹¹. These drugs are competitive inhibitors for the active site of the protease and all of the inhibitors, save one, are peptidomimetics¹¹. This means that the inhibitor can mimic the transition state of the substrate, but the peptide cannot be cleaved by the protease due to its hydroxyethylene or hydroxyethylamine core¹¹. Hence the enzyme is inhibited.

The emergence of HIV protease inhibitors on the market also brought about the emergence of mutations in the protease that were negatively selected by the virus to disrupt inhibitor binding while maintaining the enzyme function¹². These mutations of the protease that

affect inhibitor binding can also affect substrate specificity. Therefore the selective pressure is present for the substrate of the protease to co-evolve with the enzyme¹³. In this respect it has been shown that the non-active site mutations are more abundant and occur later than active site mutations; however, these mutations play a larger role in lowering the affinity of inhibitors while stabilizing enzyme-substrate binding¹⁴. The protease recognizes several sites along the gag-pol polyprotein and it has been observed that the most thermodynamically favorable inhibitors have the same volume and shape as the substrate cleavage sites. This has led to a method of a structure driven drug design for HIV protease inhibition.

1.5 HIV Protease Substrate Envelope Hypothesis: Structure Driven Design

Structure driven design of HIV protease inhibitors is based on the observation that the substrates of the protease have a rather uniform shape, despite their differing constituents¹². The border of the volume they occupy has been coined the “substrate envelope”¹². It has been observed that when the consensus volume of the bound inhibitors differs greatly from the substrate volumes that clinically relevant mutations have occurred in the areas where the inhibitors have protruded from the substrate envelope¹². This has led to the formation of the “substrate envelope hypothesis” that an inhibitor which fits the substrate envelope will resist clinically relevant mutations, and the mutations that do arise would not only reduce the affinity of the inhibitor but also the substrate, making viable escape mutations unlikely^{15,16}. Using this hypothesis to facilitate a structure based drug design approach; researchers of HIV protease have been able to analyze the binding interactions among protease and inhibitor. The success of the inhibitor darunavir in treatment experienced patients has been attributed to how well it fits in the substrate envelope¹⁶.

1.6 Structure and Functions of the HIV Protease

HIV protease (see Figure 4) is a homodimer of a consisting of two ninety-nine amino acid monomers produced by the *pol* gene that recognizes and cleaves sites in the *gag-pol* and *gag* polyproteins¹⁷. The cleavage of these polyproteins as indicated in section 1.3 is essential to the life-cycle of HIV. Each monomer contributes one conserved aspartic acid (D25) to the substrate active site, which is a pocket formed within the interface of the two monomers with the opposing aspartic acid residues necessary for cleavage opposite each other at the base of the active site¹⁸. When the molecule is dimerized it works as an aspartyl protease using the aspartic residues to cleave ten sites non-homologous sites along the polyprotein¹⁸. The flaps of the protease are flexible and are known to play a role in the binding of substrates and inhibitors by using this flexibility to accommodate substrates¹⁹.

The binding and cleavage of the substrates by the HIV protease is mediated by water molecules that facilitate the action via acid- base catalysis¹⁸. The water molecule donates a proton to the carboxyl groups of the aspartic acid and the proton is then transferred to the corresponding peptide bond completing the cleavage¹⁸. Upon binding to the substrate the protease undergoes a large conformational change¹⁷. HIV protease is a symmetric enzyme that recognizes substrates with asymmetric charge distribution and size¹⁷. The protease creates different binding surfaces for these substrates where there are at least eight residues that bind to the substrate prior to cleavage¹⁷. Within the active site of the protease all the residues of the either adjust to the substrate without binding to it or have Van der Waals interactions that are mediated by water molecules positioned between the substrate and protease¹⁷. It has also been observed that the α -carbons of the active site residues do not shift; however, the rotation of the

side-chains and rearrangement of water molecules within the site alter facilitate the binding of the substrate residues¹⁷.

1.7 Hydrophobic Sliding in HIV Protease

Forty of the ninety-nine residues of HIV protease are hydrophobic; a number of the residues line the active site or the flap region of the protease²⁰. These hydrophobic residues aid in the conformational changes that occur with HIV-1 protease by exchanging Van der Waals contacts and maintaining structurally significant hydrogen bonds throughout the conformational changes involved in opening the flap region of the protease²¹. While these residues are outside the active site they mutate along with active site mutations. The mutation of protease residues outside the active site occurs because non- active site mutations help stabilize the protease while the active site mutations decrease structural integrity¹⁴. In this way the mutations assist the protease in developing less affinity to the inhibitor while sacrificing little loss of affinity towards the substrate. Most mutations that occur in the hydrophobic core of the protease are mutations to other hydrophobic residues; this leads researchers to conclude that the hydrophobic interactions in the core are important to HIV-1 protease function²¹.

To assess the contributions of non-active site mutations on the overall function of the protease is important to understand exactly how they play a role in the movement of the protease. Engineering disulfide bridges into the protease is a proposed method to lock the protease in a particular conformation allowing for an analysis of how protease motion contributes to catalytic activity. In the non-active site of HIV-1 protease there are residue sites where the α carbon to α carbon and β carbon to β carbon distances are favorable for the position of cysteines to form a disulfide bond. To provide a comprehensive background for this type of research it is necessary to establish the effect of substituting the original residue with another. Using alanine mutagenesis

to substitute the original residue with an alanine eliminates the side chain of the residue past the β carbon and should not alter main-chain conformation, or impose extreme electrostatic or steric effects²². By analyzing the effect of this substitution on the overall catalytic activity of HIV-1 protease a better comparison of the effect of the engineered cysteine substitutions can be made.

2.0 Methods

2.1 Expression of HIV Protease

For the amplification of the HIV-1 protease with the mutations of interest Z-competent BL21 (DE3) *E. coli* cells with a pET-11a plasmid vector (Novagen) were induced with 0.33mM IPTG (isopropyl-beta-D-thiogalactoside) when the absorbance at 600nm was 0.5.

During expression the protease aggregates in inclusion bodies. To lyse these bodies the cells were resuspended using elution buffer 1 (pH 8.0; 20mM Tris pH 7.5, 1mM EDTA, 10mM DTE, 51.4mg lysozyme, 50µL 0.2PMSF in isopropanol) per pellet and let it sit on ice for 30 minutes. A cell disrupter was used to lyse the cells; the cells were then spun in Oakridge tubes for 30 minutes at 20000g. The cells were resuspended in 20mL of elution buffer 2 (pH 8.0; 20mM Tris pH 7.5, 1mM EDTA, 10mM DTE and 2M Urea). The cells are centrifuged for 30min at 20000g and resuspended in 20mL of elution buffer and spun for 30 minutes at 20000g twice and treated with elution buffer (pH 8.0; 20mM Tris pH 7.5, 1mM EDTA,7M guanidine) for one hour one ice. The protein was dialyzed against the elution buffer overnight at 4°C with two changes of buffer.

2.2 Purification

20mL of the protease protein was loaded onto a G-75 Sphedex column in 50% acetic acid at a flow rate of about 3mL/hour. Using refolding buffer (50mM Sodium Acetate pH5.5, 10% glycerol, 5% ethylene glycol, 5mM DTT) the protease was refolded using a rapid dilution technique involving a peristaltic pump slowly dripping the protease into a beaker of refolding buffer at 0°C and overnight dialysis at 4°C. After the dialysis was complete the protein was concentrated using a 2µ Amicon concentrating device until the absorbance at 280nm was greater

than 1.5. Additionally the protease was purified by fast liquid protein chromatography, to remove contaminants from the solution.

2.3 Crystallization

Protein at a concentration of 1.5mg/mL was incubated with three times excess of darunavir at room temperature for 30 minutes. Using the hanging drop method the tray with a protease-darunavir concentration of 1.2mg/mL in a buffer of 50mM sodium phosphate 7%DMSO and 30-41% ammonium sulfate yielded crystals.

2.4 Crystal Harvesting and Diffraction

A crystal was identified and collected using a Hampton loop and submerged in a cryogen of glycerol and 30% of crystallization buffer to prevent the buildup of ice under the cryostream. The crystal was diffracted using an R-Axis IV image plate mounted on a Rigaku rotating anode at -80° , a total of 200 3-minute frames were collected using a 1° oscillation.

2.5 Diffraction Data Processing

Model building was completed using COOT²³ and CCP4i²⁴. The diffraction data was processed and a molecular replacement model made using the aMoRe function. The figure was refined at a sigma value of 1.0 using COOT with data adjustment in Refmac5 for several rounds of refinement.

3.0 Results

As a first step in obtaining protein for crystallization BL21 (DE3) *z*-competent *E. coli* cells were transformed with the plasmid of choice provided by the Schiffer lab. The next step in the protocol was to amplify the transformed cells in a large scale expression over a three hour period. In Figure 6A an SDS-PAGE gel stained with coomassie is depicted. In lane one a sample of the cells is visualized without IPTG induction, the next three lanes represent specific one hour intervals after cell induction. Comparing these samples to a purified HIV protease marker shows that the induction with IPTG produced a robust expression of HIV protease. The band that is protease is referred to by an arrow and shows significant expression one hour after induction.

During expression the protease aggregated into inclusion bodies and in order to harvest the protease from the cells it was necessary to use a cell press along with the denaturing protocol assigned in the methodology. Following this step it was necessary to purify the protein from the remaining cellular debris, this was done by running the protease through a 50% acetic acid column. This step yielded several fractions over the length of the purification process that contained protease. These were run on an SDS-PAGE gel using a kaleidoscope marker and were stained with coomassie to determine if they were pure (Figure 6 panel B). The resulting gel showed one size band for all fractions indicating that the protease collected was pure enough to continue with the protocol.

During extraction from the cells the protease had been denatured and in order to continue the protease was refolded using a rapid dilution system. Once refolded, the protease was further purified by fast liquid protein chromatography (FPLC). This step ensures that the protease used for crystal trials is not contaminated with any other proteins. To further produce crystals the protein needs to be in a stable complex with substrate or inhibitor, to accomplish this protease

was incubated in three times excess of darunavir. To form crystals the protease darunavir complex is mixed with a salt buffer in a drop hung above a well with the same buffer this is called the hanging drop setup.

To identify the best set of conditions for crystallizing this mutant in crystal trials multiple buffer conditions and concentrations of protease were tested. Interestingly it was found that regardless of whether or not the protease had been purified by FPLC that the protease formed in a concentration of 1.2 mg/mL of protease and darunavir in a buffer of 50mM sodium phosphate and 30-40% ammonium sulfate (Figure 7). Once the crystallization conditions were identified by the formation of protease crystals it was then possible to harvest crystals for data collection by x-ray diffraction. Due to the size and availability of crystals formed from the protease that had not been purified by fast liquid protein chromatography, the crystal that was collected for analysis here was not from a condition using FPLC purified protein.

In order to obtain a diffraction pattern for the crystal devoid of extra water the crystal was submerged in a cryogen before diffraction at -80°C to prevent ice build-up. The resulting diffraction data was resolved using CCP4i and COOT. This revealed that that the space group of the collected protease crystal was $P2_12_12_1$, this describes the unit cell as a primitive lattice with three perpendicular twofold screw axes (Table 1). This group spacing along with the cell parameters for a, b, and c (51.98\AA , 57.38\AA , and 60.68\AA respectively) is consistent with other HIV protease-substrate complexes commonly found in literature.

In order to have a starting structure for the experimental data a molecular replacement solution was created using CCP4i's aMoRe feature. To further elucidate the structure several rounds of refinement in COOT followed by data adjustment in Refmac5 were employed. To

ensure that the experimental data were in line with the human and computer refinement it was necessary to compute the R_{work} and R_{free} values of the structure, these were obtained using Refmac5. These values give a statistical representation of the agreement between experimental and manipulated data (Table 1). The G16C, L38A mutant data yielded a structure that did not differ from the overall structure of the wild type protease (Figure 8). The protease was modeled with 121 water molecules and two sodium molecules and appears to have the usual conformations for the flap region and the active site.

To draw more significant conclusions about the overall differences in the conformation of this mutant it was compared with 1T3R, an HIV protease mutant close to the wild type with only the catalytic aspartic residues mutated to asparagines that had been incubated with darunavir. Comparison with this structure revealed that there is a disagreement in the conformation of the protease in the flap region. Several factors that may be responsible for this change include that there was a general lack of density in the area of residues 36-42, and that this crystal is not from a protein sample that had been purified by fast liquid protein chromatography, meaning that contaminants could have disrupted the x-ray diffraction and data processing.

In order to identify the cause and extent of the disruption in protein folding the residues in the area of interest are highlighted in Figure 10 along with the corresponding residues in from 1T3R. This leads to the observation that the mutations of the G16C, L38A protease (Table 2) seem to have disrupted protein folding in the flap region of the protease. These changes may also be due to the movement seen in this area of the protease or simply due to a poor quality of the crystal. However, without more data available there is no conclusive answer as to the reason.

The binding of inhibitor reveals that the overall catalytic activity of the protease has been conserved in this mutant. To ensure the feasibility of the formation of a disulfide bridge in the flap region between the 38th and 16th residues, the 38th alanine was mutated by computer simulation to a cysteine and distance measurements taken (Figure 11A). Initial measurements in space indicate that the lengths from sulfur to sulfur and from sulfur to β -carbon were about 3.58Å and 4.40Å not favorable for formation of a disulfide bridge. Due to the flexibility of the side chain of the cysteine amino acid further computer manipulation in space finds a favorable conformation of the 38th residue for the formation of a disulfide bond with the 16th residue. These residues are situated at the outer regions of the protease and would theoretically keep the flap area of the protease in a closed conformation, precluding it from binding to either substrate or inhibitor.

4.0 Discussion

By creating mutations that affect the functionality of HIV protease it is possible to visualize the movement of the protease and assess contributions of specific residues to its stability and catalytic activity. Previous studies have shown that non-active site mutations arise secondary to active site mutations and provide a mechanism for stabilizing the active site. In the treatment of HIV this means that the resistance to the HIV protease inhibitor drugs that develops overtime will become stronger as time lapses, making the need for new and better treatments of HIV a necessity.

In this study non-active site mutations were induced to explore the feasibility of disulfide bridge formation. The G16C mutation demonstrates that alteration to this area of the protease has insignificant effects on the structure of the protease. However, the mutation L38A causes a disruption in the region surrounding the flap region of the protease. Closer examination revealed that there were no complementary sections between this structure and the wild type. Despite this disagreement in structure the catalytic site is still active, meaning that the proposed mutations to cysteines that will contribute to a disulfide bridge are favorable for our purposes. Computer simulation demonstrates that it is possible to expect a disulfide bond to form between the two residues. The goal here is to restrict the overall flexibility while conserving the catalytic activity of the protease.

The flexibility of the protease is due to the large amount of hydrophobic residues in the flap region. While this flexibility would be challenged by the formation of a rather rigid disulfide bond, further mutation of the lysine 38 to a cysteine would be required to make a conclusive comment about the effect to function. This study provides the background for understanding the changes induced by creating the additional proposed mutations. The flap region would not have a normal conformation, however the active site would retain its function. This would produce a

desirable way to restrict protease function. This study shows that the formation of a disulfide bond is plausible, further studies would need to assess the effect of the responsible mutations concurrently in the protease to identify if they would indeed cause a restriction of the active site and prevent substrate access. If the active site was not restricted by this disulfide bridge then the current thinking behind the mechanism for hydrophobic sliding would need to be re-evaluated. Current approaches to drug design could be augmented by a better understanding of this region and its movement.

Due to the mechanism of hydrophobic sliding that occurs when the protease binds to substrate or inhibitor, there is some amount of thermodynamic energy being expended. If a protease mutant can be made with an engineered disulfide bond it would then be a model that could be used in the development of future protease drugs. A peptide inhibitor that does not require the expended energy of protease movement for binding would present a more favorable target for the protease than current inhibitors. Such an inhibitor would create a more thermodynamically stable protease inhibitor complex than the substrate and perhaps lead to the effective inhibition of the protease.

If this mechanism cannot be achieved then there are other HIV protease drug targets that can be examined for development, such as disrupting the hinge region of the protease by preventing dimer formation. Such a model would also inhibit HIV protease *in vivo* before it seeks out substrate or peptide inhibitors. Other areas that could be explored would include possibly using inhibitors to attract degradation complexes already present in the cell such as proteasomes. Such a mechanism could tether several ubiquitin complexes to an inhibitor and signal the protease for degradation. Using such mechanisms that already exist in the cell could possibly provide a pathway for the development of treatments without incurring many side

effects. Another way that HIV can be targeted is through using the simian animal that it is comes from as a model for treatment. As mentioned earlier, the origin of HIV has been identified as coming from various monkey species in Africa. In these species it has been noted that SIV has no apparent immunological consequence, perhaps one day it will be possible to create a vaccine using their viral form and resistance to it as a basis for development.

The wide effect of HIV on the population of the world (Figure 2) makes it a target for the development of drugs and community awareness for its treatment and prevention. The spreading awareness of HIV and methods for preventing the disease will decrease the spread of the disease. Regardless of the education and awareness of the population, the disease will continue to spread because many people do not know they have the disease until symptoms arise and continue to spread it until it is detected. Also as new, mutated forms of the virus are spread further inhibitory treatments will need to be developed until the point when a completely effective vaccine or treatment can be created. Until that time it is imperative that research into HIV drug development continues.

5.0 Figures

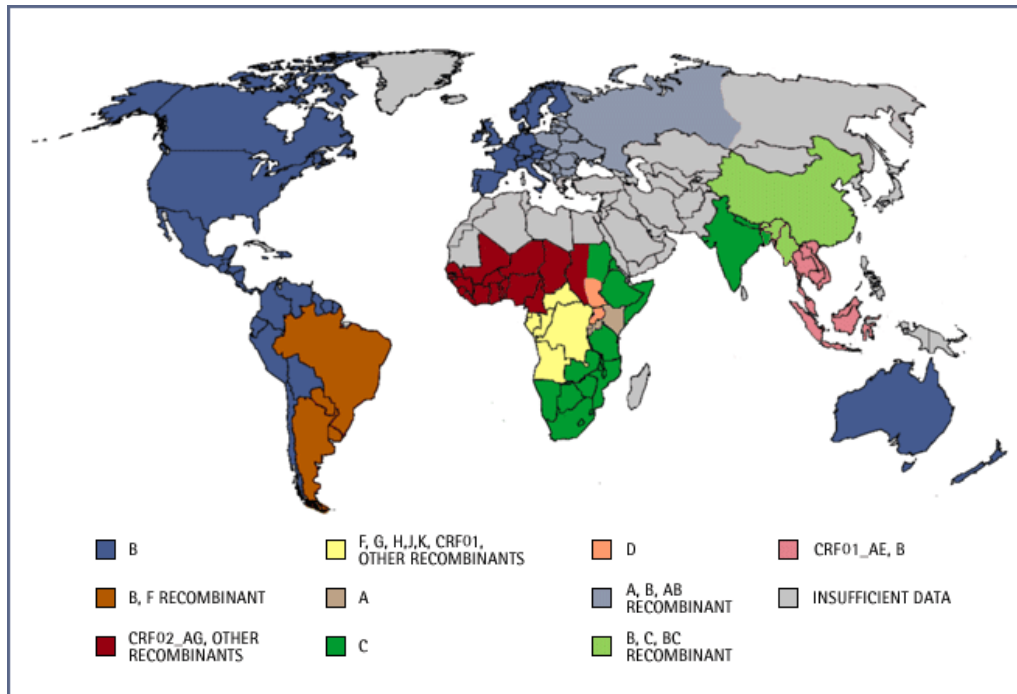


Figure 1 HIV Clades: The different clades of HIV around the world. Clade B is most prevalent in the United States and Europe and is therefore the most researched ²⁶.

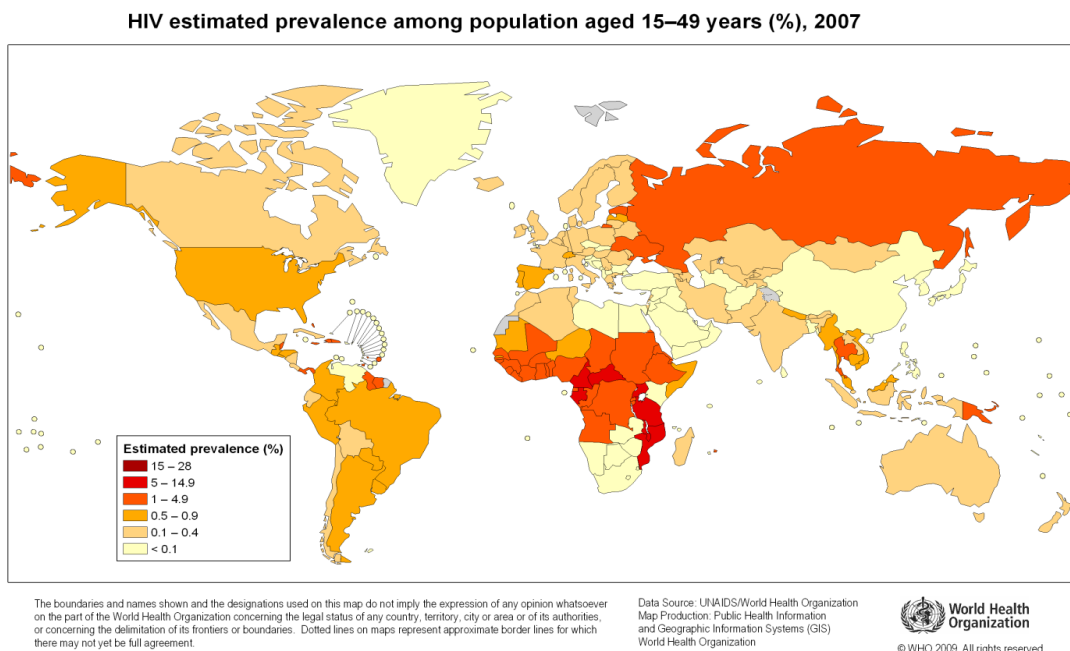


Figure 2 Prevalence of HIV in 2007: The prevalence of HIV in the world among populations aged 15-49 in 2007, the regions with the highest prevalence of HIV are Africa and Russia ²⁷.

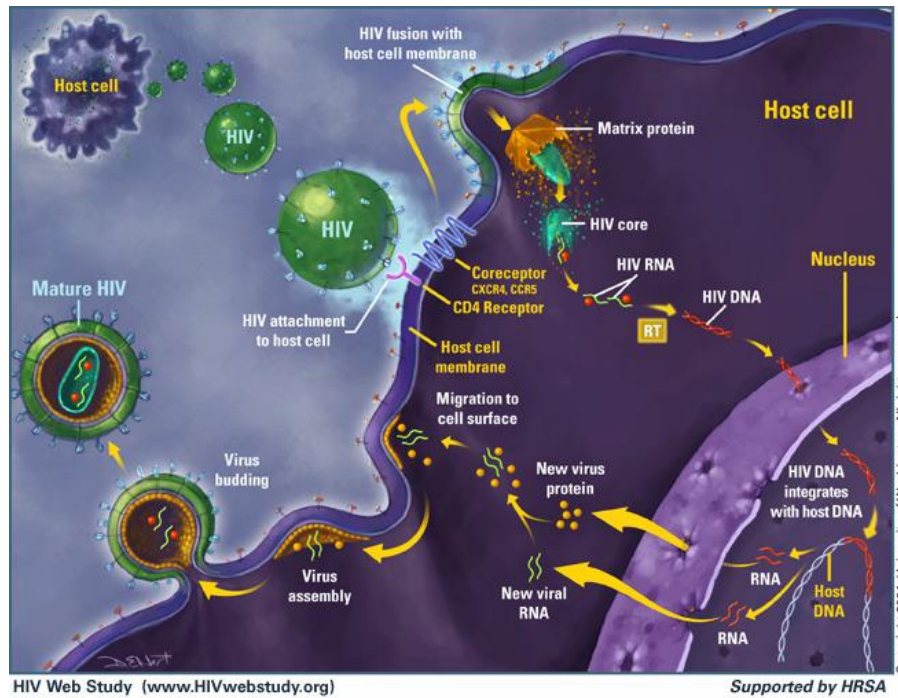


Figure 3 HIV Life Cycle Diagram: This diagram shows the life-cycle of the HIV from fusion with the cell membrane to viral assembly and budding²⁸.

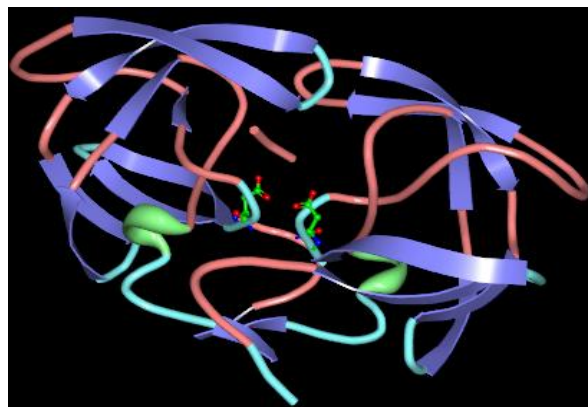


Figure 4 HIV Protease Schematic: HIV protease complexed with a tripeptide inhibitor, shown here with the catalytic aspartic acid residues highlighted^{29, 30}.

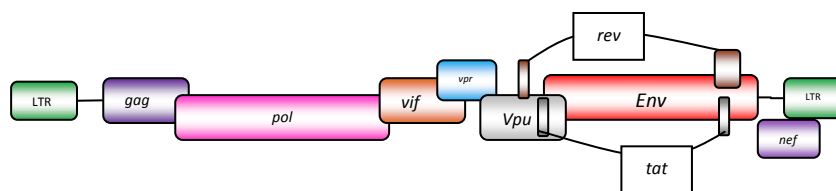


Figure 5 Genomic Organization of HIV-1: Schematic of the genome of HIV.

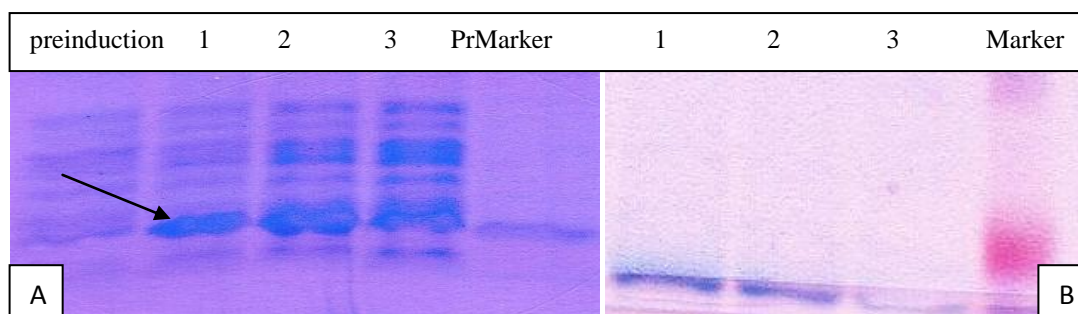


Figure 6 Induction and Purification Gels: To the left, panel A, is the induction gel with a purified protease marker, *e. Coli* cells were induced with IPTG over a three hour period. The arrow points to the band of protease after one hour of expression, which already shows strong expression of the protease in this system. To the right, panel B. The harvested protease protein was denatured and run on a 50% acetic acid column. To the right is a gel of various protease samples collected during the purification with the kaleidoscope marker on the right, the samples are fairly well purified.



Figure 7 HIV Protease Crystals: Picture of protease and darunavir crystals in the well of buffer before they were harvested for x-ray diffraction.

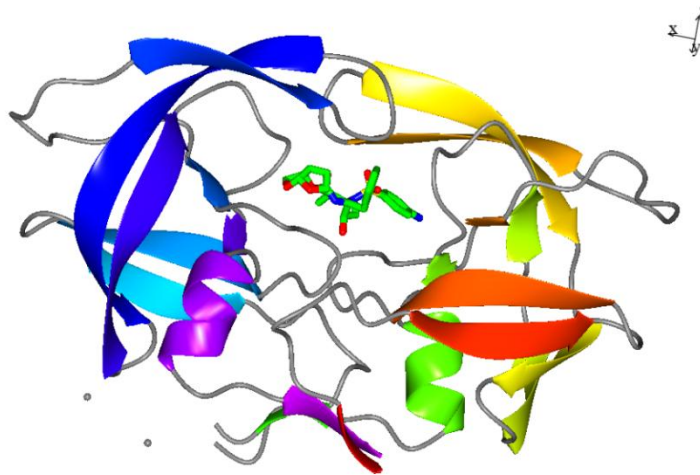


Figure 8 Obtained HIV Protease Structure: HIV protease with G16C and L38A mutations, displayed in ribbons.

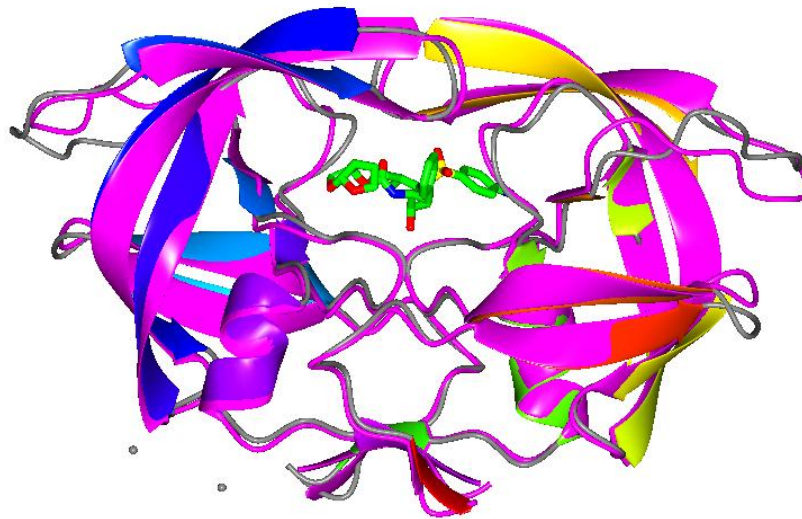


Figure 9 HIV Protease Comparison: Overlay of obtained HIV protease structure and wt-HIV protease complexed with darunavir (in magenta), notice the difference in conformation of the outer loops.

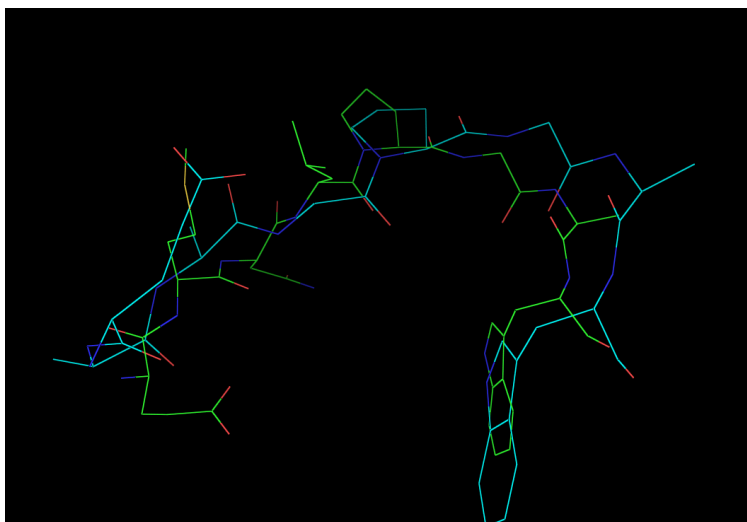


Figure 10 Overlay of Residues 35-42: Residues 35-42 of WT-protease and of the obtained structure overlaid the green strand is the wild type protease and the cyan is the obtained protease structure. The difference in side chain conformation may be due to the mutation of the 38th lysine to an alanine. Overall the overlay reveals that the conformation of this protease is different from the wild type.

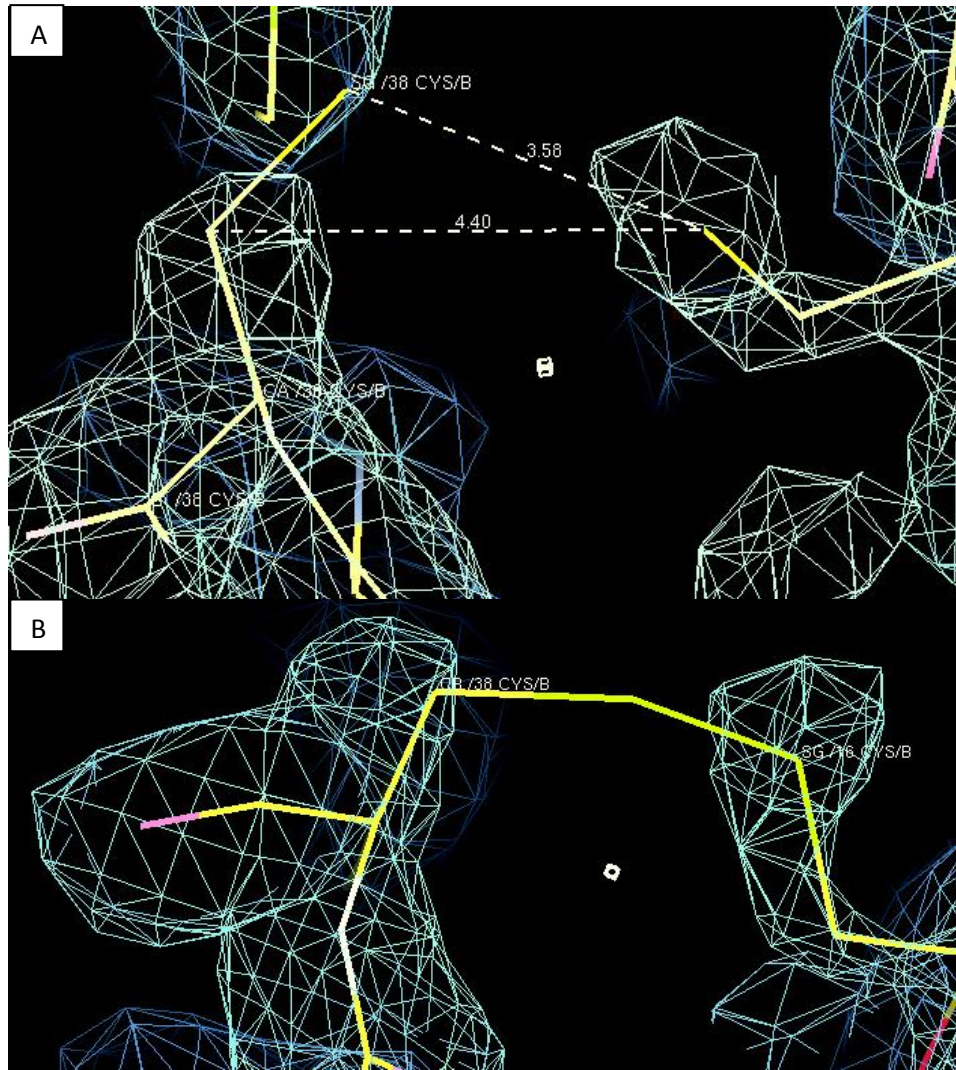


Figure 11 Cysteines in the Disulfide Bridge: By computer imaging it is possible to mutate the 38th residue from an alanine to a cysteine. Panel A is a measurement of the distance between the 16th and 38th residues. The distance is 3.58Å from sulfur to sulfur and 4.40Å from sulfur to β -carbon. Further computer simulation finds a conformation favorable for the disulfide bond formation as seen in panel B.

6.0 Tables

Table 1 Crystallographic Data Statistics: The following data was collected after refinement of the x-ray diffraction data and describes the obtained protease structure.

Space Group	P 2 ₁ 2 ₁ 2 ₁
Cell parameters	-----
a	51.98 Å
b	57.38 Å
c	60.68 Å
Rwork %	17.40%
Rfree%	22.00%
RMS deviation from ideality	-----
Bonds (Angstroms)	0.0086
Angle distance (Angstroms)	1.265
reflections used	14074
optical resolution	1.53 Å
Overall β factor	28.2 Å ²

Table 2 Mutations of the Protease: Mutations of the HIV protease construct as provided by the Schiffer lab are as follows, the two listed in bold are the mutations of interest for creating a disulfide bond.

Q7K
L33I
L63I
C67A
C95A
G16C
L38A

References

1. World Health Organization. Global summary of the AIDS epidemic december 2007. 2007.
2. HIV/AIDS in the United States [Internet]: Center for Disease Control and Prevention; cAugust 21, 2009 [cited 2009 October 3]. Available from: <http://www.cdc.gov/hiv/resources/factsheets/us.htm>.
3. Fields BN, Knipe DM, Howley PM. Fields virology. 5th ed. Philadelphia: Wolters Kluwer Health/Lippincott Williams & Wilkins; 2007. editors-in-chief, David M. Knipe, Peter M. Howley ; associate editors, Diane E. Griffin ... [et al.].; 2 v; Includes bibliographical references and index.
4. Popovic M, Sarngadharan MG, Read E, Gallo RC. Detection, isolation, and continuous production of cytopathic retroviruses (HTLV-III) from patients with AIDS and pre-AIDS. *Science* 1984 May 4;224(4648):497-500.
5. Stebbing J, Moyle G. The clades of HIV: Their origins and clinical significance. *AIDS Rev* 2003;5(4):205.
6. Ali A, Ghosh A, Nathans RS, Sharova N, O'Brien S, Cao H, Stevenson M, Rana TM. Identification of flavopiridol analogues that selectively inhibit positive transcription elongation factor (P-TEFb) and block HIV-1 replication. *Chembiochem* 2009 Aug 17;10(12):2072-80.
7. Frankel AD, Young JA. HIV-1: Fifteen proteins and an RNA. *Annu Rev Biochem* 1998;67:1-25.
8. Preston BD, Poiesz BJ, Loeb LA. Fidelity of HIV-1 reverse transcriptase. *Science* 1988 Nov 25;242(4882):1168-71.
9. Clavel F, Hance AJ. HIV drug resistance. *N Engl J Med* 2004;350:1023-1035.
10. Warnke D, Barreto J, Temesgen Z. Antiretroviral drugs. *J Clin Pharmacol* 2007 December 1;47(12):1570-9.
11. Nalam MN, Peeters A, Jonckers TH, Dierynck I, Schiffer CA. Crystal structure of lysine sulfonamide inhibitor reveals the displacement of the conserved flap water molecule in human immunodeficiency virus type 1 protease. *J Virol* 2007 Sep;81(17):9512-8.
12. Chellappan S, Kairys V, Fernandes MX, Schiffer C, Gilson MK. Evaluation of the substrate envelope hypothesis for inhibitors of HIV-1 protease. *Proteins* 2007 Aug 1;68(2):561-7.
13. Kolli M, Lastere S, Schiffer CA. Co-evolution of nelfinavir-resistant HIV-1 protease and the p1-p6 substrate. *Virology* 2006 Apr 10;347(2):405-9.

14. Muzammil S, Ross P, Freire E. A major role for a set of non-active site mutations in the development of HIV-1 protease drug resistance. *Biochemistry* 2003 Jan 28;42(3):631-8.
15. Chellappan S, Kiran Kumar Reddy GS, Ali A, Nalam MN, Anjum SG, Cao H, Kairys V, Fernandes MX, Altman MD, Tidor B, Rana TM, Schiffer CA, Gilson MK. Design of mutation-resistant HIV protease inhibitors with the substrate envelope hypothesis. *Chem Biol Drug Des* 2007 May;69(5):298-313.
16. Altman MD, Ali A, Reddy GS, Nalam MN, Anjum SG, Cao H, Chellappan S, Kairys V, Fernandes MX, Gilson MK, Schiffer CA, Rana TM, Tidor B. HIV-1 protease inhibitors from inverse design in the substrate envelope exhibit subnanomolar binding to drug-resistant variants. *J Am Chem Soc* 2008 May 14;130(19):6099-113.
17. Prabu-Jeyabalan M, Nalivaika E, Schiffer CA. How does a symmetric dimer recognize an asymmetric substrate? A substrate complex of HIV-1 protease. *J Mol Biol* 2000 Sep 1;301(5):1207-20.
18. Anderson J, Schiffer C, Lee SK, Swanstrom R. Viral protease inhibitors. *Handb Exp Pharmacol* 2009;(189)(189):85-110.
19. Louis JM, Ishima R, Torchia DA, Weber IT. HIV-1 protease: Structure, dynamics, and inhibition. *Adv Pharmacol* 2007;55:261-98.
20. Foulkes-Murzycki JE, Scott WR, Schiffer CA. Hydrophobic sliding: A possible mechanism for drug resistance in human immunodeficiency virus type 1 protease. *Structure* 2007 Feb;15(2):225-33.
21. Foulkes-Murzycki JE, Scott WR, Schiffer CA. Hydrophobic sliding: A possible mechanism for drug resistance in human immunodeficiency virus type 1 protease. *Structure* 2007 Feb;15(2):225-33.
22. Lefevre F, Remy MH, Masson JM. Alanine-stretch scanning mutagenesis: A simple and efficient method to probe protein structure and function. *Nucleic Acids Res* 1997 Jan 15;25(2):447-8.
23. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics [computer program]. *Acta Crystallographica Section D-Biological Crystallography* 60: 2126-2132 Part 12 Sp. Iss. 1: 2004. .
24. Potterton EB, Peter, Turkenburg M, Dodson E. A graphical user interface to the CCP4 program suite [computer program]. *Acta Cryst. D*59, 1131-1137: 2003. .
25. Surleraux DL, Tahri A, Verschueren WG, Pille GM, de Kock HA, Jonckers TH, Peeters A, De Meyer S, Azijn H, Pauwels R, de Bethune MP, King NM, Prabu-Jeyabalan M, Schiffer CA, Wigerinck PB. Discovery and selection of TMC114, a next generation HIV-1 protease inhibitor. *J Med Chem* 2005 Mar 24;48(6):1813-22.

26. HIV-1 Global Distribution [Internet]; c2006 [cited 2009 October 2]. Available from: <http://www.pbs.org/wgbh/pages/frontline/aids/atlas/clade.html>.
27. World Health Organization. World : HIV estimated prevalence among population aged 15-49 years (%), 2007. June 23, 2009(World):1.
28. University of Washington. HIV life cycle. 2006.
29. Louis JM, Dyda F, Nashed NT, Kimmel AR, Davies DR. Hydrophilic peptides derived from the transframe region of gag-pol inhibit the HIV-1 protease. *Biochemistry* 1998 Feb 24;37(8):2105-10.
30. Moreland JL, Gramada A, Buzko OV, Zhang Q, Bourne PE. The molecular biology toolkit (MBT): A modular platform for developing molecular visualization applications. *BMC Bioinformatics* 2005 Feb 6;6:21.