LIST OF FIGURES

Figure 16: Changes in the hydrophobic interactions of NFV with protease before and after simulation. A. hydrophobic interactions with DBM in Glide docked structure before simulation. B. After simulation

Figure 17: CASTP representation of Wild protease cavity size

Figure 18: CASTP representation of DBM cavity size

Figure 19: CASTP representation of TPM cavity size

Figure 20: RMSD trajectory of Modelled Wild protease, DBM_M and TPM_M during MD simulations. Trajectory for Wild protease (red line), DBM_M (green line) and TPM_M (purple line).

Figure 21: Residue wise RMS fluctuations of modeled Wild protease (red line), DBM_M (green line) and TPM_M (purple line). (A) chain A. (B) chain B

Figure 22: I50/I50' Distance plot in wild protease (red line), DBM_M (green line) and TPM _M (purple line)

Figure 23: Flap mutant of Wild protease (1HVP), DBM_M and TPM_M

Figure 24: A detailed view of covalent and non- covalent interactions in the comparison network of wild protease and DBM.

Figure 25: A detailed view of covalent and non- covalent interactions in the comparison network of wild protease and TPM

Figure 26 : Chemical structure of NFV-A

Figure 27: Hydrogen bonds between NFV-A and wild protease (A) and NFV resistant protease DBM (B)

Figure 28: Hydrogen bonds between NFV-A and wild protease (A) and NFV resistant protease DBM (B)

LIST OF TABLES

LIST OF ABBREVIATIONS

AIDS: Acquired Immunodeficiency Syndrome

ARS: Anti Retroviral Syndrome

ART: Anti Retroviral Treatment

DBM: Double Mutant protease

FEP: Free Energy Pertubation

FEZ-1: Fasciculation and Elongation Protein zeta-1

HAART: Highly Active Anti Retroviral Treatment

HIV: Human Immunodeficiency Virus

HTVS: High Throughput Virtual Screening

LFA-1: Lymphocyte Function Antigen-1

MD: Molecular Dynamics

NFV: Nelfinavir

NFV-A : Nelfinavir Advanced

NNRTIs : Non Nucleoside Reverse Transcriptase Inhibitors

NRTI: Nucleoside or Nucleotide Reverse Transcriptase Inhibitors

OIs: Opportunistic Infections

PDB: Protein Data Bank

PI: Protease Inhibitors

PR: Protease

RINs: Residue Interaction Network

TPM: Triple Mutant Protease

TI: Thermodynamic Integration

WHO: World Health Organisation

XP: eXtra Precision docking

COMPUTATIONAL CHARACTERIZATION OF NON-ACTIVE SITE MUTATION V77I IN HIV-1 PROTEASE: POSSIBLE CONTRIBUTION TO NELFINAVIR RESISTANCE AND DEVELOPMENT OF NEW DRUG LEADS TARGETING HIV-1 PROTEASE

ANKITA GUPTA

Delhi Technological University, Delhi, India

ABSTRACT

BACKGROUND

The Human immunodeficiency virus (HIV-1) protease is an attractive target for antiviral treatment and a number of therapeutically useful inhibitors have been designed against it. The emergence of drug resistant mutants of HIV-1 pose a serious problem for the conventional therapies been used so far. Here we have tried to study the effect of V77I mutation along with the co-occurring mutations L33F and K20T through multinanosecond molecular dynamics simulations. V77I is known to cause Nelfinavir (NFV) resistance in subtype B population of HIV-1 protease. We have reported the effect of this clinically relevant mutation on the binding of NFV and the conformational flexibility of the protease, and tried to generate derivates of potent drug Nelfinavir which can efficiently inhibit the wild and mutant proteases.

RESULT

The study proposes that V77I-L33F mutant (DBM) showed greater flexibility and the flap separation was more with respect to the wild protease. The cavity size of stabilized DBM was also found to be increased which is responsible for the decreased interaction of Nelfinavir with all the cavity residues and hence decreased its binding affinity (Glide XP score: wild= -9.3, DBM= -7.8). On the other hand the binding affinity of V77I-L33F-K20T mutant (TPM) was found to be increased for Nelfinavir (Glide XP score= -10.3). The flap separation of TPM was less and the cavity size had also reduced with respect to wild protease.

CONCLUSION

The resistant mutations had made DBM more stable in environment whereas the addition of third mutation K20T had made the protease TPM more susceptible to Nelfinavir. This lowered resistance can be the reason behind the less clinical relevance of TPM.

INTRODUCTION

Human immunodeficiency virus type 1 (HIV-1), a member of retrovirus family is one of the most dangerous viruses for mankind ("HIV epidemic -- a global update. Excerpts from the UN World AIDS Day report," 1998; Mitsuya et al., 1990). Where viral proteins like aspartic protease, reverse transcriptase and fusion proteins serve as active targets for anti-HIV-1 therapy, accumulating resistance towards different drugs in the Highly Active Anti-Retroviral Therapy (HAART) cocktails seems like an exponentially growing problem (Little et al., 2002).

HIV-1 proliferates with the support of its own homodimeric aspartic protease (HIV-1 PR), an enzyme essential for viral replication and assembly (Krausslich & Wimmer, 1988; Turner & Summers, 1999). The function of protease is assisted by characteristic flap movement which provides restricted access to the active site. The flaps adopt a semi-open conformation in unbound state, whereas they are pulled in the active site to form closed structure in bound state . Functional inhibition of HIV-1 PR leads to incomplete viral replication and therefore makes it an attractive target for anti-HIV drugs (Kohl et al., 1988). Till now seven protease inhibitors (PIs) have been approved by FDA (Wlodawer & Vondrasek, 1998). However, the evolution of resistant viral species due to genetic mutation in active and non active sites of HIV-1 PR, arising as a consequence of the selective pressure rising due to antiviral agents, serve as major problems faced by the current therapies (Leslie et al., 2005).

V77I is one of the non- active site secondary mutations causing resistance against NFV. It is highly polymorphic mutation near the cheek sheet of protease, with marked presence in subtype B virus. This minor mutation is accompanied with other primary and secondary mutations (Rhee et al., 2003). L33F is a major mutation present in the active site of protease (Jallow et al., 2009) , it shows reduced susceptibility towards NFV in the presence of other mutations. It co-occurs with V77I in large number of HIV-1 subtype B infected patient samples (Rhee et al., 2003). K20T is another mutation (Van Marck et al., 2009) co-occurring with V77I in subtype B population. The 77th,33rd and 20th aminoacids form a set of residues interacting with the 36th residue of protease, which itself is present on non-active site and its mutation causes resistance to NFV in non-subtype B viruses (Ode, Matsuyama, Hata, Neya, et al., 2007).

Here we scrutinize the behaviour of minor mutation V77I along with the co-occurring mutations L33F and K20T. We have considered two types of mutants for our study according to their actual prevalence (Rhee et al., 2003) (Table 1): first one is a double mutant, V77I-L33F (DBM); and the second is a triple mutant, V77I-L33F-K20T (TPM).

Table 1: NUMBER OF SUBTYPE-B CLINICAL ISOLATES REPORTED IN HIV DRUG RESISTANCE DATABASE AS ON 20TH FEB, 2014.

MUTANT NAME	MUTATION SET	ISOLATES
DBM	V77I-L33F	407
TPM	V77I-L33F-K20T	16

The structural roles of the mutations are studied on an atomic level with the aid of molecular-dynamics simulation of the wild and mutant HIV-1 PRs for the unbound confirmation and docked complex with NFV. The protease-ligand (NFV) interaction energies were calculated for wild and mutant proteases using MM/GBSA approach. Similar studies have been carried out before, reporting molecular mechanisms underlying several drug resistance and provide valuable insights into the mode of interaction of mutants versus wild (Batista et al., 2006; Meiselbach et al., 2007; Ode, Matsuyama, Hata, Hoshino, et al., 2007; Ode et al., 2006; Ode et al., 2005; Piana et al., 2002; Skalova et al., 2006).

REVIEW OF LITERATURE

3.1 HIV

Human Immunodeficiency Virus, as the name represents can particularly infect only **H**uman beings thereby weakens the **I**mmune system of the host and as a **V**irus can reproduce only in the cells of infected host. It is a Lentivirus, subgroup of rotavirus family (Weiss, 1993). The first cases of HIV were detected in 1983, since the infection has prevailed throughout the world population with major spread in South African populations.

How HIV is different from other viruses? HIV is very much similar to other common viruses including Influenza (causes common cold) but it presents a heavy toll on the immune system of host. The human immune system can fight most of the viruses but HIV destroys important immune cells (T cells) thereby making it deficit to fight it back (Douek et al., 2009). It simply means that once HIV infects! The infection will persist till the death of the host. HIV is transferred between the hosts through bodily fluids like blood, semen, breast milk etc, where it is present either in infected host immune cells or as free viral particle.

Although not every time (can be prevented through antiretroviral therapy), the final stages of HIV leads to AIDS i.e. Acquired Immunodeficiency Syndrome. Patients with AIDS have extremely damaged immune system which makes them prone to opportunistic infections (OIs) and Cancers. The survival time of HIV infected patient is expected to be 9-11 years after infection, with antiviral treatment (Garg et al., 2012).

3.1.1 HIV Classification

Human immunodeficiency virus type 1 (HIV-1), a member of retrovirus family is one of the most dangerous viruses for mankind (Mitsuya et al., 1990). HIV-1 has been subdivided into major group (Group M) and few minor groups. There are 39 Open Reading Frames (ORFs) in total, found in the six reading frames of HIV-1 genome, but all of these frames are not functional. Group M is the Major group and predominates in type 1 pathogenic groups. It is further divided into subtypes A-K and Circulating recombinant forms (CRF, derived from the recombination between different subtypes (Robertson et al., 1995)) (Figure 1).

Subtype B is the most prominent type 1 virus among the HIV-1 infected people, mainly in America, Japan and Europe ("HIV epidemic -- a global update. Excerpts from the UN World AIDS Day report," 1998). Whereas, Group C is found in Southern Africa, Eastern Africa, India, Nepal, and parts of China.
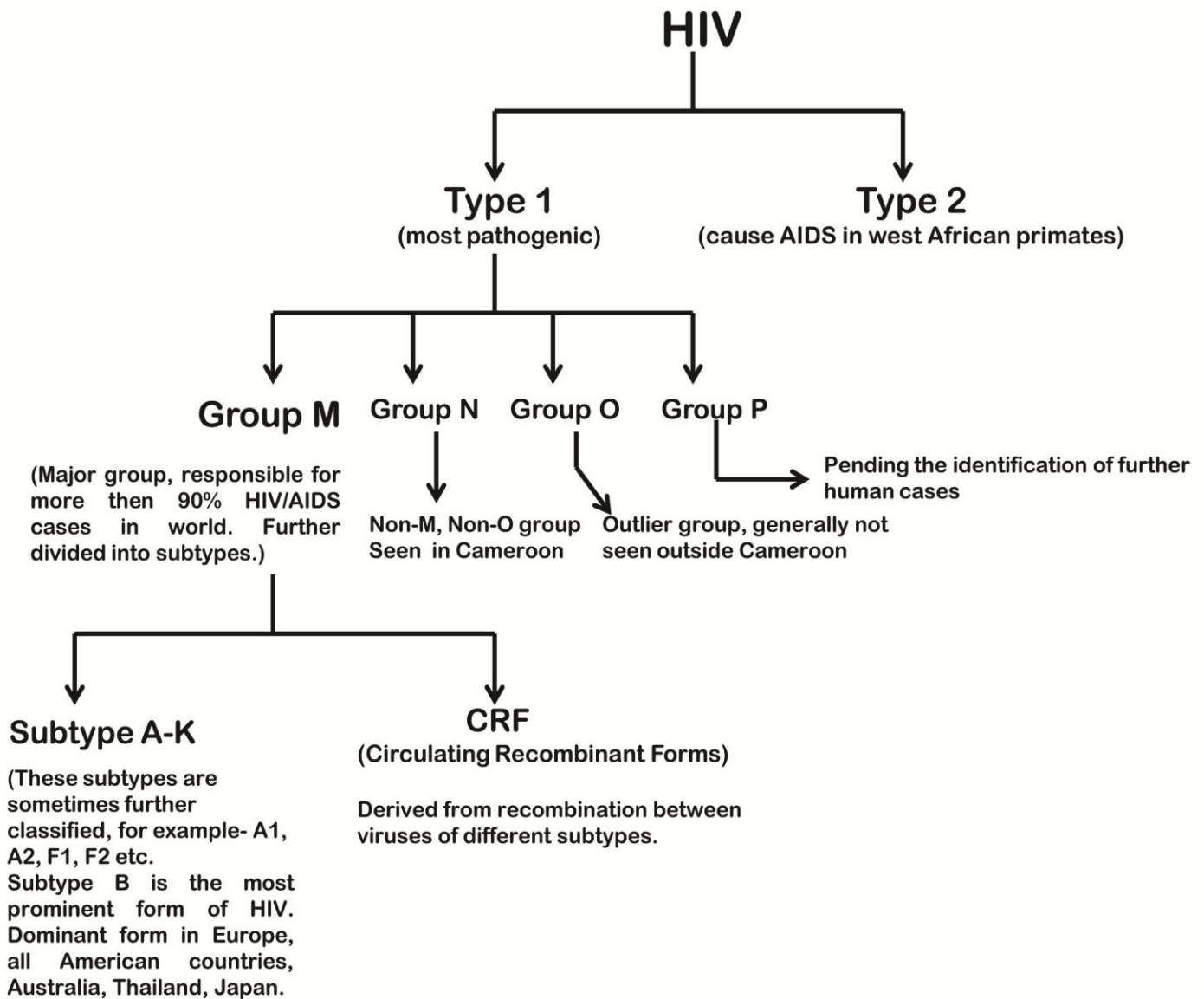
Figure 1: Classification of Human Immunodeficiency Virus.

3.1.2 Overview of HIV Infection

HIV weakens the immune system and makes the body more susceptible for other opportunistic infections. The viral infects vital immune system cells- CD$^+$ T cells, macrophages and dendritic cells (Garg et al., 2012). A cascade of intracellular events occurs after the attachment of HIV to the immune cells. These events finally lead to apoptosis of infected cells and release of new viral particles in the body.

ATTACHMENT TO LEUKOCYTE MEMBRANE

HIV enters the cell through adsorption of its glycoproteins by the receptors present on the host cell (Chan & Kim, 1998). This interaction is initiated when the viral trimeric envelope complex interacts with CD4 and the chemokine receptors on the host cell, simultaneously (Wyatt & Sodroski, 1998). The efficient cell to cell spreading of the infection is mediated through Lymphocyte Function Antigen-1 (LFA-1), which is activated when glycoprotein gp120, binds to integrin $\alpha 4\beta 7$ (Arthos et al., 2008). The attachment of CD4 to gp120 leads to structural changes of the viral envelope leading to glycoprotein gp160, gp160 has the domains corresponding to CD40 and chemokine receptors which lead to a two-pronged attachment and a more stable attachment. This allows gp41 to penetrate through the host cell membrane. Then the collapse of viral capsid is caused by interaction of gp41, HR1 and HR2. The successful binding of HIV to the target is followed by transfer of various viral enzymes- reverse transcriptase, integrase, ribonuclease, protease and the viral RNA, to the host cell.

An alternate route of viral attachment to the host is through mannose specific C-type lectin receptors (Pope & Haase, 2003). The presence of Fasciculation and elongation protein zeta-1 (FEZ-1) on neurons are thought to prevent HIV from infecting them (Haedicke et al., 2009).

REPLICATION AND TRANSCRIPTION

The single stranded RNA genome of HIV is transcribed into double stranded DNA through the enzyme reverse transcriptase, during the microtubule associated transport to the nucleus. This is followed by its integration in the host genome (Zheng et al., 2005). The process of formation of complementary DNA (cDNA) from the positive RNA viral genome is highly susceptible to errors. This may be responsible for generation of drug resistant mutations in the viral proteins. The RNA strand is simultaneously degraded by the reverse transcriptase through its ribonuclease activity. The double stranded viral DNA (complementary strand generated through DNA dependent DNA polymerase enzyme which acts on cDNA to create 'sense' strand) is finally transported to the nucleus where the integrase enzyme finally acts. Cellular transcription factors (NF-κB) are required to bring otherwise dormant viral DNA (latent stage) to active stage (Hiscott et al., 2001). The transcribed viral mRNA is spliced before it is exported to the cytoplasm where they are translate to viral protein Tat and Rev. Rev binds to new mRNAs and assist there transport without cleavage. Full length mRNAs are packaged into then produced Gag and Env proteins to finally form new virus particles. HIV-2 will preferentially bind to the mRNA that was used to create the Gag protein itself. This may mean that HIV-1 is better able to mutate (HIV-1 infection progresses to AIDS faster than HIV-2 infection and is responsible for the majority of global infections).

RECOMBINATION

One HIV-1 particle consists of two RNA genomes, among which recombination can occur during replication catalyzed by reverse transcriptase (Hu & Temin, 1990). The nascent DNA formed during reverse transcription can switch between the two copies of RNA number of times. This leads to changes or shuffling in the information flowing from parental to progeny genome. This is called copy-choice type of recombination event and may occur between 2 to 20 times in one replication cycle, throughout 1 genome (Charpentier et al., 2006). Such recombination events helps to produce

variations which later assist in escaping the host immune defense system and may also play crucial role in generating species resirstant to anti-HIV drugs (Nora et al., 2007) . For this kind of recombination events it is necessary that the two RNA genomes in infecting virus should be from different progenitor parental viruses (Chen et al., 2006). Copy-choice recombination is an adaptation for repair of genomes damaged due to chronic ongoing inflammation and reactive oxygen species produced by HIV-1 infection (Israel & Gougerot-Pocidalo, 1997; Michod et al., 2008).

ASSEMBLY AND RELEASE.

Finally the new HIV virions assemble at the plasma membrane. Gp160 is cleaved to two envelope proteins- gp41 and gp120 in Golgi complex by the action of Furin (Hallenberger et al., 1992). These glycoproteins are then transported to plasma membrane of infected cell. Gag and Pol proteins get associated with plasma membrane and the forming virion starts to bud out. The gag polyproteins are activated in the budded virion by HIV serine protease which makes the virion mature to infect other cells.

3.1.3 Stages of HIV infection

HIV infection progresses through pre-determined stages in the host body. However the rate of progress through these stages varies with infected patients on the basis of varied factors. These factors include their genetic makeup, health conditions before catching infection, length of period between infection and diagnosis, medication and quality of treatment received (Figure 2).

ACUTE INFECTION STAGE

The patient suffers through acute retroviral syndrome (ARS) within a month after infection, marked by sore throat, fever, rashes, headache, joint aches. In this stage the host immune cells (CD4+) serve as rapid manufacturer of virus particles and get destroyed in the same process. It is beneficial to begin ART at this time, there is also large risk of transmitting HIV to sexual or drug using partners since the HIV load is very high in body fluids at this stage.

CLINICAL LATENCY STAGE

Stage is also known as Asymptomatic HIV infection or Chronic HIV infection. As the latency term suggests, the virus lives in the host without causing any viral specific symptoms. The infected patients experience very mild or no symptoms of infection. ART basically helps in extending the clinical latency period for years, where the viral reproduction rate is very low but the patient can still transmit infection.

AIDS

Finally when the host immune system becomes completely damaged- CD4 cells count fall below 200 cells per cubic millimetre of blood and body becomes sensitive to opportunistic infections; the stage is known as Acquire Immunodeficiency Syndrome. It may or may not be marked by decreased T-cell count but necessarily includes development of one or more opportunistic illness.

STAGES OF HIV INFECTION

LATENCY PERIOD
Large amount of virus particles are produced.
Flu like symptoms also called "Worst flu ever"

LATENCY PERIOD
Very low level of HIV reproduction.
No or very low clinical symptoms.
Can extend for decades with help of ART.

AIDS
Heavy decrease in level of CD4+ cells.
Host becomes susceptible to other opportunistic infections.
Patient can survive for average 3 years without ART.

Figure 2 : Pictorial representation of Stages in HIV infection

3.1.4 HIV Prevalence

HIV is world's leading infectious killer since past 31 years. Almost 36 million people have died and 75 million have suffered from this infection ever since the first cases were diagnosed in 1981. As per UNAIDS GLOBAL FACTSHEET of 2013, the new infections rate has decreased, by 33% since 2001 and by 52% in children since 2001. Tuberculosis remains the major cause of death of patients suffering from HIV but the rate has decreased by 36% since 2004. It is estimated that around 32.2-38.8 million people in the world were suffering from this infection by the end of 2012. Sub- Saharan Africa is the most affected region where every 1 out of 20 adult is suffering from HIV and accounts upto 71% of HIV patients in world.
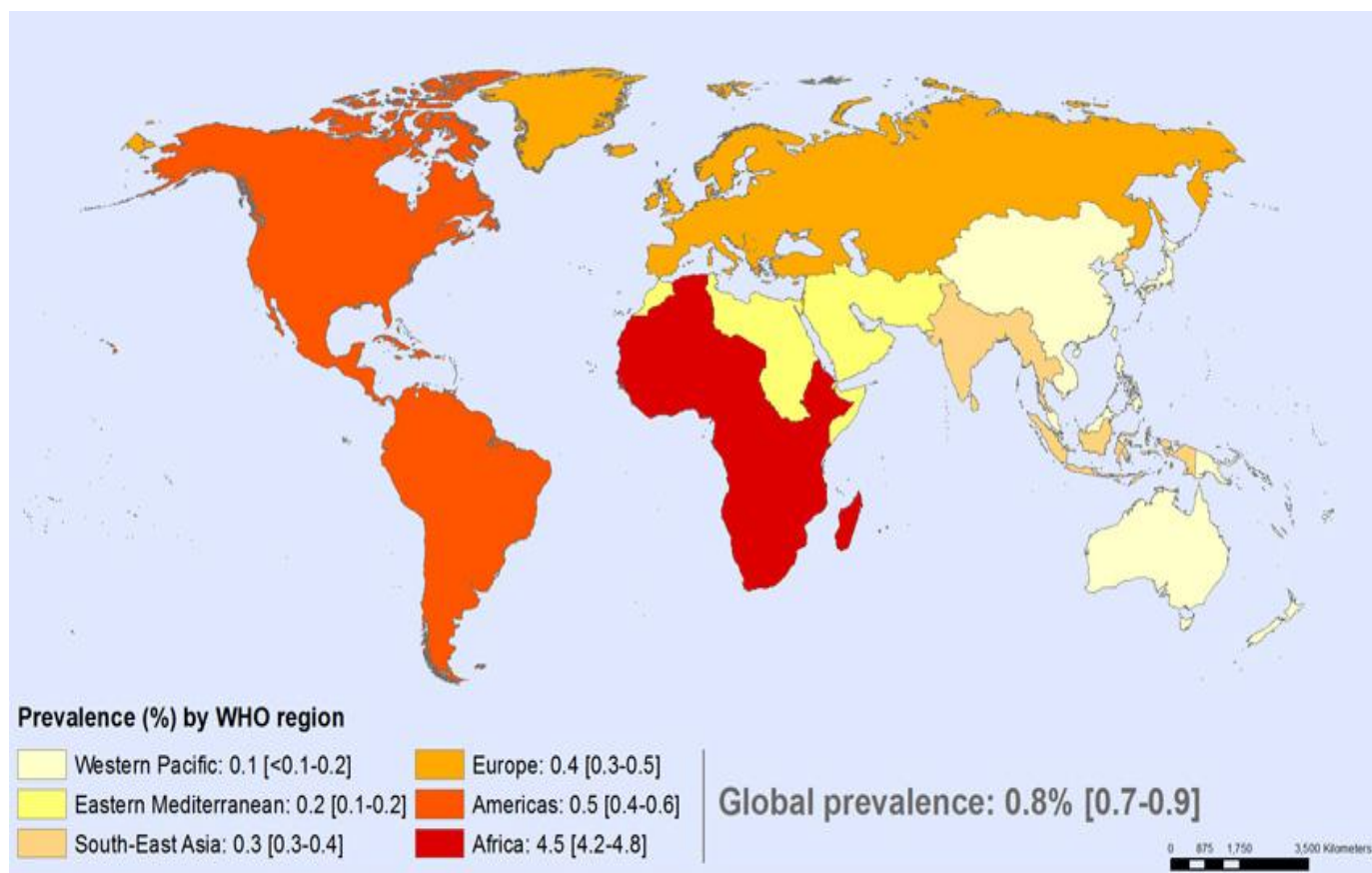
Figure 3: Prevalence (%) by WHO region, Adult HIV prevalence (14-49 years) by WHO 2013

HIV IN INDIA

The first HIV case was reported in 1986 in India and since then it has been detected in all states and union territories however the infection is not uniformly seen in whole India. The infection is more prominent in Southern regions of India mainly through heterosexual contacts. The infection is more severe in north eastern regions of India and is mainly found in injecting drug users (IDU) and sex workers (NACO, 2011). But major regions have been reported to have very low rate of infection (NACO, 2012).

3.1.5 HIV Treatment

Before the introduction of anti-retroviral therapy in 1990's, patients would progress to AIDS in just few years. But now days with the help of anti-retroviral drugs, if the infection is diagnosed at early stages the patient can expect normal life expectancy. The standard ART is a cocktail of least three anti-retroviral drugs to prevent the progress of infection in the host with prominent results shown when administered at early stages of infection. WHO suggests ARV treatment particularly for infected pregnant women and young children to sequester the spread of new infections in population. Since 1987, around 30 drugs have been approved for HIV/AIDS treatment and many are under development and clinical trials (Table 2). According to the target of the drugs and the stage in which they prevent viral growth and development, the ARVs have been classified into five classes. The

right cocktail is made by picking the right combination of ARVs from different classes. ARV classes are:

1. NRTIs

Nucleoside or Nucleotide Reverse Transcriptase Inhibitors are also known as NUKES. These drugs mimic the building blocks of viral DNA production through reverse transcriptase and hence block the formation of virus copies in host cell.

2. NNRTIs

Non Nucleoside Reverse Transcriptase Inhibitors are also known as Non-Nukes. These drugs directly act on reverse transcriptase and prevent formation of viral copies, rather than working on the genetic material of the virus.

3. PIs

Protease Inhibitors block the activity of viral serine protease. This enzyme cleaves the long viral peptides into functional proteins. Inhibition of this causes non-function viral protein production which is not capable of further infection.

4. Fusion Inhibitors

Also called Entry inhibitors, prevent the binding of virus on the CD4+ cells. It does so by binding and modifying the receptors (present either on HIV or host cell) responsible for viral attachment. This leads to unsuccessful binding of virus on host cells, and hence restricts the entry of virus in the host cells.

5. Integrase Inhibitors

The major step in life cycle of HIV infection is integration of viral DNA into the host cell where it can be expressed with the help of host machinery. This is achieved with the help of viral protein enzyme- integrase. Integrase inhibitors block this step and hence inhibit expression of viral proteins in the host system.

Other Complimentary medications are given according to the needs and exposure of the patient. These are given to cure other opportunistic infections or to prevent other side-effects like nausea, pain or diarrhoea. This kind of personalized treatment can be modified daily or weekly depending on the patient's requirements, and is known as Prophylaxis.

Table 2: Drugs employed for HIV treatment.

DRUG CLASS	GENERIC NAME	PHASE OF CLINICAL DEVELOPMENT
NRTI	Lamivudine	FDA approved
	Abacavir	FDA approved
	Zidovudine	FDA approved
	Stavudine	FDA approved
	Didanosine	FDA approved
	Emtricitabine	FDA approved
	Tenofovir	FDA approved
	Apricitabine	Phase III clinical trials
	Tenofovir Alafenamide	Phase III clinical trials
NNTRI	Delavirdine	FDA approved
	Efavirenz	FDA approved
	Etravirine	FDA approved
	Nevirapine	FDA approved
	Rilpivirine	FDA approved
	Etravirine	Approved in Jan. 2008
	Rilpivirine	Approved in May 2011
PI	Amprenavir	FDA approved
	Fosamprenavir	FDA approved
	Atazanavir	FDA approved
	Darunavir	FDA approved
	Indinavir	FDA approved
	Nelfinavir	FDA approved
	Ritonavir	FDA approved
	Saquinavir	FDA approved
	Tipranavir	FDA approved
	Brecanavir	Phase II clinical trials
Fusion Inhibitors	Enfuvirtide	FDA approved
	Maraviroc	FDA approved
	Maraviroc	Approved in Aug. 2007
	Cencriviroc	Phase II clinical trials
Integrase Inhibitors	Raltegravir	FDA approved
	Raltegravir	Approved in Oct. 2007
	Dolutegravir	Approved in Aug. 2013
	Elvitegravir	Phase III clinical trial
	Gsk126744	Phase II clinical trial

3.1.6 HIV Drug Resistance

The ability of Human Immunodeficiency virus, to resist the anti-viral drugs treatment, and reproduce in their presence is known as HIV drug resistance (HIVDR). The consequences of resistance causing mutations in HIV target proteins include treatment failure, costly second and third line treatments, introduction of drug resistant HIV strains in environment leading to requirement of new-effective anti-HIV drugs.

Where viral proteins like aspartic protease, reverse transcriptase and fusion proteins serve as active targets for anti-HIV-1 therapy, accumulating resistance towards different drugs in the Highly Active Anti-Retroviral Therapy (HAART) cocktails seems like an exponentially growing problem (Little et al., 2002). The rapid emergence of drug resistant mutants of has hindered the advantage of conventional anti-retroviral therapy for AIDS. The resistance causing mutations in serine protease of HIV has presented cross resistance and multi drug resistance in the clinical isolates (Hertogs et al., 2000; Shafer et al., 1998; Tamalet et al., 2000). The HIV-PR gene has shown genetic diversity, even in the absence of anti retroviral therapy. The 198 amino acid long protein shows variations in upto fifty different residues (Vergne et al., 2000). There are combinations of these mutations occurring in nature with no single mutation paying major role. However, substitutions in upto 11 residues in protease confer high level of resistance (Brown et al., 1999).

TESTING FOR RESISTANCE

Testing for resistance towards a drug can be performed using two methods. First one is Genotypic assay in which the genetic material of virus is screened to understand the drugs towards which it can be resistant. It is a cost-effective short process of maximum two weeks. The second one is Phenotypic assay, as the name suggests the response to medications is observed in a restricted environment. The results are easy to interpret without help of sophisticated techniques and can take upto three weeks. Helps to detect multi drug resistance when multiple drugs have failed in the treatment.

3.2 HIV PROTEASE (HIV-PR)

HIV-1 proliferates with the support of its own homodimeric aspartic protease, an enzyme essential for viral replication and assembly also referred as HIV-1 protease (HIV-1 PR) (Krausslich & Wimmer, 1988). The recognition of HIV-1 PR as a major target for antiviral therapy has led to determination of its large number of structures with slight sequence variation and different ligands. HIV-1 PR is responsible to render the non-functional polyproteins encoded by HIV-1 genome functional, by cleaving them at appropriate sites, these include gag and pol proteins, reverse transcriptase, integrase, and protease itself (Turner & Summers, 1999). It is a homodimer of 99 aminoacid long sequence that forms C2 symmetry whiles the inexistence of ligand (Kohl et al., 1988). The dimer interface forms the active site of the enzyme, which have two catalytic aspartic-acid residues. The function of protease is assisted by characteristic flap movement which provides

restricted access to the active site. The flaps are flexible anti parallel, glycine rich β-sheets of residues 45-55 from both the chains of the homodimer (Figure 4) (Navia et al., 1989; Vondrasek & Wlodawer, 2002; Wlodawer et al., 1989). Through the X-ray crystallographic studies it is known that, there are consistent structural differences between the bound and free state of protein. The flaps adopt a semi-open conformation in unbound state, whereas they are pulled in the active site to form closed structure in bound state (Louis et al., 1998; Prabu-Jeyabalan et al., 2000).
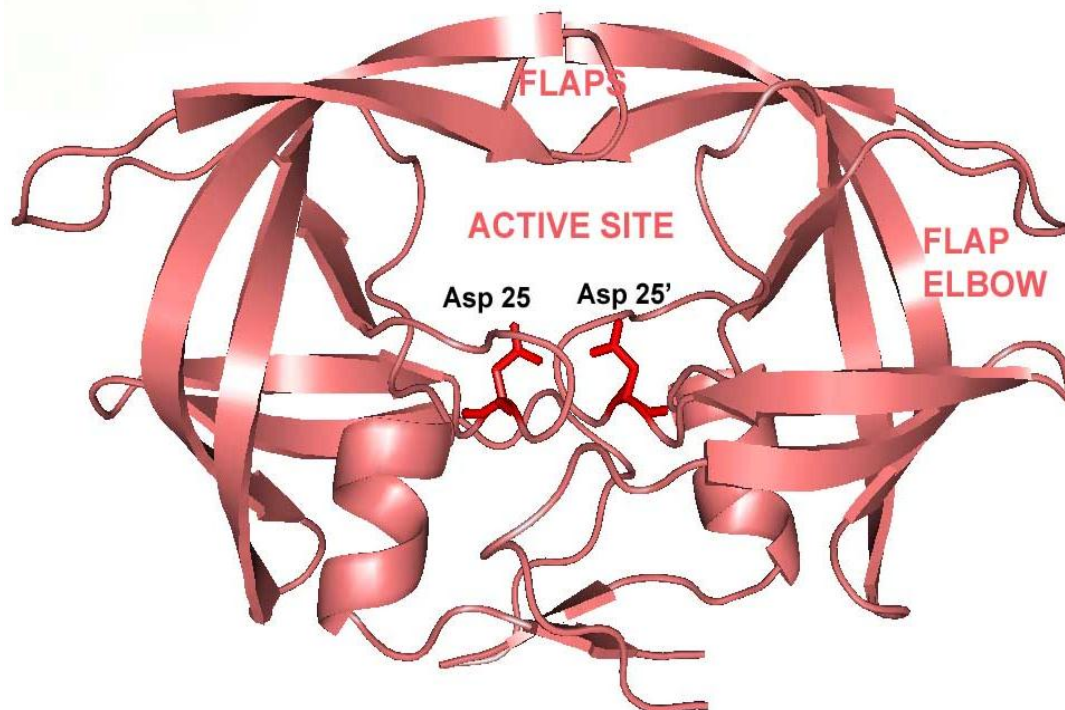
Figure 4: Bound conformation of HIV-PR (PDB: 1OHR). Ligand is removed for clarity

Functional inhibition of HIV-1 PR leads to incomplete viral replication and therefore makes it an attractive target for anti-HIV drugs (Kohl et al., 1988). Till now seven protease inhibitors (PIs) have been approved by FDA (Wlodawer & Vondrasek, 1998). However, the evolution of resistant viral species due to genetic mutation in active and non active sites of HIV-1 PR, serve as major problems faced by the current therapies. Direct resistance is caused by active site mutations also referred as primary mutations. The secondary mutations often accompanies primary mutations (accessory mutations) or they also show synergistic resistance in the presence of other secondary mutations (Arvieux & Tribut, 2005). These mutations are consequence of the selective pressure rising due to antiviral agents. Another driving force of these resistance causing mutations is the recently reported immunological pressure and the mutations are described as 'Escape mutations' (Leslie et al., 2005).

3.3 V77I, L33F and K20T mutations

V77I is one of the non active site secondary mutations causing resistance against nelfinavir (NFV). It is highly polymorphic mutation near the cheek sheet of protease, with marked presence in subtype B virus. This minor mutation is accompanied with other primary and secondary mutations (Rhee et al., 2003). L33F is a major mutation present in the active site of protease (Jallow et al., 2009) , it

17

shows reduced susceptibility towards NFV in the presence of other mutations. It is a non polymorphic mutation which provides resistance against all PIs except Indinavir and Saquinavir. L33I is a less commonly occurring mutation with similar effects to L33F, and L33V mutation has not been related to any kind of drug resistance PI therapy. L33F co-occurs with V77I in large number of HIV-1 subtype B infected patient samples, as reported in Stanford's HIV Drug Resistance Database (Rhee et al., 2003). Mutation at 20th residue is another non-polymorphic site present in the cheek turn and involved in rendering resistance against all PIs except Saquinavir and Tipranavir. K20T is most prominent mutation occurring at the 20th residue (Van Marck et al., 2009) and is found to co-occur with V77I in subtype B population. It is interesting to mention that 77th,33rd and 20th aminoacids form a set of residues interacting with the 36th residue of protease, which itself is present on non-active site and its mutation causes resistance to NFV in non-subtype B viruses (Ode, Matsuyama, Hata, Neya, et al., 2007). L23I, D30N, E35G, M46I/L/V, G48V, I54L, G73S/T/C/A, T74S, V82A/F/S/T, I84V, N88D/S and L90M are other mutations correlated to NFV resistance.

Here we scrutinize the behaviour of minor mutation V77I along with the co-occurring mutations L33F and K20T. We have considered two types of mutants for our study according to their actual prevalence (Rhee et al., 2003) (Table 1): first one is a double mutant, V77I-L33F (DBM); and the second is a triple mutant, V77I-L33F-K20T (TPM). The structural roles of the mutations are studied on an atomic level with the aid of molecular-dynamics simulation of the wild and mutant HIV-1 PRs for the unbound confirmation and docked complex with NFV. The protease-ligand (NFV) interaction energies were calculated for wild and mutant proteases using MM/GBSA approach. Similar studies have been carried out before, reporting molecular mechanisms underlying several drug resistance and provide valuable insights into the mode of interaction of mutants versus wild (Batista et al., 2006; Meiselbach et al., 2007; Ode, Matsuyama, Hata, Hoshino, et al., 2007; Ode et al., 2006; Ode et al., 2005; Piana et al., 2002; Skalova et al., 2006).

3.4 Molecular dynamics simulations

While crystallographic studies like these convincingly demonstrate the important role protein flexibility plays in ligand binding, the expense and extensive labour required to generate them have led many to seek computational techniques that can predict protein motions. In order to reproduce the actual behaviour of real molecules in motion, the energy terms are parameterized to fit quantum-mechanical calculations and experimental (for example, spectroscopic) data. Properties bonds and atomic angles like stiffness of bonds and lengths of springs, vander Waals atomic radii etc are applied in this parameterization. The clubbed form of these parameters defines 'force fields', the force fields which describe the role of the atomic forces during MD simulations. The commonly used force fields are GROMOS, CHARMM, OPLS and AMBER. Although these force fields give similar results, they have different principals of parameterizing the atomic forces. As the forces are calculated for every atom in the system, their motion is then correlated to the Newton's laws. After every such movement, the simulation time is advanced and then the atomic forces are again calculated for the new positions so as to predict the next movements in the system. Each step is of time period of around one quadrillionth of a second and hence simulations require heavy computing

access. simulations are therefore carried on supercomputers with the number of parallel processors according to the time of simulation required and the size of system being simulated. Message Passing Interface (MPI) is compatible with most of the commonly used simulation software packages. MPI facilitates execution of one task (complex task) by one software application on more than one processors working in parallel by enabling computer to computer messaging. this computation technique called MD simulation has been compared with experimental data in a number of studies and has been proved equivalent. NMR data are particularly useful, as the many receptor and ligand conformations sampled by molecular dynamics simulations can be used to predict NMR measurements like spin relaxation, permitting direct comparison between experimental and theoretical techniques. Indeed, a number of studies have shown good agreement between computational and experimental measurements of macromolecular dynamics.

3.5 *In-Silico* Docking and Screening

The availability of a protein target structure is usually helpful in identifying potential interacting drugs. Docking and screening approaches assist in explicit docking of compounds into the binding site of receptor. They predict binding mode and conformation of the compound into the receptor cavity by molecular level of analysis (Lyne, 2002). This is a difficult process and implies use of heavy algorithms to predict different poses of the ligands. For this it is necessary to know the structure and the binding site of the target protein. this can be deciphered through X-ray crystallography, NMR studies or structure prediction tools using Homology modeling and *ab-initio* techniques. Then the next step is to deduce correct pose of the compound in the active site of target protein (Taylor et al., 2002), this is done with the help of incremental algorithms capable of search optimal poses of the ligand.

Fragments are placed in the binding site of proteins and then 'grown' to fill the space available. An example of such approach has been reported by Rarey and colleagues(Rarey et al., 1996), in which the conformational space of the ligand is sampled on the basis of a discrete model and a tree-search technique is used for extending the ligand within the active site. Boehm and coworkers(Boehm et al., 2000) applied the use of needle screening to identify compounds that bind to the bacterial enzyme DNA gyrase ATP binding site. There are also an increasing number of reports on the use of Monte Carlo procedures for protein modeling and design. An early use of such procedure was described by Abagyan and Totrov(Abagyan & Totrov, 1994), which randomly selects a conformational subspace and makes a step to a new position independent of the previous position, but according to the predefined continuous probability distribution. The use of conformational ensembles and genetic algorithms to predict the bound conformations of flexible ligands to macromolecular targets was also explored A comparative evaluation of eight docking programs (DOCK, FlexX, FRED, GLIDE, GOLD, SLIDE, SURFLEX and QXP) for their capacity to recover the X-ray pose of 100 small-molecular-weight ligands was reported (Kellenberger et al., 2004). It was found that at a 1 Å r.m.s.d. threshold, docking was successful for up to 63% of cases, while at an r.m.s.d. threshold of 2 Å, the maximum success rate was 90%. Third, the system must evaluate the relative goodness-of-fit or how well the compound can bind to the receptor in comparison with other compounds. An early venture was described by Platzer and colleagues(Platzer et al., 1972), on calculating the relative standard free energy of binding of substrates to α-chymotrypsin. At that time, computational limitations did not allow the inclusion of solvation or entropic effects in the simulations. Since then, new methods have been devised which allow the basic handling of such configurations. Physical-based potentials

uses atomic force fields to model free energies of binding, and may be coupled with methods such as free energy perturbation (FEP) and thermodynamic integration (TI) for higher accuracy. Empirical-based potentials are fast and hence widely used in most docking algorithms. Such an approach requires the availability of receptor–ligand complexes with known binding affinity, and uses additive approximations of several energy terms such as van der Waals potential, electrostatic potential, hydrophobicity potential, among others, for binding free energy estimations.

METHODOLOGY

4.1 Protein preparation

The protease-drug complex structure (PDB: 1OHR) (Kaldor et al., 1997) was taken as starting structure from Protein Data Bank (Berman et al., 2000) . HIV-1 PR − Nelfinavir crystal structure was pre-processed using ViewerLite, a visualizing tool from Accelrys (Accelrys, Inc., San Diego, CA, USA). The ligand NFV and the water molecules were removed from the structure to obtain unbound structure in a traditional manner. The protein was further prepared and optimized using Schrödinger's protein preparation wizard (Schrodinger, 2011). The residues V77 and L33 were mutated to isoleucine and phenyalanine to obtain double mutant DBM (Figure 5) through the protein preparation wizard. Similarly triple mutant TPM (Figure 5) was obtained by mutating V77, L33 and K20 to Isoleucine, phenylalanine and threonine respectively. Preparation of structures involved addition of hydrogen bonds, creation of disulfide bonds, removal of bad contacts, capping of protein terminals, optimization of bond lengths, conversion of selanomethionine to methionine and cleaning the geometry of overlapping residues. Side chain prediction and refinement of selected residues was done using PRIME module provided by Schrodinger (Prime, version 2.1, Schrödinger, LLC, New York, NY, 2009). The study of flap movements is crucial to understand and compare the molecular dynamics of wild and mutant HIV-1 PRs. In the ligand bound crystal structure, the flap residues are involved in interactions with the inhibitor or natural substrate   (Miller et al., 1989) (Hornak et al., 2006). Whereas the flaps of unliganded protease were found to be highly flexible (Ishima et al., 1999) with rapid conformational changes at a time scale of less than one nanosecond. To analyse the flap motions of our mutants with respect to the wild proteases we selected the unliganded modeled structure (Weber et al., 1989) of HIV-1protease [PDB: 1HVP]. The double (DBM_2) and triple (TPM_2) mutants were prepared similar to the procedure described before. The drug nelfinavir (NFV) (CID 64143) (Figure 6), was processed before docking using LigPrep's ligand preparation protocol (Ligprep v2.5; Schrödinger, Inc.: Portland, OR, 2011). The three dimensional coordinates (tautomeric, stereochemical, ionizing variants) were generated along with their energy minimization and flexible filtering.

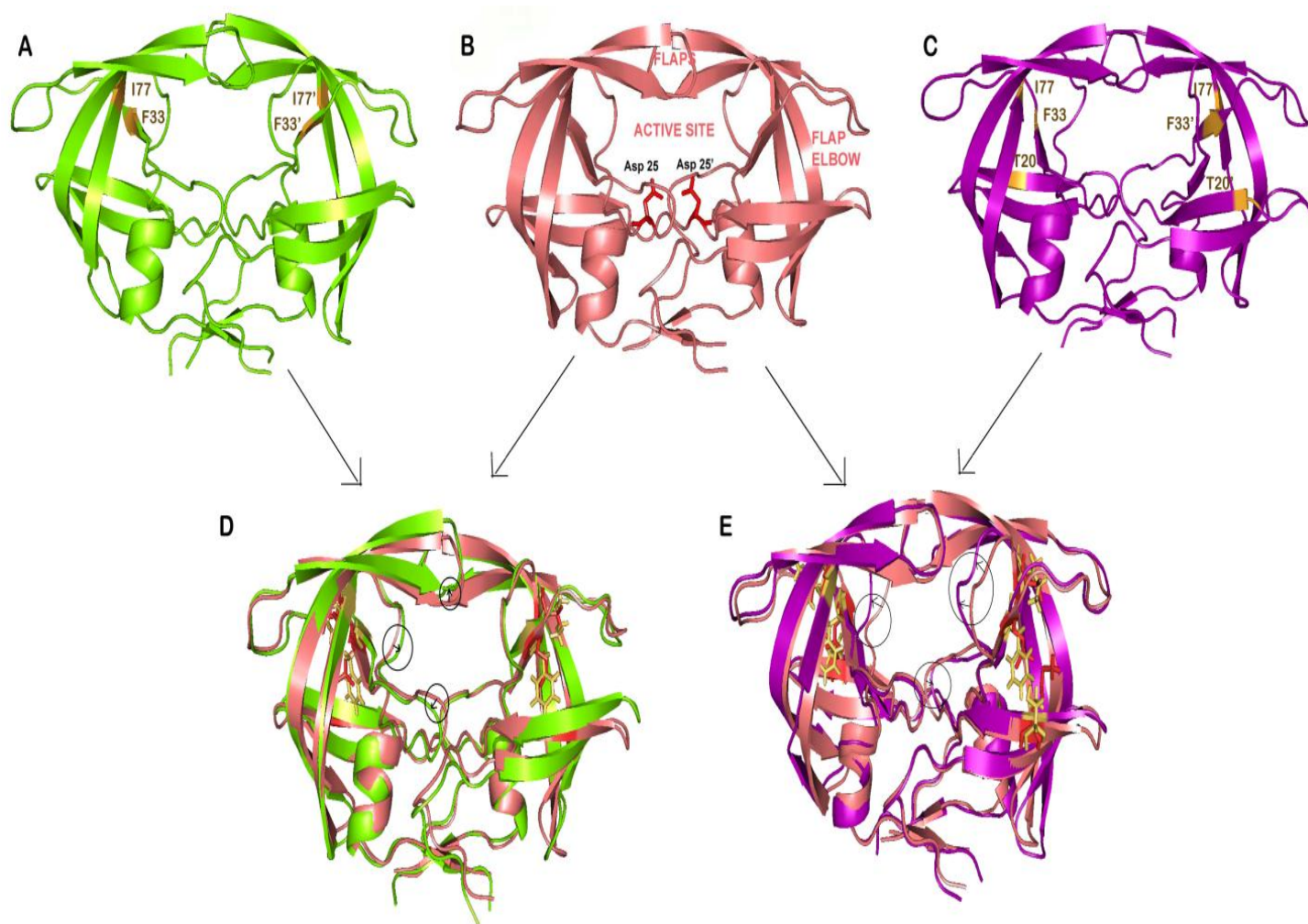**Figure 5:** (A) Structure of double mutated protease: DBM. Mutated residues V77I and L33F are shown in yellow. (B) Crystal structure of HIV-1 PR. (C) Structure of triple mutated protease: TPM. Mutated residues V77I, L33F and K20T are shown in yellow. (D) DBM superimposed on wild type HIV-1 PR. (E) TPM superimposed on wild type HIV-1 PR. Difference between wild and mutants is highlighted using arrows and circles.
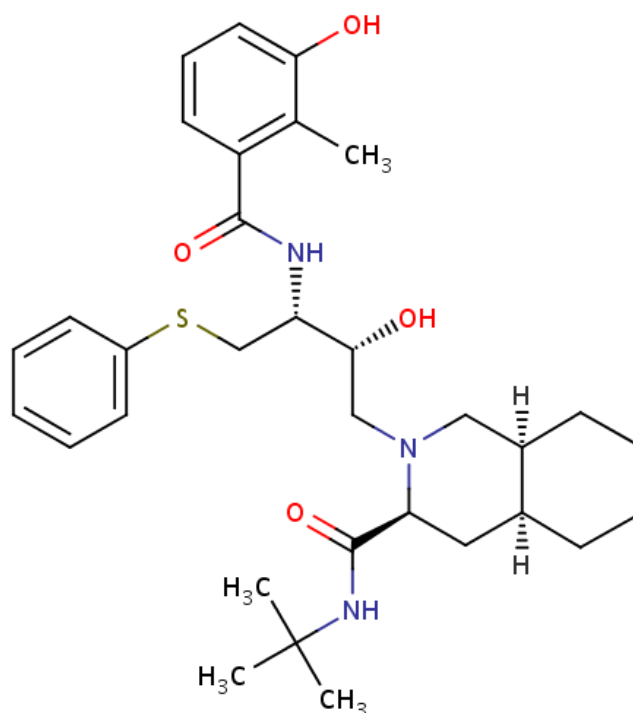
Figure 6: Molecular structure of Nelfinavir

4.2 Molecular dynamics simulations studies

MD simulations of the docked and unliganded complexes (both wild and mutant) were accomplished using Desmond Molecular Dynamics system, with Optimized Potentials for Liquid Simulations (OPLS) all-atom force field 2005 (Jorgensen et al., 1996; Kaminski et al., 2001). Prepared protein molecules were solvated in the presence of explicit solvent on a fully hydrated model with TIP4P water model in a triclinic periodic boundary box (distance between box wall and protein complex was kept 10 Å to avoid the direct interaction with its own periodic image) to generate required systems for MD simulations. Energy of prepared systems for MD simulations was minimized to 5000 steps maximum using steepest descent method until a gradient threshold (25 kcal/mol/Å) was reached, followed by L-BFGS (Low-memory Broyden-Fletcher- Goldfarb Shanno quasi-Newtonian minimizer) until a convergence threshold of 1 kcal/mol/Å was met. The default parameters in Desmond were applied for systems equilibration. The so equilibrated systems were then used for simulations at 300 K temperature and a constant pressure of 1atm, with a time step of 2fs. The long range electrostatic interactions were handled using Smooth Particle Mesh Ewald Method. Cutoff method was selected to define the short range electrostatic interactions. A cutoff of 9 Å radius (default), was used.

The prepared confirmations of NFV were docked to the stabilized mutants DBM and TPM using Glide docking software. A Glide scoring grid was prepared on the active site of the homodimer i.e. the interface of both the subunits, using receptor grid generation platform of Schrödinger (Friesner et al., 2004; Halgren et al., 2004). Keeping all the parameters default, a grid of size $20 \times 20 \times 20$ Å

with inner box size of $10 \times 10 \times 10$ Å was generated. All the Glide docking studies were performed on Intel Core 2 Duo CPU @ 3 GHz of HP origin with 1 GB DDR RAM.

4.3 Calculation of binding energies

The binding free energy was calculated according to the standard Molecular Mechanics- the generalized Born Model and Solvent Accessibility method, using Prime MM/GBSA (Lyne et al., 2006) (Prime version 2.1, 2009). NFV-protease merged structures of the XP docking protocol were used for the calculation of free energy of wild and mutant docked proteases. The binding free energy $\Delta G_{binding}$ was estimated using the following equation:

$$\Delta G_{binding} = E_{R:L} - (E_R + E_L) + \Delta G_{SA} + \Delta G_{SOLV;}$$

Where, $E_R + E_L$ is the sum of energies of unbound ligand and receptor, and $E_{R:L}$ is the energy of the docked complex. ΔG_{SA} is the difference of surface area energy of the protein-ligand complex and the sum of surface area energies of protein and ligand individually. ΔG_{SOLV} again is the difference in the GBSA salvation energy of the complex and summation of individual salvation energies of protein and ligand. Energies of the complex were calculated using Optimized Potentials for Liquid Simulations- All Atom force field (Jorgensen et al., 1996) and GB/SA continuum solvent model.

4.4 Hydrogen bond and hydrophobic interaction analysis

The hydrophobic interactions and H-bonds of the docked complexes were analysed using Ligplot program (Wallace et al., 1995). The parameters identified to define the H-bonds between ligand and the protein complex were the acceptor-donor atoms with distances less than 3.3 Å, hydrogen acceptor atom with distances maximum 2.7 Å and the acceptor-H-donor angle of 90° or more. Ligand bound protein structures from Glide and Representative Structures from Desmond were selected for carrying out interaction studies.

4.5 SYFPEITHI epitope prediction analysis

Epitope prediction was done using Syfpeithi Database of MHC ligands and peptide motifs (Rammensee et al., 1999) . The algorithm relies on the scoring of binding motifs. From the first amino acid of the protein, its sequence is divided into octamers, nonamers and decamers. The score of each oligomer is then calculated according to the summation of scores of individual amino acids. The amino acids are scored based on their observed frequencies. Most frequently occurring residues in the anchor positions are given value 10, followed by 8 given to the residues occurring in significant number of ligands. Likewise, residues regarded unfavorable for binding have a coefficient of -1 to -3. We have used MHC class HLA-A3 for our analysis.

4.6 Calculation of volume and surface are of HIV-1 PR cavity

We used CASTp online server (Dundas et al., 2006) to estimate the cavity volume and surface area of wild and mutant proteases. CASTp works on the principles of Alpha Shape Theory (Liang et al., 1998) for detection and measurement of pockets in protein which are inaccessible to the solvent outside. The probe of radius 1.4 Å is used for cavity measurement.

4.7 Generation of combinatorial library

After analyzing the mechanism of resistance, next step was to develop derivatives of Nelfinavir which could be used as new drug leads and could bind to both wild and mutant proteases. In order to do so Leadgrow module of Vlife MDS was used. The tool allows creating a library of compounds based on a common template. The template should contain substitution sites, the tool then applies permutation and combination to create a vast library of compounds. The greater the number of substitution sites, bigger is the library. Library was prepared using original NFV as template and two substitution sites. In order to generate the library, Leadgrow was selected from the module dropdown of Vlife MDS. The template was introduced in the tool along with the substitution sites. Compounds were created by using various combinations of substitution groups like various atoms, alkyl group, alkenes, ketones, acids, aromatic rings, cyclic groups, -OCH3, -OCH2CH3, NH2 at all substitution sites. The tool then generates all possible combinations and the compounds are saved in .mol2 format. The library was further processed using Ligprep module of Schrodinger Maestro. The prepared library was screened with wild protease and the mutant protease to find the most capable NFV derivative, capable of inhibiting both the proteases.

4.8 Residue Interaction Network Comparison

The representative structure from MD simulations of DBM and TPM were retrieved and PDB structure 1OHR were used for development of residue interaction networks and there comparison. The residues are represented as nodes and the interactions between them as edges. The networks were visualized and compared using Rinalyzer plugin of Cytoscape 2.8.1. The combined comparison networks of wild versus DBM and wild versus TPM were generated based on the superposition alignment of their 3D structures. The comparison networks are described with different type of edges. Different type describes the interaction preserved or lost between the two structures compared. Terms like betweenness centrality and closeness centrality were used to describe the interaction network. Betweenness centrality of a node represent the amount of control that node have on the interactions made by other (neighbouring) nodes in the same network. Closeness centrality is the measure of how quickly information flows from a node to other reachable nodes in the network.

RESULTS

5.1 Molecular Dynamics Study

To study the structural changes in closed confirmation of HIV-1 PR due to mutation we considered the NFV-docked crystal structure of HIV-1 PR, 1OHR. NFV was removed before mutating the residues and then the MD analysis of unliganded- wild, double (DBM) and triple (TPM) was performed for 20ns. The Root Mean Square Deviation of DBM and TPM was more stable with respect to that of wild protease, with standard deviation of 0.28, 0.23 and 0.16 of wild, DBM and TPM respectively (Figure 7). To observe and compare the movement of residues, we plotted the RMS fluctuation plot for both the subunits of the wild and mutant proteases (Figure 8). The Root Mean Square Fluctuation is the measure of average atomic mobility of the backbone atoms during the molecular dynamics (MD) simulations. The residues of DBM deviated more from the wild than that of TPM, especially in chain A. The flap residues (residue 33- 62) of the wild protease were more flexible in comparison to the mutants DBM and TPM, indicating that there was relatively strong interactions between the flap of the mutants which made them more stable in close conformation than the wild. The representative structures of DBM and TPM were then selected for studying their interaction with NFV. The mutants were docked with all the stable confirmations on NFV using Glide module of Schrodinger and was compared with the NFV-docked crystal structure of wild HIV-1 PR. The wild protease showed a great affinity with NFV, with XP docking score of -9.32 (Table 3). This strong interaction was mediated by a number of hydrophobic interactions from both the chains of wild type proteases and a single hydrogen bond between Gly27 of A chain with oxygen atom of NFV (Figure 9A, 10A). Prime/MM-GBSA free binding energy of the wild docked structure was calculated to be -38.93kcal/mol. Instead of further stabilizing this docked structure, we compared these interactions between NFV and wild protease with our reference crystal structure, 1OHR. Here, there were stronger interactions stabilized by 18 hydrophobic interactions (Figure 10B) from both the chains of protease and 4 hydrogen bonds. The catalytic site residues Asp 25 (A) and Asp25 (B) through their delta oxygen atoms made hydrogen bonds with NFV, of length 2.63Å and 2.77Å respectively. The delta oxygen atom of Asp 30(A) made 2.90Å long hydrogen bond with NFV, and Gly27(A) made 3.26Å long hydrogen bond with nitrogen atom of NFV (Figure 9B) (Appendix 2, 3, 4).
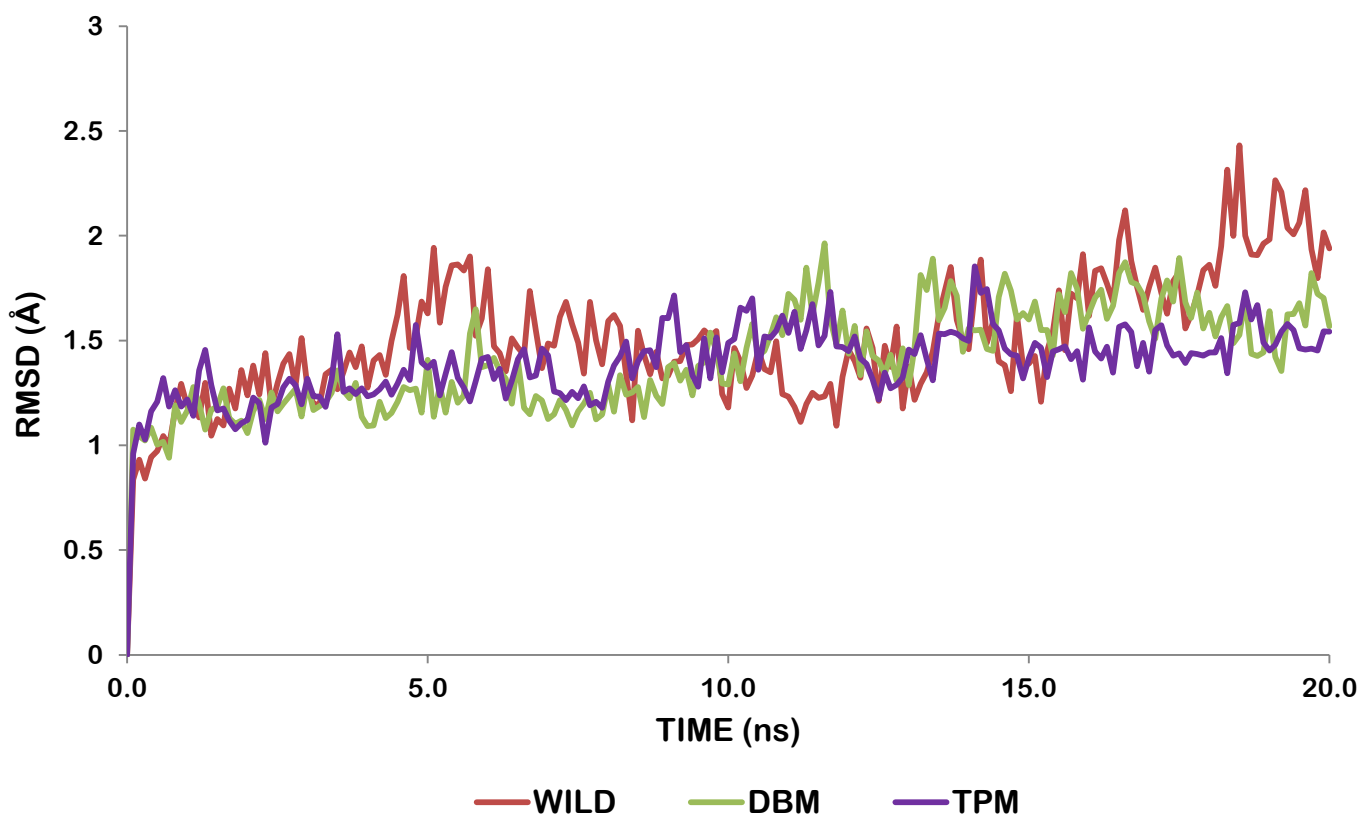
Figure 7: RMSD trajectory of Wild protease, DBM and TPM during MD simulations. Trajectory for Wild protease (red line), DBM (green line) and TPM (purple line)

Table 3: Docking score

Docking Score	Wild	DBM	TPM
Glide XP	-9.32	-7.8	-10.31
MM/GBSA(Kcal/mol)	-38.98	-11.08	-42.66
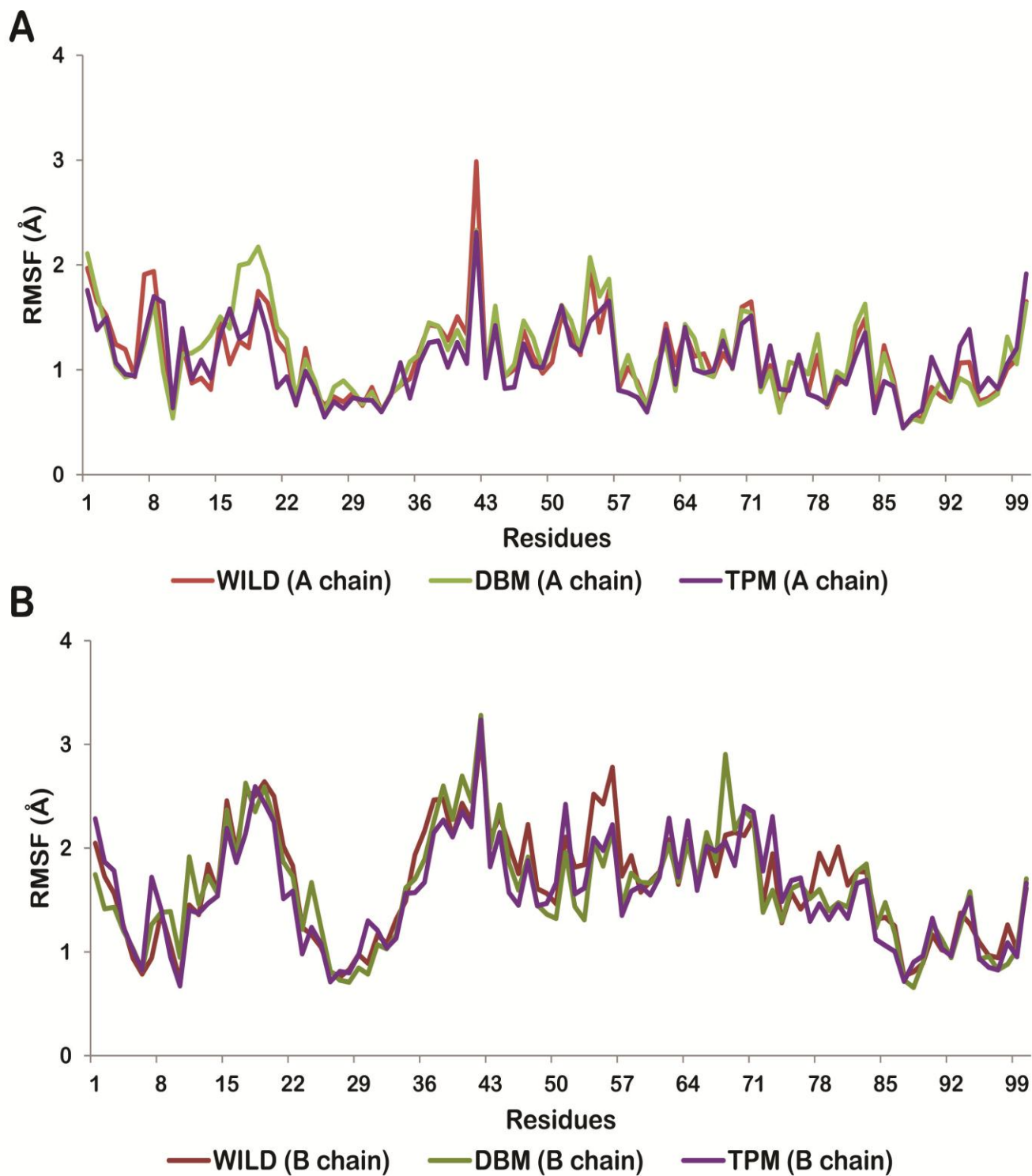
Figure 8: Residue wise RMS fluctuations of Wild protease (red line), DBM (green line) and TPM (purple line). (A) chain A. (chain B)

Figure 9: Changes in the hydrogen bonds of NFV with protease before and after simulation. A. hydrogen bonds with wild protease in Glide docked structure. B. in PDB structure 1OHR.

Figure 10: Changes in the hydrophobic interactions of NFV with protease before and after simulation. A. hydrophobic interactions with wild protease in Glide docked structure. B. in PDB structure 1OHR.
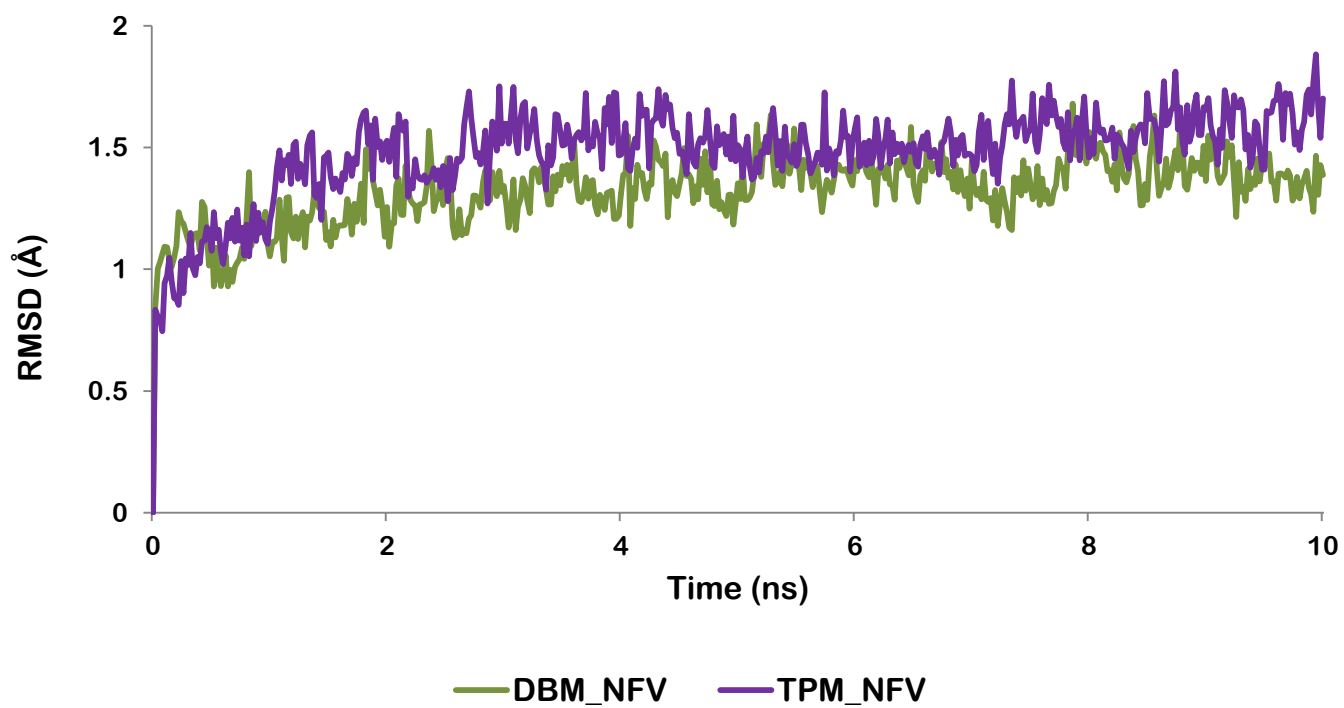
Figure 11: RMSD trajectory of NFV docked DBM and TPM during MD simulations. Trajectory for DBM (green line) and TPM (purple line)
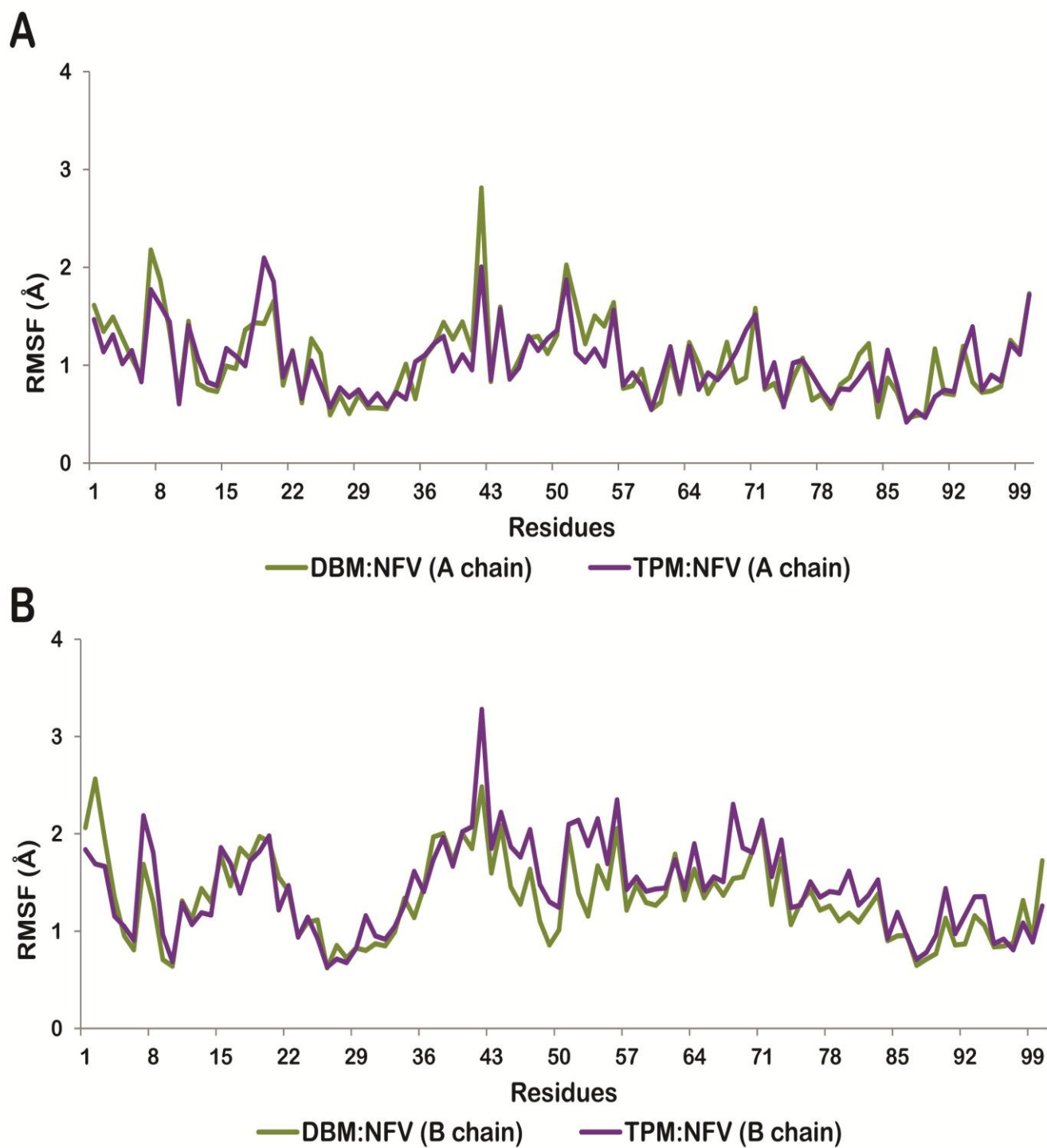
Figure 12: Residue wise RMS fluctuations of NFV docked- DBM (green line) and TPM (purple line). (A) chain A (B) chain B

31

Similar strategy was applied to study the binding interactions of mutants, DBM and TPM. The docking affinity of DBM was found to be decreased with the docking score of -7.78. The Prime/MM-GBSA free binding energy of docked-DBM had also decreased by 27.90kcal/mol to -11.08kcal/mol (Table 3). The binding of NFV was mediated through thirteen hydrophobic interactions from both the chains in double mutant protease and six hydrogen bonds between protease and NFV. The interactions were made by residues- Gly27(A), Asp29(A), Gly27(B), Asp25(B) and Asp29(B) (Figure 13A, 14A). The docked structure was stabilized *in-silico* through 10ns molecular dynamics simulations. The structure was stable throughout the simulation with RMSD standard deviation of 0.132 (Figure 11). The decrease in flexibility of flap and active site residues of the docked-DBM in comparison to the undocked-DBM was an obvious result (Figure 12). The binding interactions were reduced to four long hydrogen bonds and sixteen week hydrophobic interactions. The alignment of NFV had also changed reducing its surface area interaction with the cavity residues. The hydrogen bonds were formed by active site residues of B chain only- Asp25(B) and Gly27(B) only (Figure 13B, 14B).

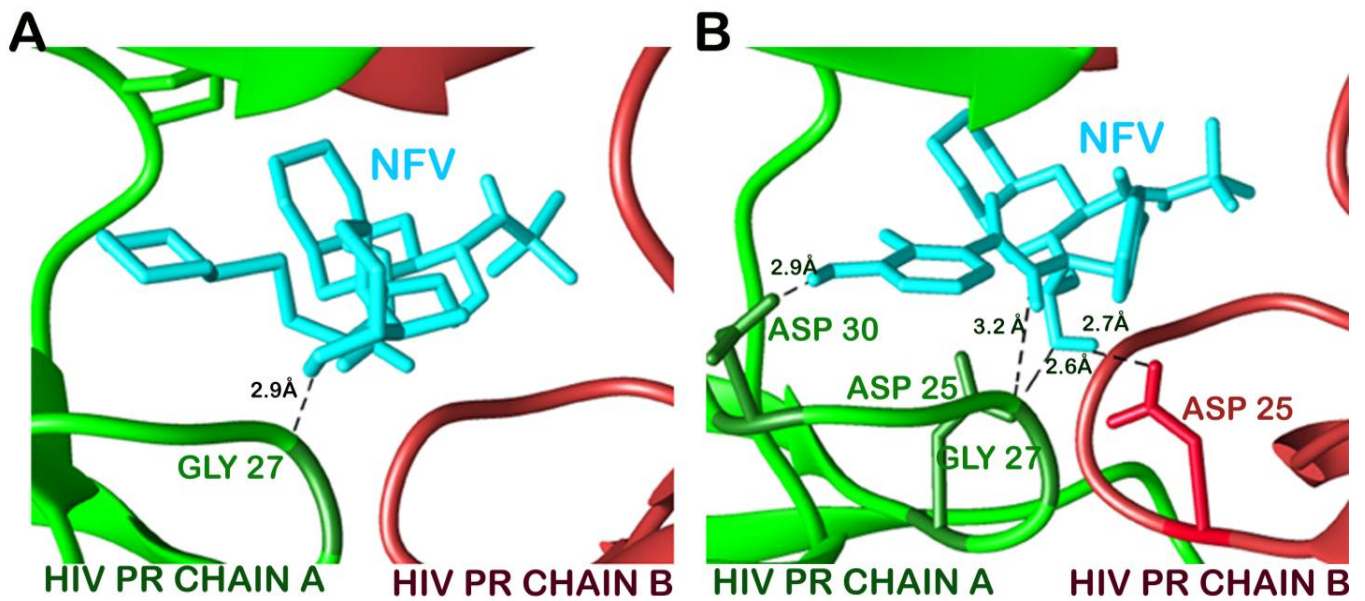**Figure 13:** Changes in the hydrogen bonds of NFV with protease before and after simulation. A. hydrogen bonds with DBM in Glide docked structure before simulation . B. After simulation.
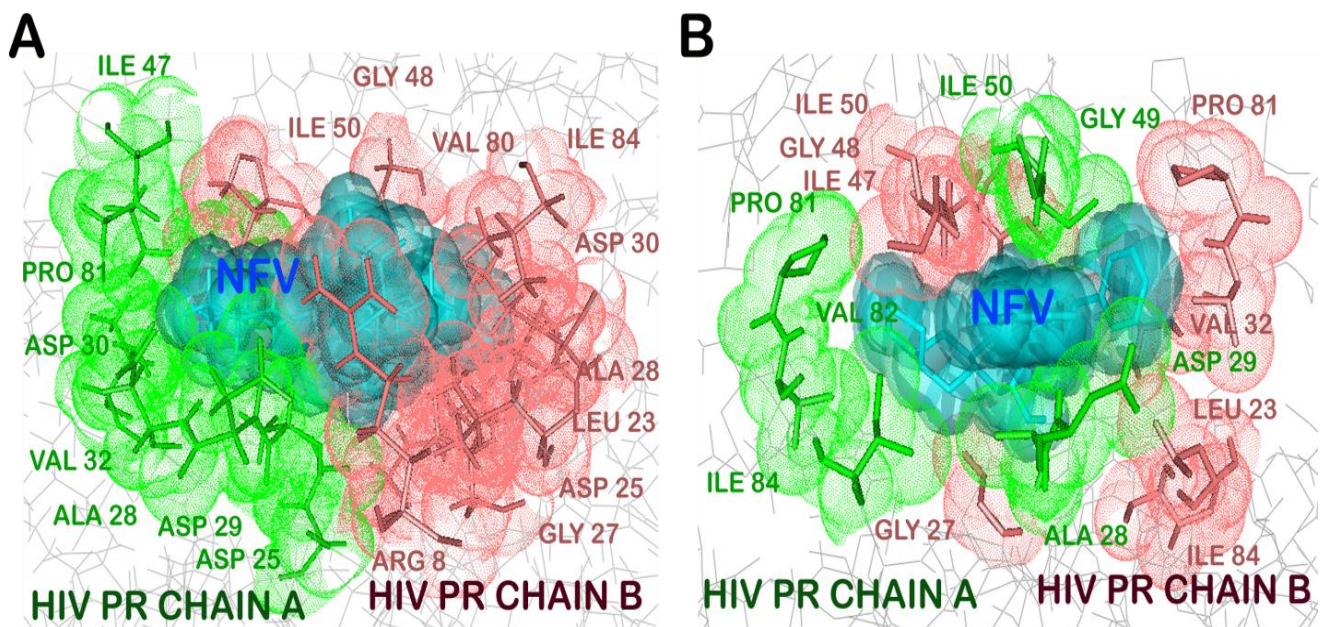
Figure 14: Changes in the hydrophobic interactions of NFV with protease before and after simulation. A. hydrophobic interactions with DBM in Glide docked structure before simulation. B. After simulation.

The simulation and docking studies of triple mutant, TPM were notable different from expected resistant proteases and therefore justified its lower clinical presence with respect to DBM. The initial docking scores (Glide XP score) of stable mutant, TPM with NFV was found to be -10.314. The Prime/MM-GBSA free binding energy was stabilized by 3.68kcal\ml to -42kcal\mol, with respect to the wild protease (Table 3). This binding was supported by five hydrogen bonds made by residues-Arg8(A), Asp25(A), Ile50(A) and Gly48(B); and thirteen hydrophobic interactions (Figure 15A, 16A). However the number of interactions had significantly reduced after molecular dynamic simulations of 10ns. The NFV-TPM complex was stable throughout the simulation trajectory with RMSD standard deviation of 0.183 (Figure 11). The RMSF values of individual residues had also reduced with respect to undocked-TPM (Figure 12). The binding was supported by only two hydrogen bonds made by Asp25(A) and twelve hydrophobic interactions by the cavity residues. From this we could suggest that otherwise resistant K20T mutation have the potential to reduce the resistance of TPM in comparison to DBM (Figure 15B, 16B).

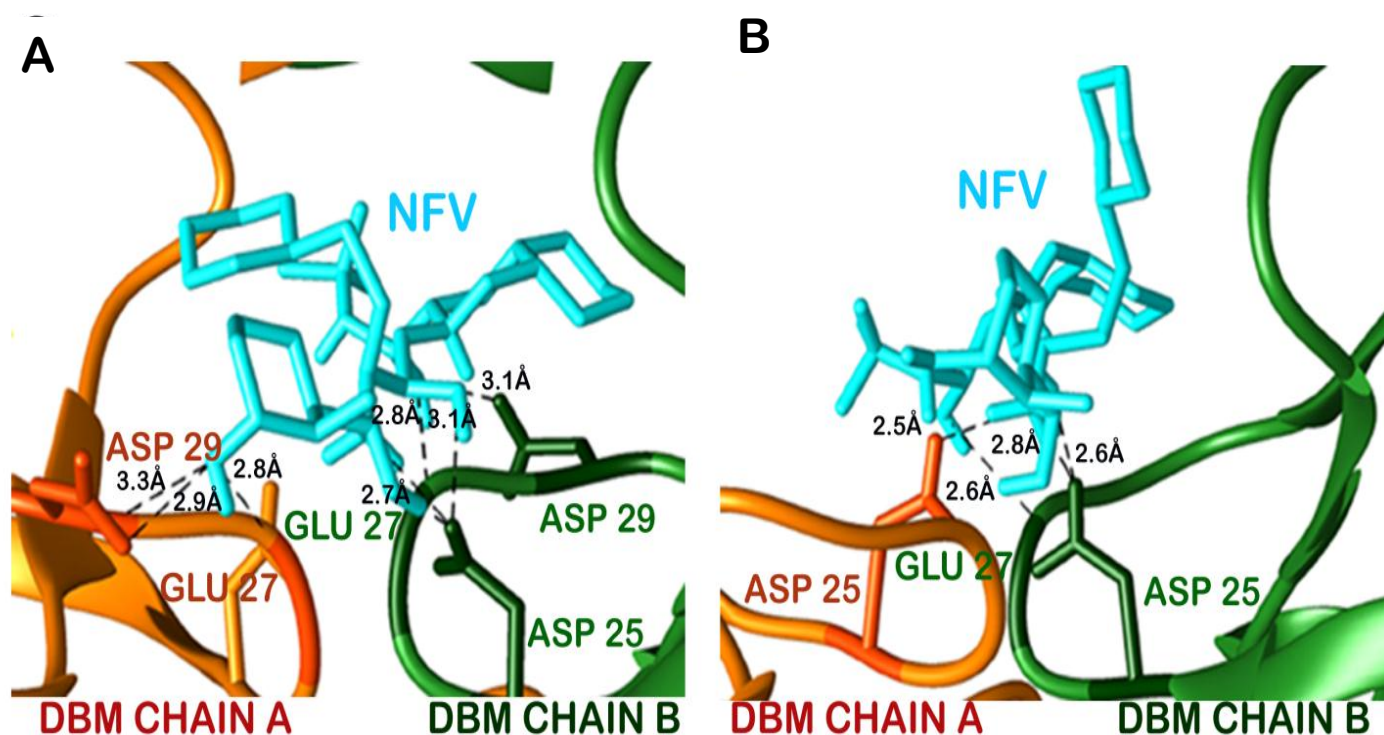**Figure 15:** Changes in the hydrogen bonds of NFV with protease before and after simulation. A. hydrogen bonds with TPM in Glide docked structure before simulation . B. After simulation.
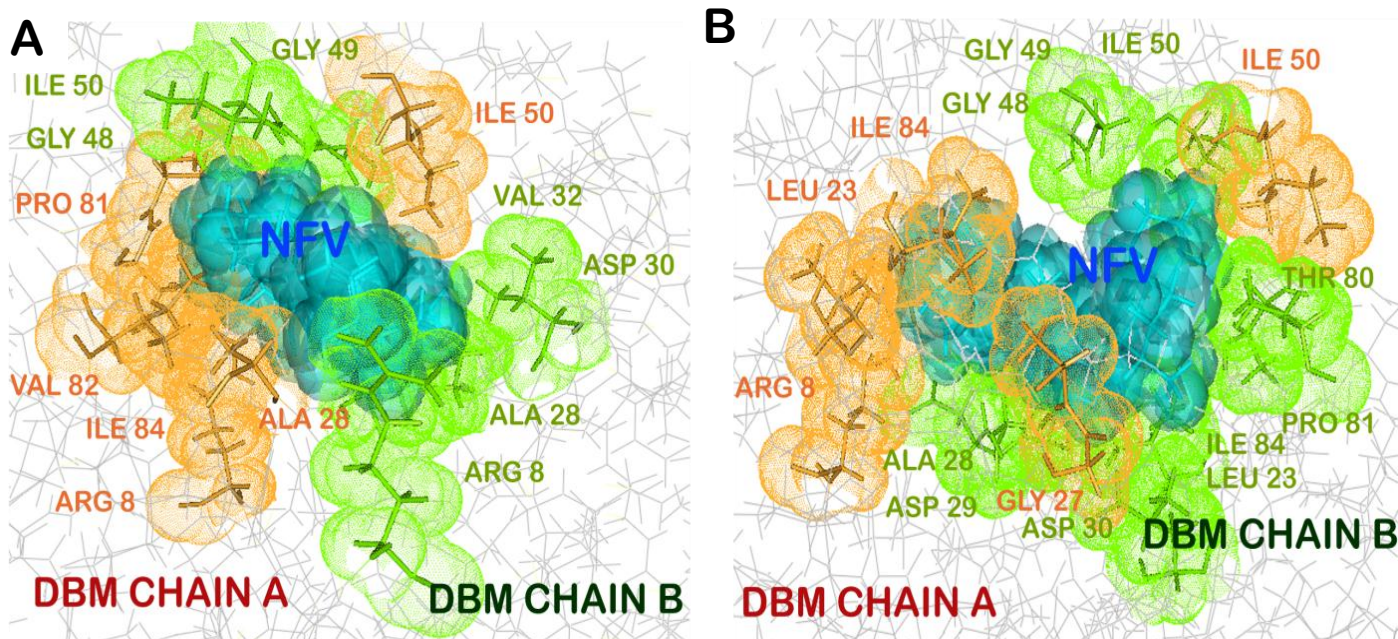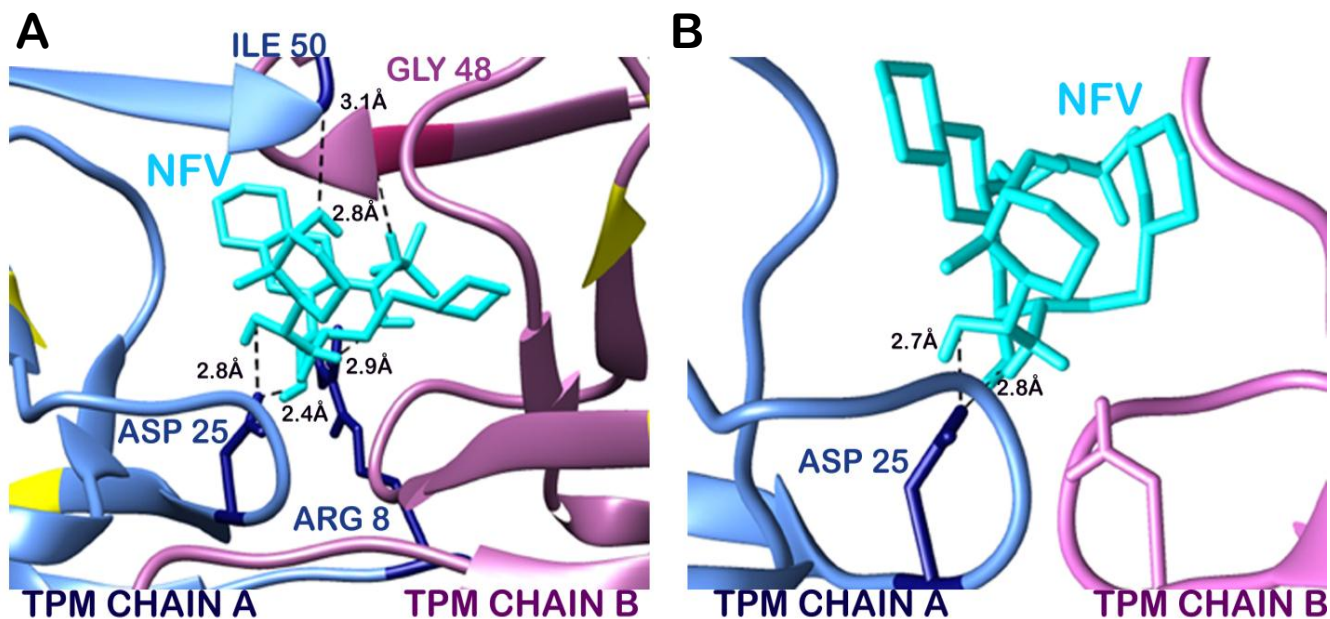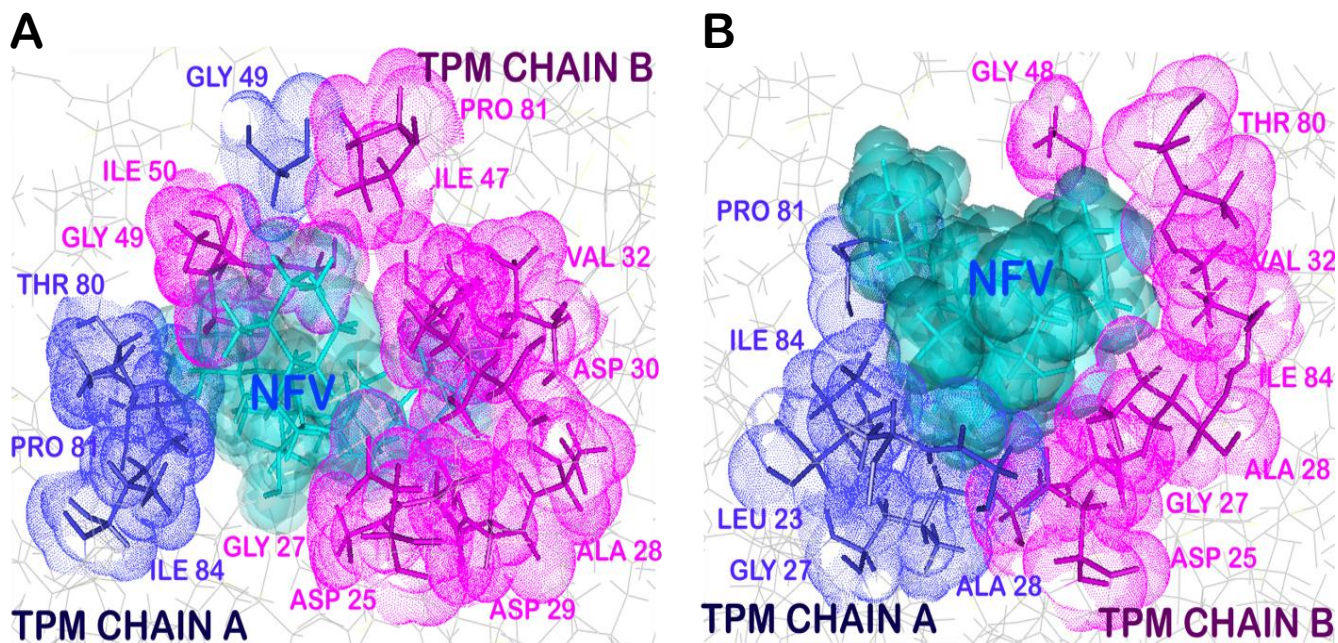

Figure 16: Changes in the hydrophobic interactions of NFV with protease before and after simulation. A. hydrophobic interactions with DBM in Glide docked structure before simulation. B. After simulation.

4.2 Possible intra-molecular interactions by V77I, L33F, K20T and the neighbouring residues, and their effect on cavity size

To further investigate the role of these clinically significant mutations on the structure of protease and to understand the reason behind the marked variation in interactions and docking score, the cavity size and volume of mutant and wild proteases was studied.

The pocket volume and surface area of wild protease was found to be 1186.1 Å^3 and 705.9 Å^2 respectively (Supplementary Data 1) (Figure 17). We tried to analyze the number of possible hydrogen bonds which could be formed by the mutated and their neighbouring residues. Lys20 of both the wild protease chains made two hydrogen bonds with their corresponding Ile13 residues. Leu33 formed two hydrogen bonds Leu76 and Gly78 in both the chains. Val77 made two hydrogen bonds with Arg57 in both the chains. Thirty-nine intermolecular hydrogen bonds could possibly be formed by V77, L33, K20 and neighbouring aminoacids (18-22, 31-35, 75-79) including above interactions (Appendix 1).

V77I mutation in combination with L33F (DBM), presented increase in the size of binding cavity to 1375.5 Å^3 volume and 732.10 Å^2 area. This increase in cavity size probably is the reason behind decreased docking affinity of NFV to DBM (due to decrease in contact surface area of ligand and active site residues) (Figure 18). L33F mutation had caused positional change of neighbouring residue Glu34, which caused formation of an extra hydrogen bond between Glu34 and Lys20 in DBM. Total forty-eight hydrogen bonds could be formed by I77, F33, K20 and their neighbouring residues in DBM.

As expected, the binding pocket volume and area of triple mutant had reduced with respect to wild, thereby increasing the contact surface area between ligand and active site residues. The pocket volume and surface area of TPM was found to be 1042.5 Å^3 and 634.3 Å^2 respectively (Figure 19). Total thirty-nine hydrogen bonds were formed by I77, F33, T20 and their neighbouring residues in TPM. Thr20 formed an additional hydrogen bond with Gly21 in B chain, whereas hydrogen bond between 33rd residue and Gly34 in B chain, and Leu76 in A chain were lost due to mutations.
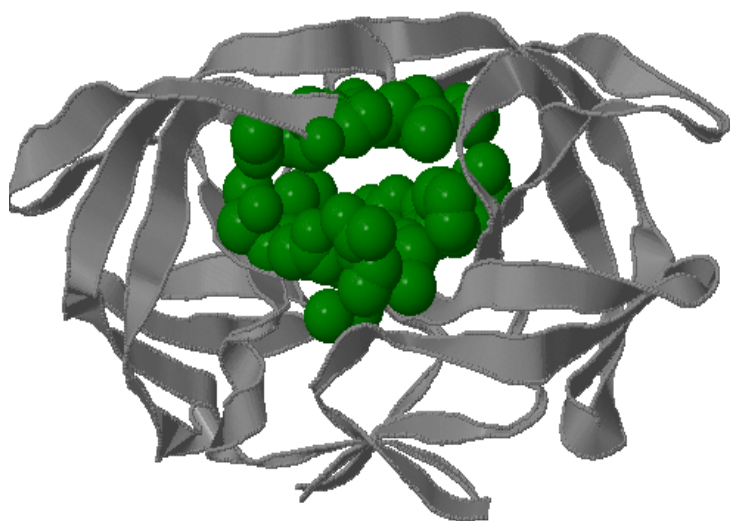
Figure 17: CASTP representation of Wild protease cavity size.
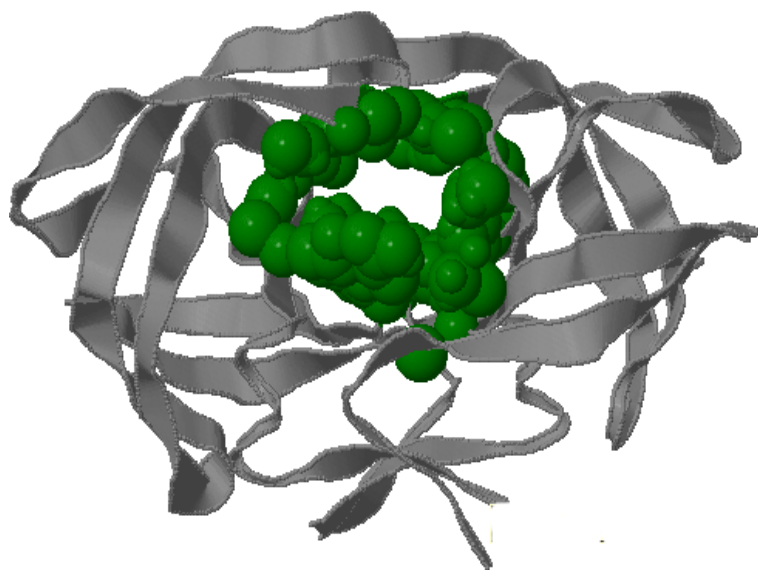
Figure 18: CASTP representation of DBM cavity size.
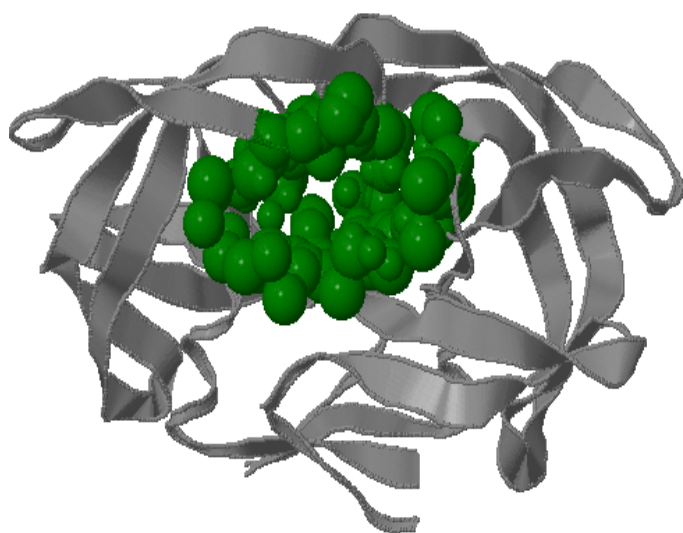
Figure 19: CASTP representation of TPM cavity size.

Table 4: BINDING CAVITY SIZE AND AREA

CAVITY	Wild	DBM	TPM
VOLUME ($\mathring{A}^3$)	1186.1	1375.5	1042.5
AREA ($\mathring{A}^2$)	705.9	732.1	634.3

4.3 Comparison of flap movements of double and triple mutant with the wild type protease

To study the effect of V77I mutation along with L33F and K20T on the flap movements of HIV protease, we considered semi open modelled structure HIV-1PR: 1HVP (Weber et al., 1989). We performed small molecular dynamics simulations of 5ns to view the flap opening mechanism of mutants and compare them with wild protease. 1HVP was processed and mutated similar to the technique followed before. The mutants of this modelled protease are abbreviated as DBM_M (V77I, L33F) and TPM_M (V77I, L33F, K20T). The wild and mutated structures were stable throughout the trajectory. RMSD standard deviation of DBM_M was least with the value of 0.3, representing its more stable nature in comparison to wild (standard deviation: 0.45) and TPM_M (standard deviation: 0.53). The RMSD trajectory of all three structures has been shown in Figure 20. The RMSF plots of both the chains were plotted for wild, DBM_M and TPM_M (Figure 21A, 21B). Though not a much difference was observed between the mutants and wild, but B chain of DBM_M was highly flexible with RMSF of flap residues reaching till 6.66 Å. This indicates wider opening of the flap residues of DBM_M.

To verify this proposal we calculated the distance between I50(A)-Cα and I50(B)-Cα atoms (Toth & Borics, 2006). The transition between semiopen and open confirmations of protease flaps is characterized by interaction between I50 residue located on the tip of the flaps. The flaps of DBM_M separated to the maximum distance of 26.8Å between I50 Cα atoms, before 1 ns of simulation time (Figure 22). This separation occurs early in wild protease at around 400ps with the separation of 25.7Å between I50 Cα atoms. DBM_M and wild protease regain their semiopen open confirmations after 1.13ns. The closing of DBM_M after 2.03ns is not in sync with the wild protease (Appendix 5). The flaps of TPM_M showed entirely different trend throughout the simulation time. They retain their closed to semiopen transition for about 2.25ns. This distance plot suggests that flaps DBM_M are more flexible with respect to NFV-susceptible wild protease and less clinically prominent TPM_M proteases. The flap movement of wild and mutants was visualized after specific time intervals during the simulation and has been presented in Figure 23.
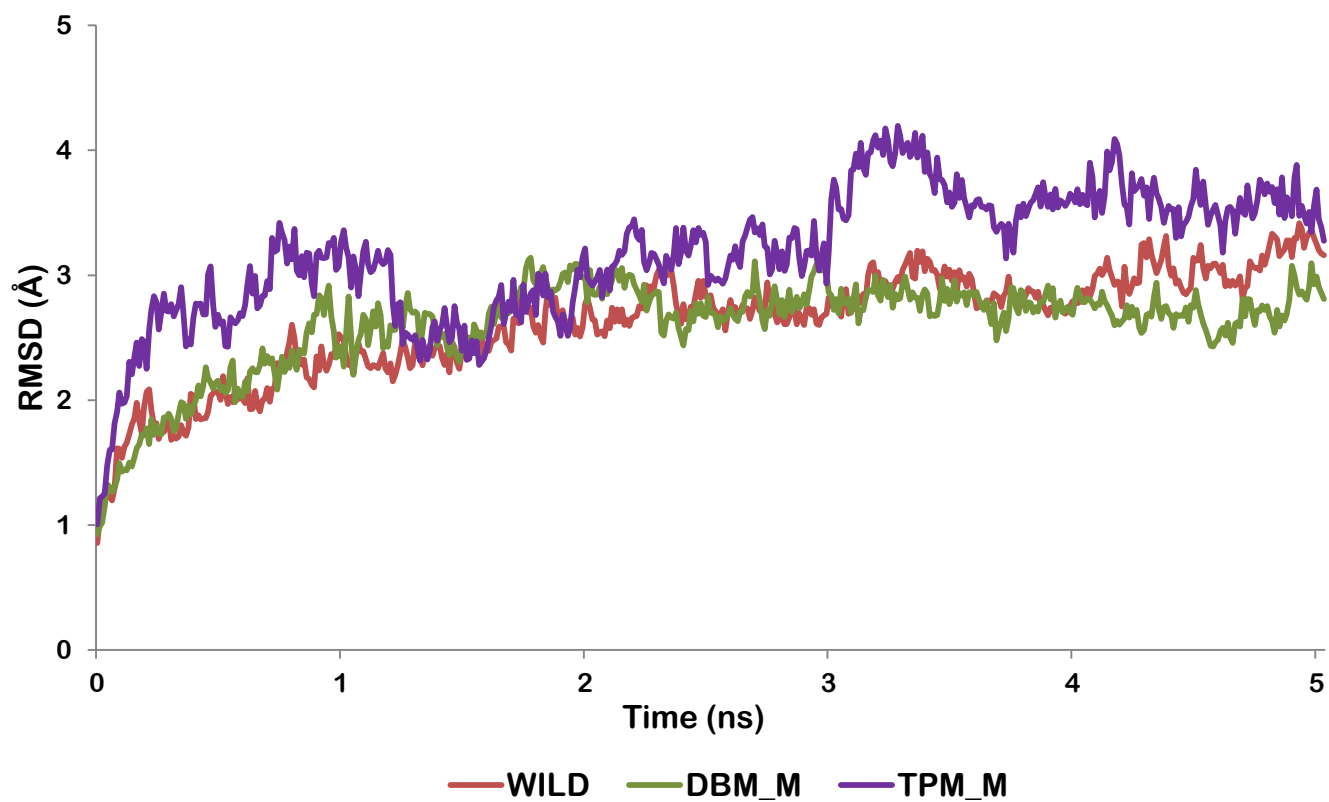
Figure 20: RMSD trajectory of Modeled Wild protease, DBM_M and TPM_M during MD simulations. Trajectory for Wild protease (red line), DBM_M (green line) and TPM_M (purple line).
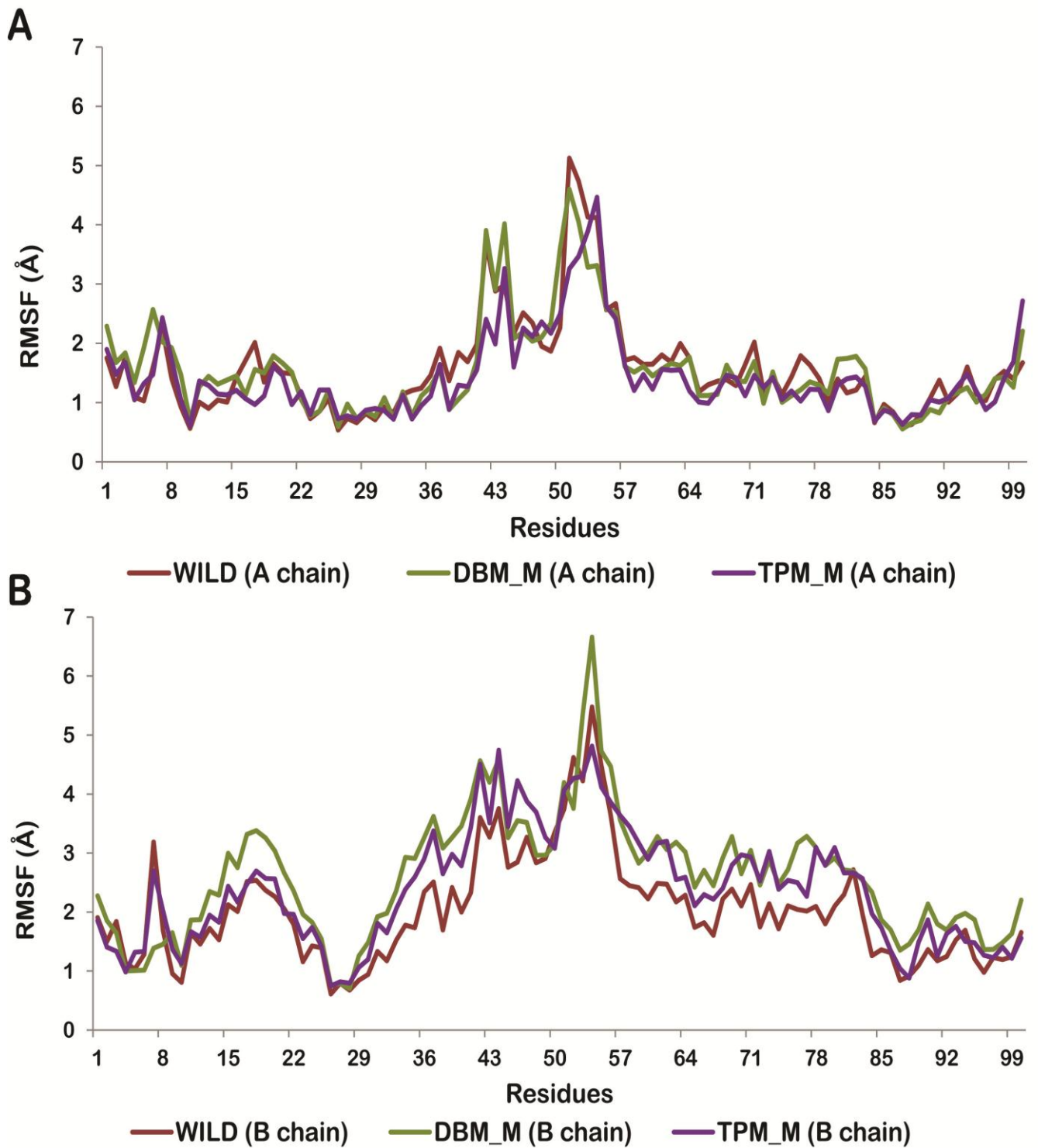
Figure 21: Residue wise RMS fluctuations of modelled Wild protease (red line), DBM_M (green line) and TPM_M (purple line). (A) Chain A. (B) Chain B
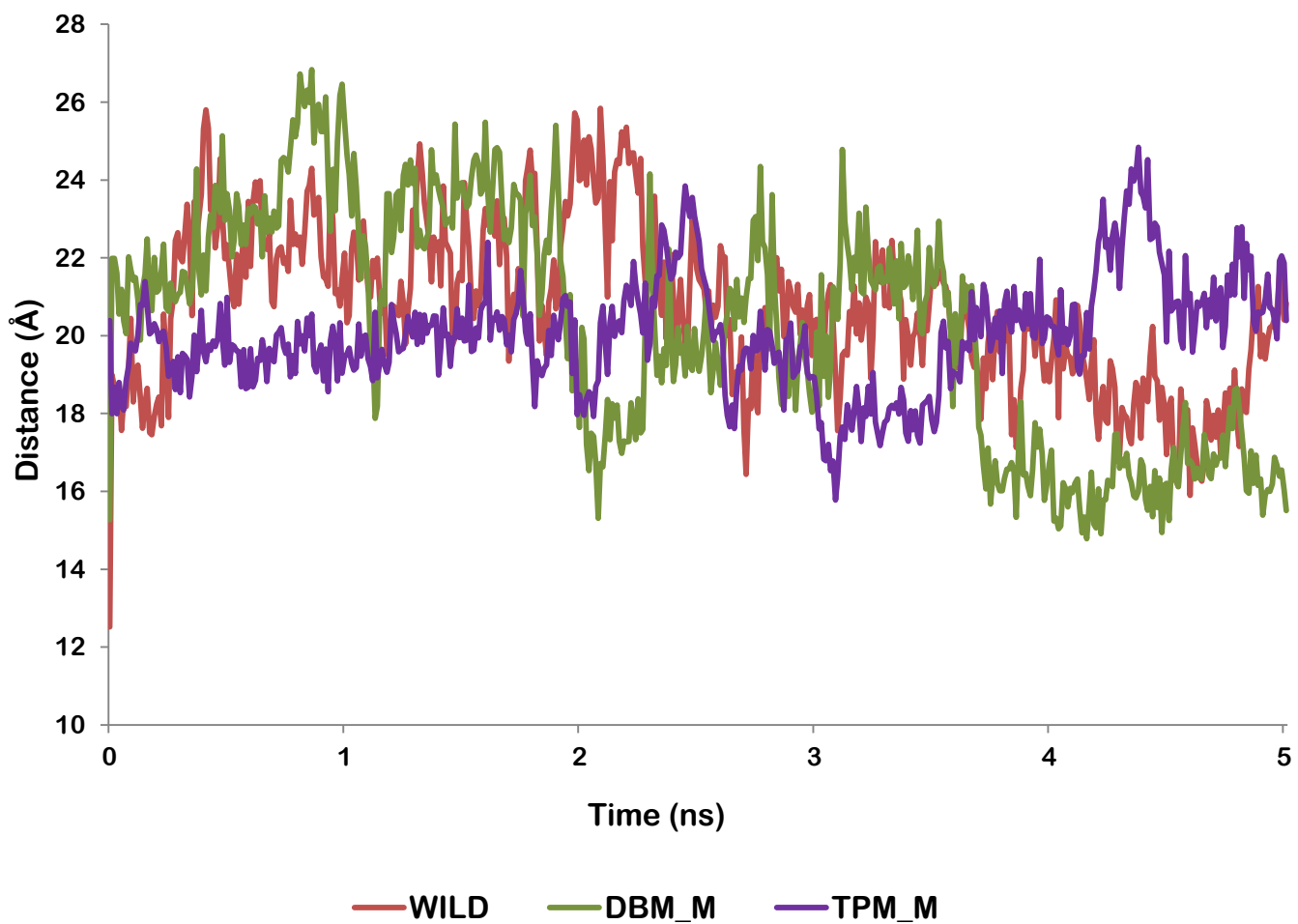
Figure 22: I50/I50' Distance plot in wild protease (red line), DBM_M (green line) and TPM _M (purple line)
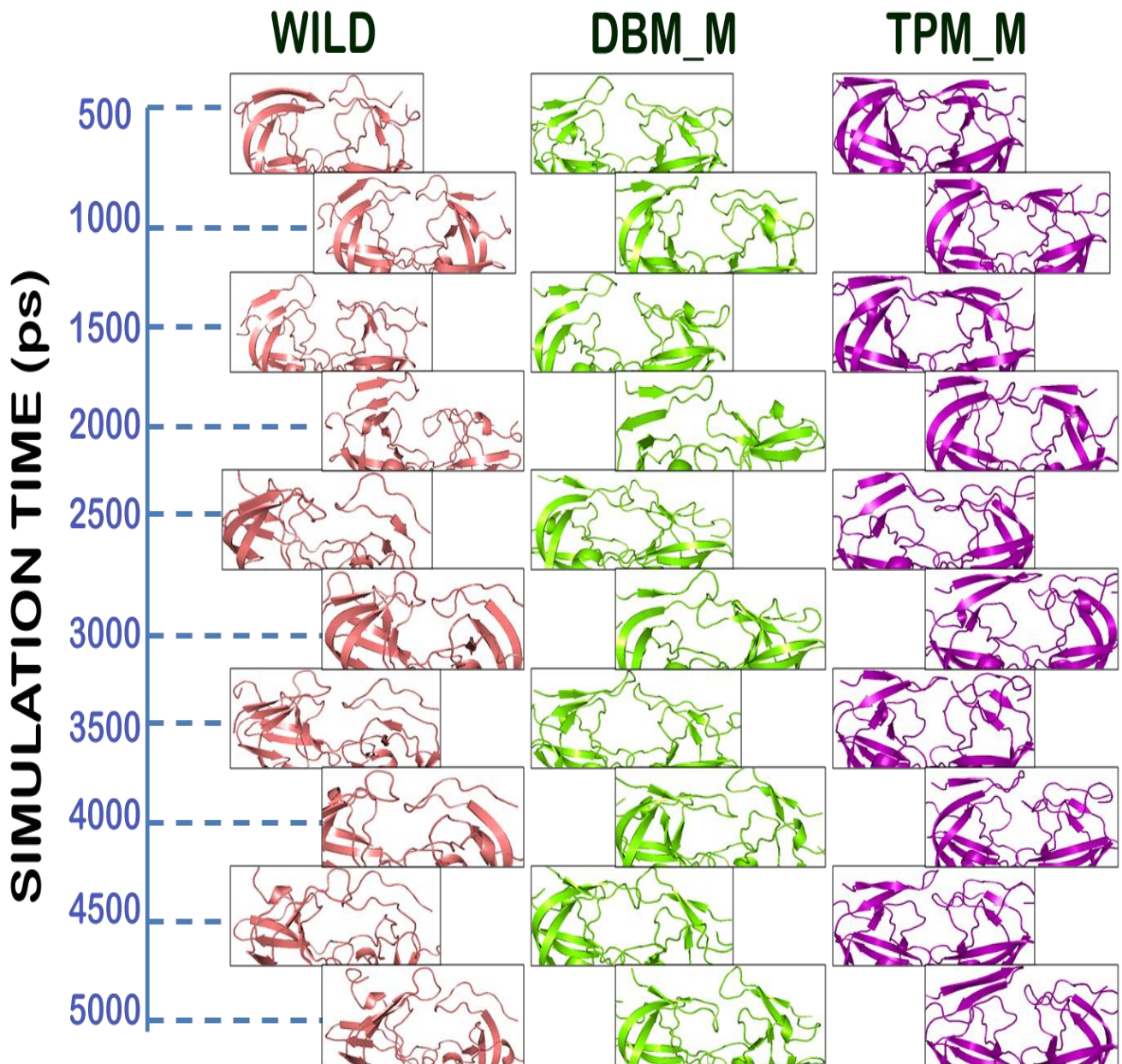
Figure 23: Flap mutant of Wild protease (1HVP), DBM_M and TPM_M

4.4 Residue Interaction Network comparison

Analysis of residue networks of a protein structure has been widely performed to gain knowledge of the residues playing key roles "hubs" in the protein structure. Here we have tried to study the relations between this small world networks and NFV resistant mutations and thereby tried to explore the effects of these mutations on neighbouring and other important residues. The representative structures of DBM and TPM from the molecular dynamics study were retrieved for the analysis. The RINs of these mutated representative structures were then compared with the RIN of wild protease (PDB id: 1OHR).

The edges of the network represent the covalent and non-covalent interaction among the residues. The solid lines represent interacts preserved in both wild and mutated structures, dashed line were interactions which were present in wild protease but were lost due to mutations in DBM or TPM. The dotted lines represent the newly formed interaction in the mutated proteases. The different types of interactions are colour coded in different colours to have greater understanding of the molecular structure of proteases. Hydrogen bond between main chain atoms have been shown in dark blue, H-bond between main chain atom and side chain atoms is shown in light blue. Ionic interaction between side chain atoms is shown in orange and the polar bonds between main chain atoms is shown in forest green. The inter-atomic contacts between main chain atoms are shown in dark purple, between main chain – side chain atoms in light purple and between two side chain atoms in light orange. The mutated residues are labelled in red colour. The detailed legend is shown in Appendix . For simplicity, only the important residues are presented in comparison networks in Figure 24.

It has been observed that residues which are functionally important carry a high closeness value and the residues playing major role in stabilizing protein structure have a high shortest betweenness value. To apply this in our study shortest path betweenness and closeness centrality of each node of DBM, TPM and wild protease was calculated and then compared. To have a better and easier understanding the shortest path betweenness is represented as node size: the node size increases with the betweenness centrality; and the Closeness centrality is described with the help of node colour: lighter the node colour higher is the closeness centrality.

COMPARISION NETWORK BETWEEN WILD PROTEASE AND DBM

The residue interaction network shows that there is a considerable change in the interaction pattern caused due to mutations. A detailed understanding of effect of these changes can help to decipher the cause behind resistance. The flap residues Ile 54 and Ile 47 had lost ionic and interatomic interactions with Asp 30, Val 32 and Thr 80. The loss of these strong interactions may be playing important role in increased flap flexibility of DBM. Changes in the form of interactions between the residues can be clearly observed. Moreover, new interactions were made between Lys 20 and Glu 34 and mutated Phe 33 in DBM. The catalytic residue Asp 25 also made a stronger hydrogen bond with Ala 28 in DBM with respect to an inter atomic contact in wild protease.
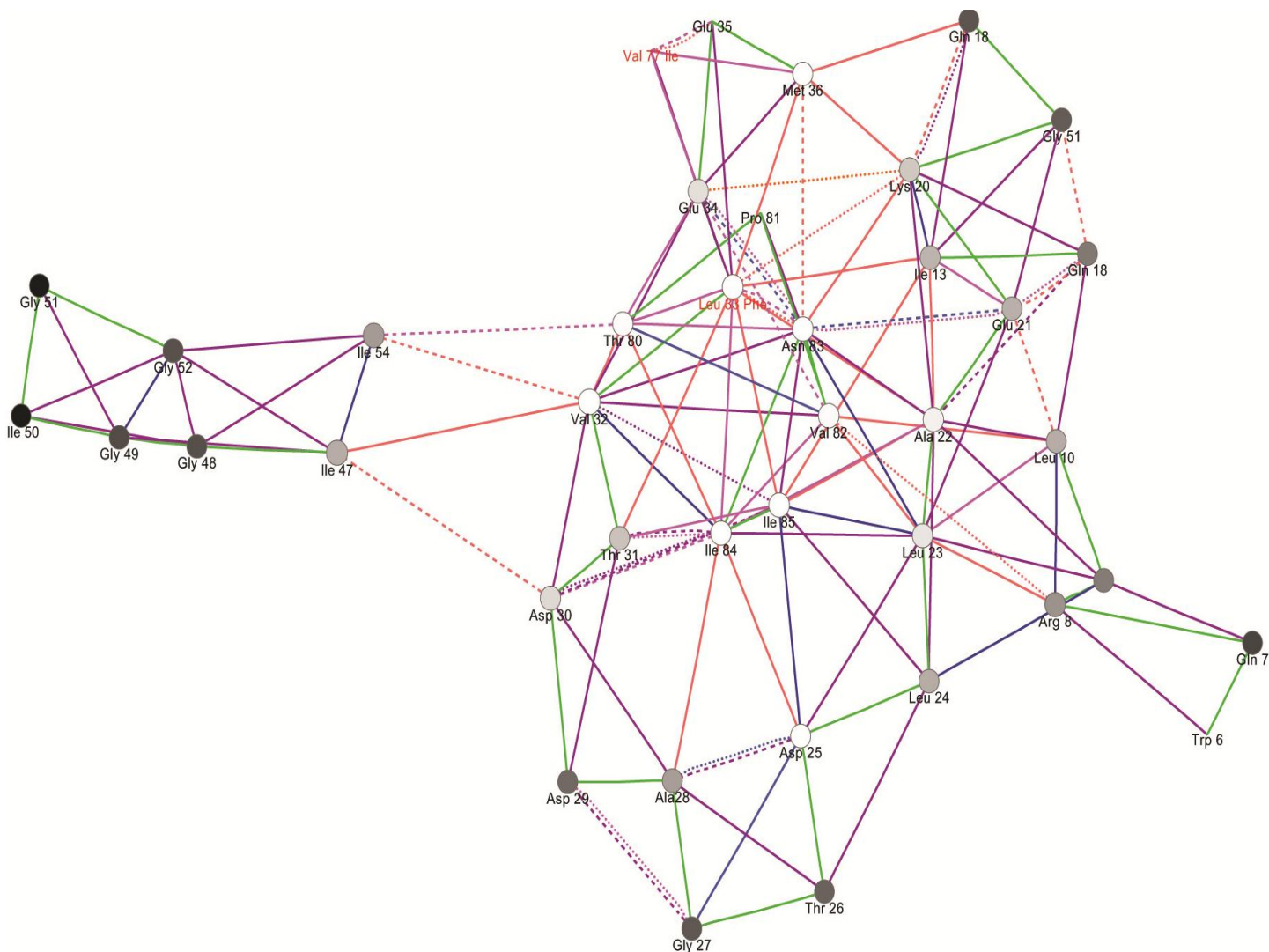
Figure 24: A detailed view of covalent and non- covalent interactions in the comparison network of wild protease and DBM.

COMPARISION NETWORK BETWEEN WILD PROTEASE AND DBM

The comparison network between TPM and wild protease shows that the interaction between the flap residues and cheek residues has been re-conserved. It presents that K20T mutation brings the flexibility of flap residues back to normal. The interaction pattern between other residues is somewhat similar between wild and TPM considering minor changes in the type of interactions, which were observed in DBM also. Met 36 is an important residue as mutation over here has shown to provide resistance against NFV in earlier studies. Interactions of Met 36 with mutated residues Ile 77 and Thr 20 were lost in TPM.
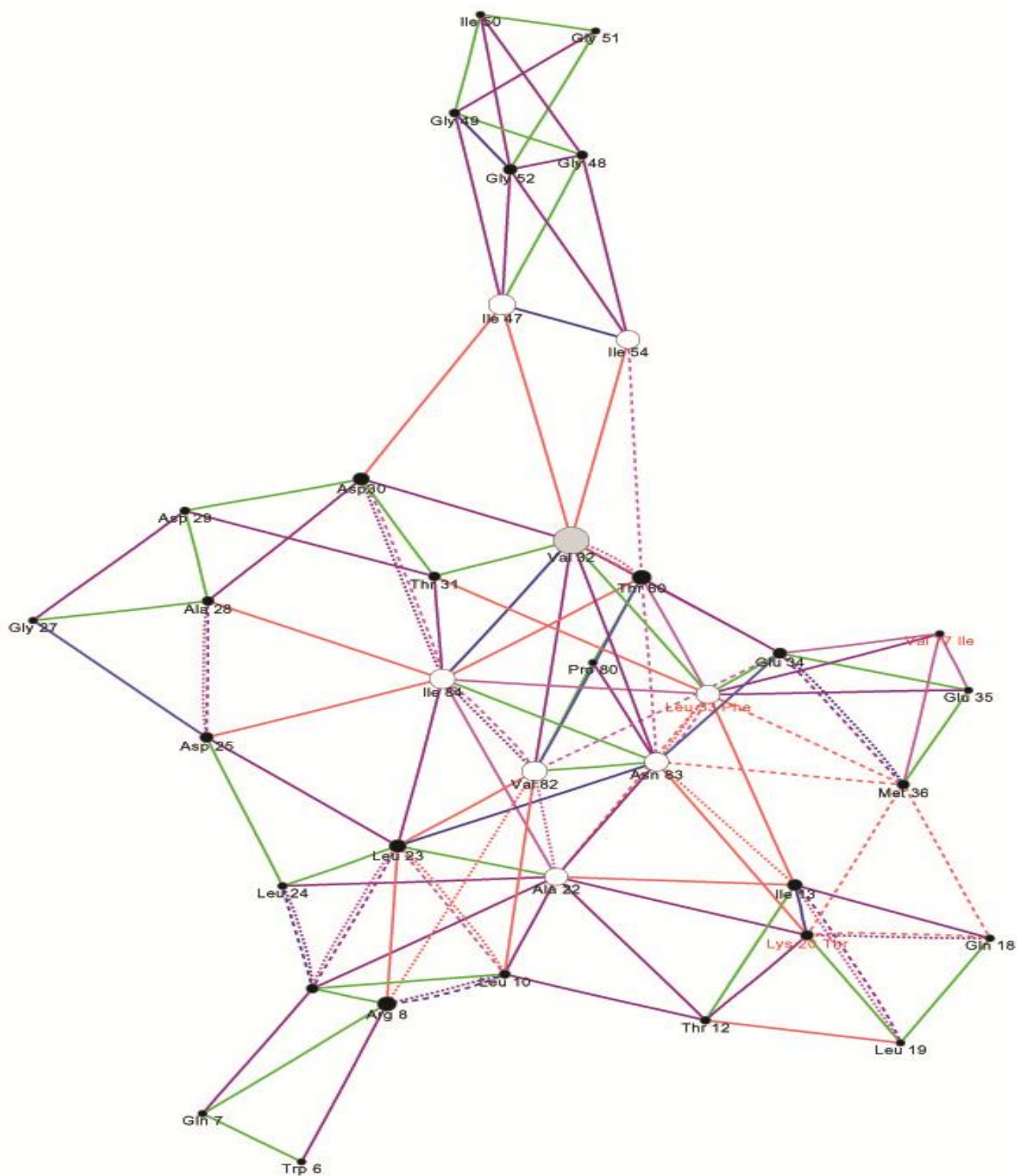
Figure 25: A detailed view of covalent and non- covalent interactions in the comparison network of wild protease and TPM.

4.5 Putative selective mechanism of resistant mutations

Along with the selective pressure arising with anti-viral treatment, immunological pressure has also been reported as a sound theory behind the emergence of resistant mutations in HIV-1 PR. The human immune system proteins, HLA (human leukocyte antigens) bind to the intracellular epitopes arising from digestion of viral proteins. HLA are responsible for presenting these epitopes on cell surface and hence triggering an immune response against the virus. Escape mutations hinders strong binding of HLA to the epitopes and thereby assist in bypassing the immune response. This is also credible in case of V77I mutation as it is located as an anchoring residue in the epitope recognized by HLA-A3 (John et al., 2005), (nonamer- LIGPTPVNI).   The score representing probability of binding and presenting of the peptide by HLA-A3 was seen to be decreased to a small extent, suggesting that the emergence of mutation V77I is preferentially due to selective pressures imposed by anti-viral therapy, and less likely due to immunological pressure.

4.6 Combinatorial library analysis

Our aim was to identify new derivatives of NFV which could be more efficient against resistant strains. For this purpose, a combinatorial library were generated based on the template of NFV. The substitution was made by different alkanes, atoms, aromatic compounds and rings. The library consisted of 35,000 compounds. In order to identify the compounds that could bind with DBM and wild protease the prepared library was docked with the crystal structure of HIV-1 PR, both wild and DBM, to screen compounds with high binding affinity. A total of 423 compounds with glide score more than -5 (in magnitude) were screened and subjected to extra precision (XP) docking protocol of glide. The top scoring compound was selected on the basis of Glide Score and Emodel score. Emodel score is used to select the top ranked pose of each ligand and present it to the user. The compound, *(3S,4aS,8aR)-2-[(2S)-3-amino-2-[(R)-[(3-hydroxy-2-methylphenyl)formamido](phenoxy)methyl]propyl]-N-tert-butyl-decahydroisoquinoline-3-carboxamide*, possessed the highest glide XP score of -13.0 Kcal/mol and Emodel score of -127.59 Kcal/mol when docked with wild protease and XP score of -10.681 Kcal/mol and Emodel score of -100.76 Kcal/mol when docked with mutant DBM (Table 5). For our convenience we will use the name NFV-A, describing it as advanced form of NFV. Structure of NFV-A is shown in figure 26.
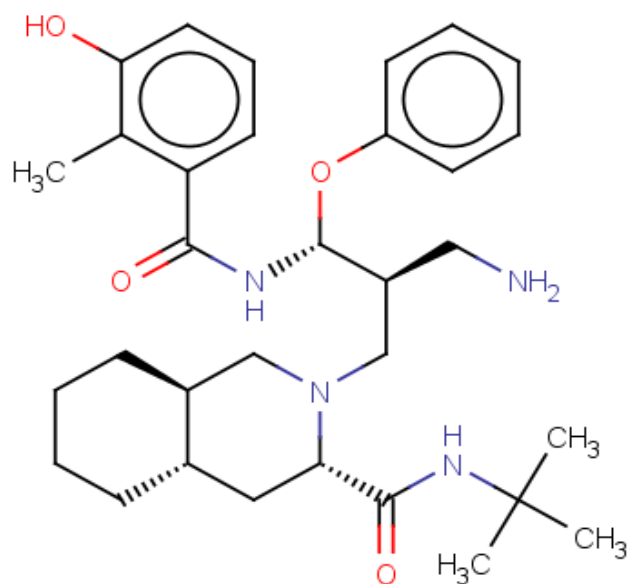
Figure 26 : Chemical structure of NFV-A.

Analysis of the components of Glide score for wild protease revealed that van der Waals energy (E_{vdw}) had the largest contribution (Table 6). On an average, contribution of van der Waals interaction energy was largest (-47.20 Kcal/mol) while coulomb energy (E_{coul}) also contributed a considerable value of -32.30 Kcal/mol. The contributions of the other terms, lipophilic interaction (Lipo = -3.08), hydrogen bonding (hbond = -0.48), penalty for freezing rotatable bonds (E_{rotb} = 0.61) and term for polar interaction at active site (E_{site} = 0.31) were negligible. For, mutant complex the values were, -36.22 Kcal/mol for van der Waals interaction energy, while coulomb energy (E_{coul}) also contributed a considerable value of -24.59 Kcal/mol. The contributions of the other terms, lipophilic interaction (Lipo = -3.15), hydrogen bonding (hbond = -0.38), penalty for freezing rotatable bonds (E_{rotb} = 0.61) and term for polar interaction at active site (E_{site} = -0.26).

Table 5: Binding affinity scores and energies of NFV-A in complex with native and DBM

| Complex | Glide Score | | Glide | Potential | Glide |
	HTVS	XP	Emodel	Energy	Energy
NFV-A-Wild	-8.00	-13.00	-127.59	180.9	-78.61
NFV-A-DBM	-6.43	-10.68	-100.76	200.29	-60.81

Table 6: Division of Glide scores into its various components.

Complex	Lipo	hbond	E_{vdw}	E_{coul}	E_{rotb}	E_{site}
NFV-A-Wild	-3.08	-0.48	-47.20	-32.30	0.61	-0.31
NFV-A-DBM	-3.15	-0.38	-36.22	-24.59	0.61	-0.26

NFV-A can be seen to find more strongly and stable with wild and NFV resistant protease, DBM. The binding is stabilized through interaction with major residues in the catalytic site of protease. The hydrogen and hydrophobic interactions made between in NFV-A and protease (wild and DBM) are shown in figure 28 . The hydrogen bonds between NFV-A and wild protease were formed by Asp 30 and Glu 27 of A chain; and Asp 25 and Ile 50 from B chain. The hydrophobic interactions involved were through Leu 23, Asp 25, Ala 28, Asp 29, Val 32, Gly 48, Gly 49 and Val 82 of A chain; and Gly 27, Ala 28, Asp 30, Val 32, Gly 48, Gly 49, Pro 81 and Val 82 of B chain.

The hydrogen bonds between NFV-A and NFV resistant DBM were formed by Asp 25 of A chain; and Asp 25 and Gly 48 from B chain. The hydrophobic interactions involved were through Arg 8, Leu 23, Asp 25, Asp 30, Val 32, Gly 48, Gly 49, Ile 50 and Val 82 of A chain; and Leu 23, Asp 25, Val 32, Gly 49, Ile 50, Pro 81 and Val 82 of B chain (Figure 27).
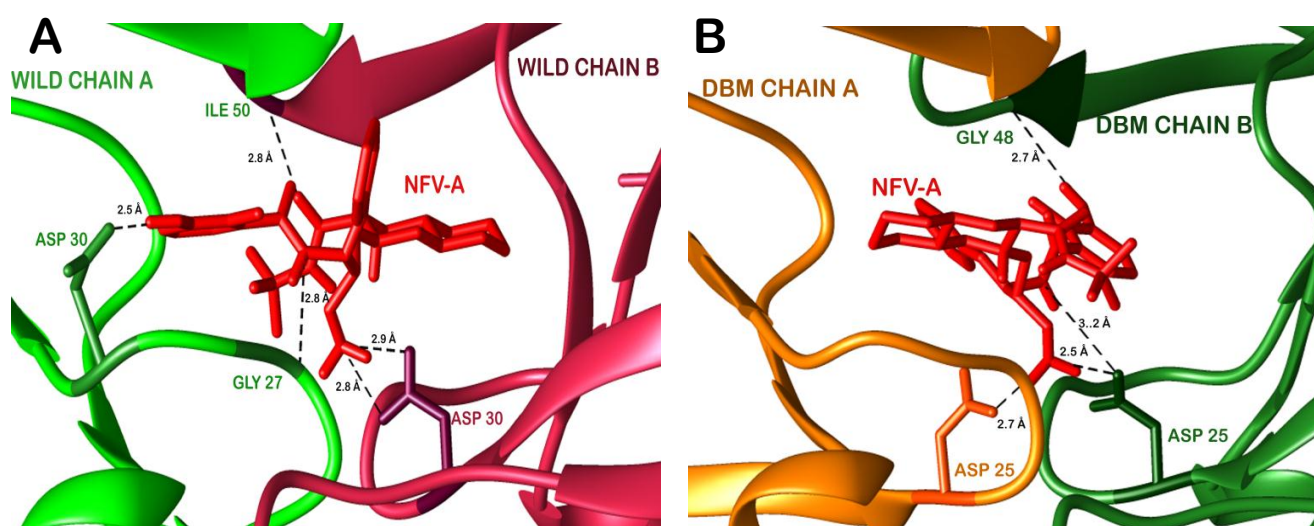
Figure 27: Hydrogen bonds between NFV-A and wild protease (A) and NFV resistant protease DBM (B)
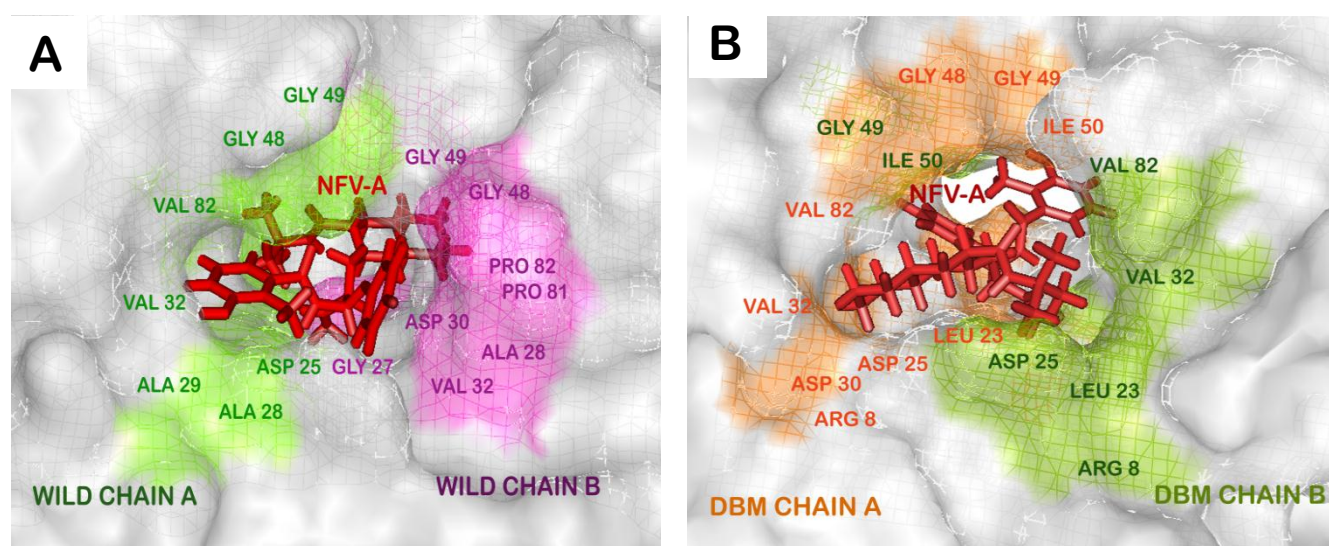
Figure 28: Hydrogen bonds between NFV-A and wild protease (A) and NFV resistant protease DBM (B)

DISCUSSION AND FUTURE PERSPECTIVE

HIV is the most harmful virus which causes heavy toll on the human immune system ("HIV epidemic -- a global update. Excerpts from the UN World AIDS Day report," 1998). In 2012, 2.3 million people were reported with HIV infection all over world. AIDS related deaths have fallen considerable since 2005 (33%), 1.6 million people died from AIDS related causes in 2012 (Kranzer et al., 2012).

Many drugs have been discovered to fight HIV infection and has been successfully used for slowing the infection process (Brinkhof et al., 2009; Rosen & Fox, 2011). But the accumulation of resistant mutations in the HIV targets have rendered these drugs unsuccessful or less effective (Vandenhende et al., 2014). There are total 50 mutation sites seen in protease which confer resistance against one or more protease inhibitors (Vergne et al., 2000). Nelfinavir is one of the FDA approved Protease inhibitor against which resistance has been seen in many clinical isolates. L23I, D30N, E35G, M46I/L/V, G48V, I54L, G73S/T/C/A, T74S, V82A/F/S/T, I84V, N88D/S, V77I, K20T/R, L33F and L90M are mutations correlated to NFV resistance (Rhee et al., 2003). The effect of these mutations has been studied extensively to understand the molecular mechanism of resistance (Meiselbach et al., 2007; Ode et al., 2006; Ode et al., 2005; Piana et al., 2002; Skalova et al., 2006).

This study explains the molecular mechanism through which V77I mutation in protease cause resistance towards NFV. Because this is a non-active site accessory mutation and clinically occurs with other resistant mutants, we have considered two types of mutant proteases- DBM and TPM. DBM stands for double mutant protease (V77I-L33F), and is more clinically prominent than TPM (V77I-L33F-K20T) (Rhee et al., 2003). DBM showed lower binding affinity towards NFV, and the mutant was more stable than the wild type. The flap opening confirmation of DBM_M suggests wider separation of flaps and more flexibility. Therefore the mutation has its effect on the equilibrium of closed and semiopen confirmations of protease and could be one reason behind the resistance showed by DBM. Further, the increased cavity size of DBM, justifies the decreased binding affinity of mutant protease for NFV due decrease in contact surface area. The residue interaction networks comparison showed that there was decrease in the interactions between flap residues with cheek residues of protease. The absence of these bonds lead to increased flexibility of flap and therefore the drug was not able to stably bind inside the cavity for longer duration.TPM showed increased affinity towards drug and therefore may be reason behind its less clinical prevalence. The interactions between flap residues and cheek residues which were lost in DBM were seen in TPM. This helped TPM to regain its limited flexibility as in wild protease. The decreased pocket size and stable flaps suggests that the combination of three mutations made HIV-1 PR non-resistant towards NFV and hence should not be selected by nature. However the clinical presence of these three mutations together suggests that, the mutant protease in nature may have been made resistant due to the presence of other mutations. The study was further extended to find a Nelfinavir derivative capable of inhibiting both wild protease and DBM. A library of NFV derivatives was generated and NFV-A is recognised as drug effectively binding to both the proteases.

Future prospective lies in recognising other co-occurring mutations which enhance the effect of V77I and understanding the molecular mechanism behind them. Development of drugs capably binding to resistant and wild proteases would help to treat and eradicate HIV, and would also considerably reduce the treatment costs.

REFERENCES

Abagyan, R., & Totrov, M. (1994). Biased probability Monte Carlo conformational searches and electrostatic calculations for peptides and proteins. *J Mol Biol, 235*(3), 983-1002.

Arthos, J., Cicala, C., Martinelli, E., Macleod, K., Van Ryk, D., Wei, D.,Fauci, A. S. (2008). HIV-1 envelope protein binds to and signals through integrin alpha4beta7, the gut mucosal homing receptor for peripheral T cells. *Nat Immunol, 9*(3), 301-309.

Arvieux, C., & Tribut, O. (2005). Amprenavir or fosamprenavir plus ritonavir in HIV infection: pharmacology, efficacy and tolerability profile. *Drugs, 65*(5), 633-659.

Batista, P. R., Wilter, A., Durham, E. H., & Pascutti, P. G. (2006). Molecular dynamics simulations applied to the study of subtypes of HIV-1 protease common to Brazil, Africa, and Asia. *Cell Biochem Biophys, 44*(3), 395-404.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Bourne, P. E. (2000). The Protein Data Bank. *Nucleic Acids Res, 28*(1), 235-242.

Boehm, H. J., Boehringer, M., Bur, D., Gmuender, H., Huber, W., Klaus, W., Mueller, F. (2000). Novel inhibitors of DNA gyrase: 3D structure based biased needle screening, hit validation by biophysical methods, and 3D guided optimization. A promising alternative to random screening. *J Med Chem, 43*(14), 2664-2674.

Brinkhof, M. W. G., Pujades-Rodriguez, M., & Egger, M. (2009). Mortality of Patients Lost to Follow-Up in Antiretroviral Treatment Programmes in Resource-Limited Settings: Systematic Review and Meta-Analysis. *PLoS ONE, 4*(6), e5790.

Brown, A. J., Korber, B. T., & Condra, J. H. (1999). Associations between amino acids in the evolution of HIV type 1 protease sequences under indinavir therapy. *AIDS Res Hum Retroviruses, 15*(3), 247-253.

Chan, D. C., & Kim, P. S. (1998). HIV entry and its inhibition. *Cell, 93*(5), 681-684.

Charpentier, C., Nora, T., Tenaillon, O., Clavel, F., & Hance, A. J. (2006). Extensive recombination among human immunodeficiency virus type 1 quasispecies makes an important contribution to viral diversity in individual patients. *J Virol, 80*(5), 2472-2482.

Chen, J., Powell, D., & Hu, W. S. (2006). High frequency of genetic recombination is a common feature of primate lentivirus replication. *J Virol, 80*(19), 9651-9658.

Douek, D. C., Roederer, M., & Koup, R. A. (2009). Emerging concepts in the immunopathogenesis of AIDS. *Annu Rev Med, 60*, 471-484.

Dundas, J., Ouyang, Z., Tseng, J., Binkowski, A., Turpaz, Y., & Liang, J. (2006). CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. *Nucleic Acids Res, 34*(Web Server issue), W116-118.

Friesner, R. A., Banks, J. L., Murphy, R. B., Halgren, T. A., Klicic, J. J., Mainz, D. T., Shenkin, P. S. (2004). Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J Med Chem, 47*(7), 1739-1749.

Garg, H., Mohl, J., & Joshi, A. (2012). HIV-1 induced bystander apoptosis. *Viruses, 4*(11), 3020-3043.

Haedicke, J., Brown, C., & Naghavi, M. H. (2009). The brain-specific factor FEZ1 is a determinant of neuronal susceptibility to HIV-1 infection. *Proc Natl Acad Sci U S A, 106*(33), 14040-14045.

Halgren, T. A., Murphy, R. B., Friesner, R. A., Beard, H. S., Frye, L. L., Pollard, W. T., & Banks, J. L. (2004). Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *J Med Chem, 47*(7), 1750-1759.

Hallenberger, S., Bosch, V., Angliker, H., Shaw, E., Klenk, H. D., & Garten, W. (1992). Inhibition of furin-mediated cleavage activation of HIV-1 glycoprotein gp160. *Nature, 360*(6402), 358-361.

Hertogs, K., Bloor, S., Kemp, S. D., Van den Eynde, C., Alcorn, T. M., Pauwels, R., Larder, B. A. (2000). Phenotypic and genotypic analysis of clinical HIV-1 isolates reveals extensive protease inhibitor cross-resistance: a survey of over 6000 samples. *Aids, 14*(9), 1203-1210.

Hiscott, J., Kwon, H., & Genin, P. (2001). Hostile takeovers: viral appropriation of the NF-kappaB pathway. *J Clin Invest, 107*(2), 143-151.

HIV epidemic -- a global update. Excerpts from the UN World AIDS Day report. (1998). *Health Millions, 24*(1), 3-5.

Hornak, V., Okur, A., Rizzo, R. C., & Simmerling, C. (2006). HIV-1 protease flaps spontaneously open and reclose in molecular dynamics simulations. *Proc Natl Acad Sci U S A, 103*(4), 915-920.

Hu, W. S., & Temin, H. M. (1990). Retroviral recombination and reverse transcription. *Science, 250*(4985), 1227-1233.

Ishima, R., Freedberg, D. I., Wang, Y. X., Louis, J. M., & Torchia, D. A. (1999). Flap opening and dimer-interface flexibility in the free and inhibitor-bound HIV protease, and their implications for function. *Structure, 7*(9), 1047-1055.

Israel, N., & Gougerot-Pocidalo, M. A. (1997). Oxidative stress in human immunodeficiency virus infection. *Cell Mol Life Sci, 53*(11-12), 864-870.

Jallow, S., Alabi, A., Sarge-Njie, R., Peterson, K., Whittle, H., Corrah, T., Janssens, W. (2009). Virological response to highly active antiretroviral therapy in patients infected with human immunodeficiency virus type 2 (HIV-2) and in patients dually infected with HIV-1 and HIV-2 in the Gambia and emergence of drug-resistant variants. *J Clin Microbiol, 47*(7), 2200-2208.

John, M., Moore, C. B., James, I. R., & Mallal, S. A. (2005). Interactive selective pressures of HLA-restricted immune responses and antiretroviral drugs on HIV-1. *Antivir Ther, 10*(4), 551-555.

Jorgensen, W. L., Maxwell, D. S., & Tirado-Rives, J. (1996). Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J Am Chem Soc, 118*(45), 11225-11236.

Kaldor, S. W., Kalish, V. J., Davies, J. F., 2nd, Shetty, B. V., Fritz, J. E., Appelt, K., . . . Tatlock, J. H. (1997). Viracept (nelfinavir mesylate, AG1343): a potent, orally bioavailable inhibitor of HIV-1 protease. *J Med Chem, 40*(24), 3979-3985.

Kaminski, G. A., Friesner, R. A., Tirado-Rives, J., & Jorgensen, W. L. (2001). Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides†. *The Journal of Physical Chemistry B, 105*(28), 6474-6487.

Kellenberger, E., Rodrigo, J., Muller, P., & Rognan, D. (2004). Comparative evaluation of eight docking tools for docking and virtual screening accuracy. *Proteins, 57*(2), 225-242.

Kohl, N. E., Emini, E. A., Schleif, W. A., Davis, L. J., Heimbach, J. C., Dixon, R. A., . . . Sigal, I. S. (1988). Active human immunodeficiency virus protease is required for viral infectivity. *Proc Natl Acad Sci U S A, 85*(13), 4686-4690.

Kranzer, K., Govindasamy, D., Ford, N., Johnston, V., & Lawn, S. D. (2012). Quantifying and addressing losses along the continuum of care for people living with HIV infection in sub-Saharan Africa: a systematic review. *J Int AIDS Soc, 15*(2), 17383.

Krausslich, H. G., & Wimmer, E. (1988). Viral proteinases. *Annu Rev Biochem, 57*, 701-754.

Leslie, A., Kavanagh, D., Honeyborne, I., Pfafferott, K., Edwards, C., Pillay, T.,Goulder, P. (2005). Transmission and accumulation of CTL escape variants drive negative associations between HIV polymorphisms and HLA. *J Exp Med, 201*(6), 891-902.

Liang, J., Edelsbrunner, H., & Woodward, C. (1998). Anatomy of protein pockets and cavities: measurement of binding site geometry and implications for ligand design. *Protein Sci, 7*(9), 1884-1897.

Little, S. J., Holte, S., Routy, J. P., Daar, E. S., Markowitz, M., Collier, A. C., . . . Richman, D. D. (2002). Antiretroviral-drug resistance among patients recently infected with HIV. *N Engl J Med, 347*(6), 385-394.

Louis, J. M., Dyda, F., Nashed, N. T., Kimmel, A. R., & Davies, D. R. (1998). Hydrophilic peptides derived from the transframe region of Gag-Pol inhibit the HIV-1 protease. *Biochemistry, 37*(8), 2105-2110.

Lyne, P. D. (2002). Structure-based virtual screening: an overview. *Drug Discov Today, 7*(20), 1047-1055.
]
Lyne, P. D., Lamb, M. L., & Saeh, J. C. (2006). Accurate prediction of the relative potencies of members of a series of kinase inhibitors using molecular docking and MM-GBSA scoring. *J Med Chem, 49*(16), 4805-4808.

Meiselbach, H., Horn, A. H., Harrer, T., & Sticht, H. (2007). Insights into amprenavir resistance in E35D HIV-1 protease mutation from molecular dynamics and binding free-energy calculations. *J Mol Model, 13*(2), 297-304.

Michod, R. E., Bernstein, H., & Nedelcu, A. M. (2008). Adaptive value of sex in microbial pathogens. *Infect Genet Evol, 8*(3), 267-285.

Miller, M., Schneider, J., Sathyanarayana, B. K., Toth, M. V., Marshall, G. R., Clawson, L., . . . Wlodawer, A. (1989). Structure of complex of synthetic HIV-1 protease with a substrate-based inhibitor at 2.3 A resolution. *Science, 246*(4934), 1149-1152.

Mitsuya, H., Yarchoan, R., & Broder, S. (1990). Molecular targets for AIDS therapy. *Science, 249*(4976), 1533-1544.

Navia, M. A., Fitzgerald, P. M., McKeever, B. M., Leu, C. T., Heimbach, J. C., Herber, W. K., . . . Springer, J. P. (1989). Three-dimensional structure of aspartyl protease from human immunodeficiency virus HIV-1. *Nature, 337*(6208), 615-620.

Nora, T., Charpentier, C., Tenaillon, O., Hoede, C., Clavel, F., & Hance, A. J. (2007). Contribution of recombination to the evolution of human immunodeficiency viruses expressing resistance to antiretroviral treatment. *J Virol, 81*(14), 7620-7628.

Ode, H., Matsuyama, S., Hata, M., Hoshino, T., Kakizawa, J., & Sugiura, W. (2007). Mechanism of drug resistance due to N88S in CRF01_AE HIV-1 protease, analyzed by molecular dynamics simulations. *J Med Chem, 50*(8), 1768-1777.

Ode, H., Matsuyama, S., Hata, M., Neya, S., Kakizawa, J., Sugiura, W., & Hoshino, T. (2007). Computational characterization of structural role of the non-active site mutation M36I of human immunodeficiency virus type 1 protease. *J Mol Biol, 370*(3), 598-607.

Ode, H., Neya, S., Hata, M., Sugiura, W., & Hoshino, T. (2006).Computational simulations of HIV-1 proteases--multi-drug resistance due to nonactive site mutation L90M. *J Am Chem Soc, 128*(24), 7887-7895.

Ode, H., Ota, M., Neya, S., Hata, M., Sugiura, W., & Hoshino, T. (2005). Resistant mechanism against nelfinavir of human immunodeficiency virus type 1 proteases. *J Phys Chem B, 109*(1), 565-574.

Piana, S., Carloni, P., & Rothlisberger, U. (2002). Drug resistance in HIV-1 protease: Flexibility-assisted mechanism of compensatory mutations. *Protein Sci, 11*(10), 2393-2402.

Platzer, K. E., Momany, F. A., & Scheraga, H. A. (1972). Conformational energy calculations of enzyme-substrate interactions. II. Computation of the binding energy for substrates in the active site of -chymotrypsin. *Int J Pept Protein Res, 4*(3), 201-219.

Pope, M., & Haase, A. T. (2003). Transmission, acute HIV-1 infection and the quest for strategies to prevent infection. *Nat Med, 9*(7), 847-852.

Prabu-Jeyabalan, M., Nalivaika, E., & Schiffer, C. A. (2000). How does a symmetric dimer recognize an asymmetric substrate? a substrate complex of HIV-1 protease. *J Mol Biol, 301*(5), 1207-1220.

Rammensee, H., Bachmann, J., Emmerich, N. P., Bachor, O. A., & Stevanovic, S. (1999). SYFPEITHI: database for MHC ligands and peptide motifs. *Immunogenetics, 50*(3-4), 213-219.

Rarey, M., Kramer, B., Lengauer, T., & Klebe, G. (1996). A fast flexible docking method using an incremental construction algorithm. *J Mol Biol, 261*(3), 470-489.

Rhee, S. Y., Gonzales, M. J., Kantor, R., Betts, B. J., Ravela, J., & Shafer, R. W. (2003). Human immunodeficiency virus reverse transcriptase and protease sequence database. *Nucleic Acids Res, 31*(1), 298-303.

Robertson, D. L., Hahn, B. H., & Sharp, P. M. (1995). Recombination in AIDS viruses. *J Mol Evol, 40*(3), 249-259.

Rosen, S., & Fox, M. P. (2011). Retention in HIV care between testing and treatment in sub-Saharan Africa: a systematic review. *PLoS Med, 8*(7), e1001056.

Schrodinger, L. (2011). Schrodinger Software Suite. *New York: Schrödinger, LLC*.

Shafer, R. W., Winters, M. A., Palmer, S., & Merigan, T. C. (1998). Multiple concurrent reverse transcriptase and protease mutations and multidrug resistance of HIV-1 isolates from heavily treated patients. *Ann Intern Med, 128*(11), 906-911.

Skalova, T., Dohnalek, J., Duskova, J., Petrokova, H., Hradilek, M., Soucek, M., . . . Hasek, J. (2006). HIV-1 protease mutations and inhibitor modifications monitored on a series of complexes. Structural basis for the effect of the A71V mutation on the active site. *J Med Chem, 49*(19), 5777-5784.

Tamalet, C., Pasquier, C., Yahi, N., Colson, P., Poizot-Martin, I., Lepeu, G., . . . Izopet, J. (2000). Prevalence of drug resistant mutants and virological response to combination therapy in patients with primary HIV-1 infection. *J Med Virol, 61*(2), 181-186.

Taylor, R. D., Jewsbury, P. J., & Essex, J. W. (2002). A review of protein-small molecule docking methods. *J Comput Aided Mol Des, 16*(3), 151-166.

Toth, G., & Borics, A. (2006). Flap opening mechanism of HIV-1 protease. *J Mol Graph Model, 24*(6), 465-474.

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

Van Marck, H., Dierynck, I., Kraus, G., Hallenberger, S., Pattery, T., Muyldermans, G., Hertogs, K. (2009). The impact of individual human immunodeficiency virus type 1 protease mutations on drug susceptibility is highly influenced by complex interactions with the background protease sequence. *J Virol, 83*(18), 9512-9520.

Vandenhende, M.-A., Bellecave, P., Recordon-Pinson, P., Reigadas, S., Bidet, Y., Bruyand, M.,. Masquelier, B. (2014). Prevalence and Evolution of Low Frequency HIV Drug Resistance Mutations Detected by Ultra Deep Sequencing in Patients Experiencing First Line Antiretroviral Therapy Failure. *PLoS ONE, 9*(1), e86771.

Vergne, L., Peeters, M., Mpoudi-Ngole, E., Bourgeois, A., Liegeois, F., Toure-Kane, C., Delaporte, E. (2000). Genetic diversity of protease and reverse transcriptase sequences in non-subtype-B human immunodeficiency virus type 1 strains: evidence of many minor drug resistance mutations in treatment-naive patients. *J Clin Microbiol, 38*(11), 3919-3925.

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

Wallace, A. C., Laskowski, R. A., & Thornton, J. M. (1995). LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng, 8*(2), 127-134.

Weber, I. T., Miller, M., Jaskolski, M., Leis, J., Skalka, A. M., & Wlodawer, A. (1989). Molecular modeling of the HIV-1 protease and its substrate binding site. *Science, 243*(4893), 928-931.

Weiss, R. A. (1993). How does HIV cause AIDS? *Science, 260*(5112), 1273-1279.

Wlodawer, A., Miller, M., Jaskolski, M., Sathyanarayana, B. K., Baldwin, E., Weber, I. T., Kent, S. B. (1989). Conserved folding in retroviral proteases: crystal structure of a synthetic HIV-1 protease. *Science, 245*(4918), 616-621.

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

Wyatt, R., & Sodroski, J. (1998). The HIV-1 envelope glycoproteins: fusogens, antigens, and immunogens. *Science, 280*(5371), 1884-1888.

Zheng, Y. H., Lovsin, N., & Peterlin, B. M. (2005). Newly identified host factors modulate HIV replication. *Immunol Lett, 97*(2), 225-234.

APPENDIX

1. INTRA-MOLECULAR INTERACTIONS PRESENT IN WILD-PROTEASE, DBM AND TPM.

INTRA-MOLECULAR INTERACTIONS BY WILD-PROTEASE RESIDUES (18-22, 31-35, 75-79)

Donor	Acceptor	Hydrogen	D..A	D-H..A
VAL 11.A N	ALA 22.A O	VAL 11.A H	2.758	1.868
ILE 13.A N	LYS 20.A O	ILE 13.A H	2.756	1.838
ILE 15.A N	GLN 18.A O	ILE 15.A H	2.857	1.915
GLN 18.A N	ILE 15.A O	GLN 18.A H	2.856	1.910
LYS 20.A N	ILE 13.A O	LYS 20.A H	2.978	2.056
ALA 22.A N	VAL 11.A O	ALA 22.A H	2.708	1.829
VAL 32.A N	ILE 84.A O	VAL 32.A H	3.041	2.181
LEU 33.A N	LEU 76.A O	LEU 33.A H	2.842	1.879
GLU 34.A N	ASN 83.A OD1	GLU 34.A H	2.798	1.838
ARG 57.A N	VAL 77.A O	ARG 57.A H	2.802	1.833
TYR 59.A N	VAL 75.A O	TYR 59.A H	2.800	1.843
VAL 75.A N	TYR 59.A O	VAL 75.A H	2.852	1.903
LEU 76.A N	THR 31.A O	LEU 76.A H	2.798	1.848
VAL 77.A N	ARG 57.A O	VAL 77.A H	2.784	1.863
GLY 78.A N	LEU 33.A O	GLY 78.A H	2.870	1.908
THR 80.A N	GLY 78.A O	THR 80.A H	2.856	2.013
ASN 83.A ND2	GLU 21.A O	ASN 83.A HD22	3.115	2.170
ILE 84.A N	VAL 32.A O	ILE 84.A H	2.813	1.844
VAL 11.B N	ALA 22.B O	VAL 11.B H	2.711	1.798
ILE 13.B N	LYS 20.B O	ILE 13.B H	2.762	1.809
ILE 15.B N	GLN 18.B O	ILE 15.B H	2.922	1.985
GLN 18.B N	ILE 15.B O	GLN 18.B H	2.789	1.905
LYS 20.B N	ILE 13.B O	LYS 20.B H	2.805	1.917
ALA 22.B N	VAL 11.B O	ALA 22.B H	2.845	1.963
VAL 32.B N	ILE 84.B O	VAL 32.B H	2.968	2.097
LEU 33.B N	LEU 76.B O	LEU 33.B H	2.756	1.815
GLU 34.B N	ASN 83.B OD1	GLU 34.B H	2.860	1.890
ARG 57.B N	VAL 77.B O	ARG 57.B H	2.963	2.020
ARG 57.B NH1	GLU 35.B OE1	ARG 57.B 2HH1	3.031	2.098
ARG 57.B NH1	GLU 35.B OE2	ARG 57.B 2HH1	2.967	2.110
ARG 57.B NH2	GLU 35.B OE2	ARG 57.B 2HH2	2.937	2.075
TYR 59.B N	VAL 75.B O	TYR 59.B H	2.860	1.886
VAL 75.B N	TYR 59.B O	VAL 75.B H	3.003	2.052
LEU 76.B N	THR 31.B O	LEU 76.B H	2.935	1.973
VAL 77.B N	ARG 57.B O	VAL 77.B H	2.873	1.952
GLY 78.B N	LEU 33.B O	GLY 78.B H	3.119	2.181
THR 80.B N	GLY 78.B O	THR 80.B H	2.765	1.913
ASN 83.B ND2	GLU 21.B O	ASN 83.B HD22	2.946	1.983
ILE 84.B N	VAL 32.B O	ILE 84.B H	2.778	1.814

Intra-molecular Interactions by DBM residues (18-22, 31-35, 75-79)

Donor	Acceptor	Hydrogen	D..A	D-H..A
VAL 11.A N	ALA 22.A O	VAL 11.A H	2.784	1.786
ILE 13.A N	LYS 20.A O	ILE 13.A H	2.742	1.867
ILE 15.A N	GLN 18.A O	ILE 15.A H	3.111	2.123
GLN 18.A N	ILE 15.A O	GLN 18.A H	3.093	2.142
LYS 20.A N	ILE 13.A O	LYS 20.A H	3.034	2.058
LYS 20.A NZ	GLU 34.A OE2	LYS 20.A HZ1	2.792	1.835
ALA 22.A N	VAL 11.A O	ALA 22.A H	2.934	1.965
THR 31.A OG1	ASN 88.A OD1	THR 31.A HG1	3.166	2.269
VAL 32.A N	ILE 84.A O	VAL 32.A H	2.943	1.989
PHE 33.A N	LEU 76.A O	PHE 33.A H	2.889	1.983
GLU 34.A N	ASN 83.A OD1	GLU 34.A H	2.615	1.606
ARG 57.A N	ILE 77.A O	ARG 57.A H	2.791	1.792
ARG 57.A NH1	GLU 35.A OE1	ARG 57.A HH12	3.277	2.466
ARG 57.A NH1	GLU 35.A OE2	ARG 57.A HH12	2.589	1.609
ARG 57.A NH2	GLU 35.A OE1	ARG 57.A HH22	2.679	1.676
ARG 57.A NH2	GLU 35.A OE2	ARG 57.A HH22	3.545	2.797
TYR 59.A N	VAL 75.A O	TYR 59.A H	3.001	2.025
VAL 75.A N	TYR 59.A O	VAL 75.A H	2.844	1.845
LEU 76.A N	THR 31.A O	LEU 76.A H	2.923	1.914
ILE 77.A N	ARG 57.A O	ILE 77.A H	2.710	1.740
GLY 78.A N	PHE 33.A O	GLY 78.A H	3.174	2.202
ASN 83.A ND2	GLU 21.A O	ASN 83.A HD22	3.590	2.769
ASN 83.A ND2	GLU 34.A OE2	ASN 83.A HD21	2.709	1.708
ILE 84.A N	VAL 32.A O	ILE 84.A H	2.819	1.827
GLY 86.A N	THR 31.A OG1	GLY 86.A H	2.749	1.758
VAL 11.B N	ALA 22.B O	VAL 11.B H	2.944	1.945
ILE 13.B N	LYS 20.B O	ILE 13.B H	2.756	1.752
ILE 15.B N	GLN 18.B O	ILE 15.B H	3.519	2.541
GLN 18.B NE2	ILE 15.B O	GLN 18.B HE22	3.257	2.679
LYS 20.B N	ILE 13.B O	LYS 20.B H	3.046	2.038
LYS 20.B NZ	GLU 34.B OE1	LYS 20.B HZ2	2.515	1.533
ALA 22.B N	VAL 11.B O	ALA 22.B H	3.045	2.055
THR 31.B OG1	ASN 88.B OD1	THR 31.B HG1	2.895	2.018
VAL 32.B N	ILE 84.B O	VAL 32.B H	2.756	1.818
PHE 33.B N	LEU 76.B O	PHE 33.B H	2.741	1.756
GLU 34.B N	ASN 83.B OD1	GLU 34.B H	2.926	1.956
ARG 57.B N	ILE 77.B O	ARG 57.B H	2.700	1.708
ARG 57.B NH1	GLU 35.B OE2	ARG 57.B HH12	2.901	1.901
ARG 57.B NH2	GLU 35.B OE1	ARG 57.B HH22	2.660	1.662
TYR 59.B N	VAL 75.B O	TYR 59.B H	2.969	1.966
VAL 75.B N	TYR 59.B O	VAL 75.B H	3.031	2.041
LEU 76.B N	THR 31.B O	LEU 76.B H	2.769	1.829
ILE 77.B N	ARG 57.B O	ILE 77.B H	2.764	1.780
GLY 78.B N	PHE 33.B O	GLY 78.B H	3.289	2.380
ASN 83.B ND2	GLU 21.B O	ASN 83.B HD22	3.533	2.888
ASN 83.B ND2	GLU 34.B OE1	ASN 83.B HD21	2.933	1.936
ILE 84.B N	VAL 32.B O	ILE 84.B H	2.595	1.585
GLY 86.B N	THR 31.B OG1	GLY 86.B H	2.831	1.825

Intra-molecular Interactions by TPM residues (18-22, 31-35, 75-79)

Donor	Acceptor	Hydrogen	D..A	D-H..A
VAL 11.A N	ALA 22.A O	VAL 11.A H	3.159	2.191
ILE 13.A N	THR 20.A O	ILE 13.A H	2.979	1.990
THR 20.A N	ILE 13.A O	THR 20.A H	3.223	2.215
ALA 22.A N	VAL 11.A O	ALA 22.A H	2.982	2.022
THR 31.A OG1	ASN 88.A OD1	THR 31.A HG1	2.764	1.837
VAL 32.A N	ILE 84.A O	VAL 32.A H	2.951	2.122
PHE 33.A N	LEU 76.A O	PHE 33.A H	2.611	1.629
GLU 34.A N	ASN 83.A OD1	GLU 34.A H	3.322	2.344
MET 36.A N	GLU 34.A O	MET 36.A H	2.817	1.918
ARG 57.A N	ILE 77.A O	ARG 57.A H	3.092	2.123
ARG 57.A NE	GLU 35.A OE1	ARG 57.A HE	3.468	2.682
ARG 57.A NH2	GLU 35.A OE1	ARG 57.A HH21	2.771	1.766
TYR 59.A N	VAL 75.A O	TYR 59.A H	3.133	2.125
VAL 75.A N	TYR 59.A O	VAL 75.A H	3.286	2.334
LEU 76.A N	THR 31.A O	LEU 76.A H	3.189	2.245
ILE 77.A N	ARG 57.A O	ILE 77.A H	2.835	1.894
GLY 78.A N	PHE 33.A O	GLY 78.A H	2.821	1.853
ASN 83.A N	GLU 21.A O	ASN 83.A H	3.556	2.567
ASN 83.A ND2	GLU 21.A O	ASN 83.A HD22	3.430	2.558
ILE 84.A N	VAL 32.A O	ILE 84.A H	2.803	1.896
VAL 11.B N	ALA 22.B O	VAL 11.B H	3.154	2.162
ILE 13.B N	THR 20.B O	ILE 13.B H	2.892	1.899
ILE 15.B N	GLN 18.B O	ILE 15.B H	3.087	2.151
THR 20.B N	ILE 13.B O	THR 20.B H	3.033	2.024
GLU 21.B N	THR 20.B OG1	GLU 21.B H	2.868	2.049
ALA 22.B N	VAL 11.B O	ALA 22.B H	3.123	2.208
THR 31.B OG1	ASN 88.B OD1	THR 31.B HG1	2.840	1.979
VAL 32.B N	ILE 84.B O	VAL 32.B H	2.887	1.973
PHE 33.B N	LEU 76.B O	PHE 33.B H	2.768	1.778
GLU 34.B N	ASN 83.B OD1	GLU 34.B H	2.789	1.799
ARG 57.B N	ILE 77.B O	ARG 57.B H	2.795	1.790
ARG 57.B NH1	GLU 35.B OE1	ARG 57.B HH12	2.630	1.624
ARG 57.B NH2	GLU 35.B OE2	ARG 57.B HH22	2.821	1.825
TYR 59.B N	VAL 75.B O	TYR 59.B H	2.844	1.840
VAL 75.B N	TYR 59.B O	VAL 75.B H	2.904	1.897
LEU 76.B N	THR 31.B O	LEU 76.B H	3.037	2.033
ILE 77.B N	ARG 57.B O	ILE 77.B H	2.928	1.974
ASN 83.B ND2	GLU 21.B O	ASN 83.B HD22	3.222	2.229
ILE 84.B N	VAL 32.B O	ILE 84.B H	2.703	1.730

2. RMSD VALUES OF WILD PROTEASE, DBM AND TPM

TIMESTEP	TIME(ns)	WILD	TPM	DBM
0	0.0	0.00	0.00	0.00
160	0.2	0.93	1.10	1.04
360	0.4	0.94	1.16	1.08
560	0.6	1.04	1.32	1.02
760	0.8	1.17	1.26	1.19
960	1.0	1.19	1.22	1.16
1160	1.2	1.13	1.36	1.20
1360	1.4	1.05	1.29	1.17
1560	1.6	1.09	1.18	1.27
1760	1.8	1.18	1.08	1.10
1960	2.0	1.24	1.12	1.06
2160	2.2	1.24	1.20	1.22
2360	2.4	1.18	1.18	1.25
2561	2.6	1.39	1.28	1.20
2761	2.8	1.28	1.29	1.27
2961	3.0	1.28	1.32	1.30
3161	3.2	1.19	1.23	1.19
3361	3.4	1.36	1.31	1.25
3561	3.6	1.35	1.26	1.26
3761	3.8	1.37	1.24	1.30
3961	4.0	1.27	1.23	1.09
4161	4.2	1.43	1.27	1.21
4361	4.4	1.50	1.24	1.15
4561	4.6	1.81	1.36	1.28
4761	4.8	1.52	1.57	1.27
4961	5.0	1.63	1.37	1.41
5161	5.2	1.58	1.24	1.28
5361	5.4	1.86	1.44	1.30
5561	5.6	1.83	1.28	1.24
5761	5.8	1.52	1.31	1.65
5961	6.0	1.84	1.42	1.38

6161	6.2	1.44	1.37	1.36
6361	6.4	1.51	1.31	1.20
6561	6.6	1.43	1.46	1.18
6761	6.8	1.55	1.33	1.24
6961	7.0	1.49	1.43	1.13
7161	7.2	1.61	1.25	1.22
7361	7.4	1.58	1.26	1.09
7562	7.6	1.34	1.28	1.20
7762	7.8	1.50	1.21	1.12
7962	8.0	1.59	1.30	1.29
8162	8.2	1.57	1.43	1.34
8362	8.4	1.12	1.32	1.25
8562	8.6	1.43	1.45	1.13
8762	8.8	1.45	1.37	1.24
8962	9.0	1.33	1.61	1.37
9162	9.2	1.40	1.45	1.31
9362	9.4	1.48	1.33	1.24
9562	9.6	1.55	1.51	1.40
9762	9.8	1.55	1.52	1.45
9962	10.0	1.18	1.49	1.29
10162	10.2	1.40	1.66	1.31
10362	10.4	1.33	1.70	1.58
10562	10.6	1.36	1.52	1.45
10762	10.8	1.50	1.55	1.61
10962	11.0	1.23	1.54	1.72
11162	11.2	1.11	1.46	1.60
11362	11.4	1.25	1.67	1.62
11562	11.6	1.23	1.52	1.96
11762	11.8	1.09	1.47	1.48
11962	12.0	1.46	1.45	1.44
12162	12.2	1.32	1.41	1.33
12362	12.4	1.47	1.31	1.43
12563	12.6	1.47	1.35	1.34

12763	12.8	1.57	1.29	1.31
12963	13.0	1.38	1.46	1.29
13163	13.2	1.29	1.53	1.81
13363	13.4	1.45	1.31	1.89
13563	13.6	1.73	1.53	1.66
13763	13.8	1.60	1.53	1.71
13963	14.0	1.46	1.50	1.56
14163	14.2	1.89	1.73	1.55
14363	14.4	1.59	1.58	1.45
14563	14.6	1.38	1.46	1.82
14763	14.8	1.62	1.43	1.60
14963	15.0	1.39	1.42	1.60
15163	15.2	1.21	1.47	1.55
15363	15.4	1.54	1.45	1.43
15563	15.6	1.45	1.47	1.64
15763	15.8	1.70	1.45	1.75
15963	16.0	1.62	1.56	1.63
16163	16.2	1.84	1.41	1.74
16363	16.4	1.69	1.35	1.67
16563	16.6	2.12	1.58	1.87
16763	16.8	1.76	1.38	1.77
16963	17.0	1.76	1.35	1.59
17163	17.2	1.73	1.57	1.70
17363	17.4	1.78	1.43	1.69
17564	17.6	1.56	1.39	1.68
17764	17.8	1.72	1.43	1.73
17964	18.0	1.86	1.44	1.63
18164	18.2	1.95	1.51	1.61
18364	18.4	2.00	1.57	1.49
18564	18.6	2.00	1.73	1.70
18764	18.8	1.91	1.67	1.43
18964	19.0	1.98	1.45	1.64
19164	19.2	2.21	1.54	1.35

19364	19.4	2.01	1.55	1.63
19564	19.6	2.22	1.46	1.57
19764	19.8	1.80	1.45	1.72
19964	20.0	1.94	1.54	1.57

3. RMSD VALUES OF WILD, DBM AND TPM DOCKED WITH NFV

TIMESTEP	TIME(ns)	NFV:DBM	NFV:TPM
0	0	0.00	0.00
200.04	0.20	0.88	1.02
400.08	0.40	1.02	1.10
600.12	0.60	1.02	1.16
800.16	0.80	1.18	1.18
1000.2	1.00	1.20	1.22
1200.24	1.20	1.51	1.32
1400.28	1.40	1.29	1.39
1600.32	1.60	1.32	1.27
1800.36	1.80	1.64	1.15
2000.4	2.00	1.53	1.21
2200.44	2.20	1.37	1.09
2400.48	2.40	1.41	1.18
2600.52	2.60	1.45	1.36
2800.56	2.80	1.43	1.27
3000.6	3.00	1.60	1.16
3200.64	3.20	1.50	1.21
3400.68	3.40	1.43	1.34
3600.72	3.60	1.53	1.38
3800.76	3.80	1.56	1.24
4000.8	4.00	1.47	1.37
4200.84	4.20	1.65	1.26
4400.88	4.40	1.66	1.26
4600.92	4.60	1.52	1.29
4800.96	4.80	1.42	1.49
5001	5.00	1.51	1.44
5201.04	5.20	1.49	1.26
5401.08	5.40	1.59	1.33
5601.12	5.60	1.47	1.37
5801.16	5.80	1.46	1.35
6001.2	6.00	1.58	1.32
6201.24	6.20	1.55	1.30
6401.28	6.40	1.52	1.24

6601.32	6.60	1.49	1.28
6801.36	6.80	1.62	1.32
7001.4	7.00	1.47	1.32
7201.44	7.20	1.52	1.20
7401.48	7.40	1.60	1.22
7601.52	7.60	1.70	1.21
7801.56	7.80	1.49	1.32
8001.6	8.00	1.46	1.44
8201.64	8.20	1.58	1.31
8401.68	8.40	1.54	1.27
8601.72	8.60	1.44	1.30
8801.76	8.80	1.53	1.50
9001.8	9.00	1.66	1.78
9201.84	9.20	1.61	1.49
9401.88	9.40	1.45	1.38
9601.92	9.60	1.69	1.46
9801.96	9.80	1.49	1.46
10000	10.00	1.70	1.67

4. HYDROGEN AND HYDROPHOBIC INTERACTION RESIDUES RESPONSIBLE FOR NFV DOCKING IN DBM, TPM AND WILD PROTEASE.

Complex	Hydrogen Bond	Hydrophobic Interaction
Wild	Asp 25 (A), Asp 25(B), Asp 30 (A), Gly 27 (A)	Gly 49 (A)(B), Asp 29 (A), Ala 28(A), Leu 23 (B), Ile 84 (A) (B), Ile 50 (A)(B), Val 32 (B), Pro 81 (A)(B), Gly 27 (B) Val 82 (A)(B), Gly 48 (B)
DBM	Gly 27(B), Asp 25 (B)	Ile 84 (A)(B), Ala28(A)(B), Gly 27 (A), Leu 23 (A)(B), Pro 81 (B), Thr 54 (B), Gly 49 (B), Ile 50 (A), Gly 48(B), Arg 8(A), Asp 29 (B,) Asp 30 (B)
TPM	Asp 25(A)(B)	Asp 25(B), Ala 28(A), Leu 23 (A)(B), Val 32(B), Gly 48(B), Pro 81(B), Thr 80 (B), Ile 84 (A)(B)

5. DISTANCE BETWEEN I50 OF A-CHAIN AND B-CHAIN IN 5ns OF MD SIMULATION

FRAME	TIME (ns)	TPM_M	WILD	DBM_M
18	0.18	20.105083	17.458467	20.64528
38	0.38	19.747227	23.035646	21.6803
58	0.58	18.637041	21.500481	22.34336
78	0.78	19.492489	22.345636	25.54836
98	0.98	20.333569	20.762596	26.17736
118	1.18	19.063282	19.520779	23.64928
138	1.38	20.509741	22.537935	23.92057
158	1.58	20.652126	21.100128	23.63934
178	1.78	19.783588	24.312254	23.44402
198	1.98	20.405014	25.72109	19.53445
218	2.18	20.07478	25.238594	16.97504
238	2.38	21.796947	21.970781	22.22749
258	2.58	20.105846	20.47533	19.82522
278	2.78	20.028631	19.69668	22.4097
298	2.98	18.911804	21.056171	18.51631
318	3.18	18.69602	20.04405	21.66754
338	3.38	18.357105	18.883821	21.23491
358	3.58	18.638906	20.188974	20.05277
378	3.78	19.955084	20.622522	16.21681
398	3.98	19.92346	18.839596	15.77297
418	4.18	20.743917	19.290247	16.49675
438	4.38	24.837502	18.596037	16.00123
458	4.58	22.059246	18.062082	18.28542
478	4.78	21.833754	17.991043	17.61212
498	4.98	21.924614	21.61314	16.39381

6. LEGEND FOR RESIDUE INTERACTION NETWORKS

EDGE COLOR LEGEND IN RESIDUE INTERACTION NETWORKS

Edge Color	interaction
———	hbond:mc-mc
———	hbond:mc-sc
———	iac:mc-mc
———	iac:mc-sc
———	iac:sc-sc
———	ionic:sc-sc
———	pb:mc-mc

hbond : hydrogen bond

iac : interactomic contacts

ionic : ionic bonds

pb : polar bonds

mc : main chain

sc : side chain

EDGE LINE STYLE LEGEND IN RESIDUE INTERACTION NETWORKS

Edge Line Style	BelongsTo
———	both
- - - - - -	net1
··············	net2

net 1: residue interaction network of wild protease

net 2 : residue interaction network of mutant protease

NODE COLOR LEGEND IN RESIDUE INTERACTION NETWORKS

NODE SIZE LEGEND IN RESIDUE INTERACTION NETWORKS