Thermodynamics and Biophysical Basis of Drug Resistance in HIV-1 Protease via Multiscale Simulations

M.Sc. Thesis

By

HARIDEV S R

DISCIPLINE OF PHYSICS

INDIAN INSTITUTE OF TECHNOLOGY INDORE

JUNE 2018

Thermodynamics and Biophysical Basis of Drug Resistance in HIV-1 Protease via Multiscale Simulations

A THESIS

Submitted in partial fulfillment of the requirements for the award of the degree

of

Master of Science

by

HARIDEV S R

DISCIPLINE OF PHYSICS

INDIAN INSTITUTE OF TECHNOLOGY INDORE

JUNE 2018

INDIAN INSTITUTE OF TECHNOLOGY INDORE

CANDIDATE'S DECLARATION

I hereby certify that the work which is being presented in the thesis entitled **Thermodynamics and Biophysical Basis of Drug Resistance in HIV-1 Protease via Multiscale Simulations** in the partial fulfillment of the requirements for the award of the degree of **MASTER OF SCIENCE** and submitted in the **DISCIPLINE OF PHYSICS, Indian Institute of Technology Indore**, is an authentic record of my own work carried out during the time period from JULY 2016 to JUNE 2018 under the supervision of Dr. Parimal Kar Assistant Professor, Discipline of Bioscience and Biomedical Engineering, IIT Indore and Dr. Somaditya Sen Associate Professor, Department of Physics, IIT Indore.

The matter presented in this thesis has not been submitted by me for the award of any other degree of this or any other institute.

Signature of the student with date

(HARIDEV S R)

This is to certify that the above statement made by the candidate is correct to the best of my/our knowledge.

Signature of the Supervisor of
M.Sc. thesis (with date)
(DR.PARIMAL KAR)

Signature of the Supervisor of
M.Sc. thesis (with date)
(DR.SOMADITYA SEN)

HARIDEV S R has successfully given his/her M.Sc. Oral Examination held on **21-JUNE-2018**.

Signature(s) of Supervisor(s) of MSc thesis
Date:

Convener, DPGC
Date:

Signature of PSPC Member #1
Date:

Signature of PSPC Member #1
Date:

ACKNOWLEDGEMENTS

I would like to extend my heartfelt thanks to my advisor Dr. Parimal Kar for his continuous support, patience, and immense knowledge. As a newcomer to the field of computational biophysics, it is his encouragement and friendly nature that help me to do the project with great enthusiasm. The weekly journal club and the group discussions in the lab have created an excellent platform for me to learn many things as a part of my MSc project. It was indeed a great atmosphere in his lab for learning new things.

I would also like to take this opportunity to thank the rest of my thesis committee members, Dr. Somaditya Sen, Prof. Sarika Jalan and Dr. Sharad Gupta for their insightful comments which help me to widen my knowledge in the field.

I thank my fellow lab mates- Mrs. Nisha Amarnath Jonniya, Mr. Rajarshi Roy, Mr. MD Fulbabu SK and Mr. Nima Tshering Bhutia for stimulating discussions and helping me continuously whenever I got stuck in my work. Finally, timely birthday celebrations and nonacademic talks with them have created a homely environment to rejoice.

Abstract

Drug resistance due to mutation has sharply limited the effectiveness of HIV-1 protease inhibitors in AIDS therapy. It is critically important to understand the molecular basis of drug resistance for designing new drugs. Elucidating the dynamic nature and thermodynamic basis of binding of drugs to wild-type and mutant variants of protease could be insightful, for the development of resistance-evading drugs. In this study, we have conducted molecular dynamics simulations in combination with the free energy calculation for elucidating the mechanism of binding of the inhibitor TMC-126 to HIV-1 protease. Five mutant variants (A28S, V32I, M46L, I50V, and MDR20) and HIV-2 protease are also considered. The popular and widely used Molecular Mechanics-Poisson Boltzmann Surface Area (MM-PBSA) method is utilized to calculate the free energy of binding and the normal mode analysis is performed for estimating the entropic contribution to the binding free energy.

From our study it is observed that for all cases, the binding is mainly driven by the van der Waals interactions. Furthermore, it is observed that the intermolecular electrostatic interactions and the nonpolar solvation free energy also contribute favourably to the binding free energy. However, the intermolecular electrostatic interaction is over-compensated by the unfavorable polar solvation free energy.

The inhibitor is found to be losing its potency against all five mutant variants. A significant decrease in the binding free energy is observed for A28S and V32I mutations. Our study suggests that the mutation-induced drug resistance arises mainly because of decrease in intermolecular electrostatic interactions compared to the wild-type. Over all, the current study elucidates the biophysical basis of drug resistance and may help in designing new drugs that can be effective against mutant variants.

TABLE OF CONTENTS

LIST OF FIGURES

and PR2 compared to its WT.

LIST OF TABLES

ACRONYMS

AIDS - Acquired Immunodeficiency Syndrome

HIV - Human Immunodeficiency Virus

DNA - Deoxyribonucleic acid

PR - Protease

MD -Molecular Dynamics

MDR -Multidrug Resistance

WT -Wild Type

AMBER -Assisted Model Building with Energy Refinement

GAFF -General Amber Force Field

MM-PBSA - Molecular Mechanics Poisson Boltzmann Surface area

CHARMM -Chemistry at HARvard using Molecular Mechanics

OPLS -Optimized Potentials for Liquid Simulations

GB -Generalised Born

FEP - Free Energy Perturbation

TI -Thermodynamic Integration

PBC -Periodic Boundary Condition

PME -Particle Mesh Ewald

RMSD -Root Mean Square Deviation

RMSF -Root Mean Square Fluctuation

DCCM -Dynamic Cross Correlation Matrix

Chapter 1

INTRODUCTION

1.1 HIV/AIDS

In 1981 the first case of person having collection of opportunistic infections and tumours that were normally suppressed by immune system was identified. Later, more number of persons having the same conditions were reported. On further studies it is understood that the opportunistic infections are associated with the marked decrease in the circulating T-helper lymphocytes, so the term Acquired Immunodeficiency Syndrome (AIDS) was coined to describe the state of opportunistic infection due to the suppression of immune system. In 1983 one retrovirus isolated from the patients blood called as 'Human Immunodeficiency Virus type 1 (HIV-1)' is found to be the reason for the disease.

The virus targets the immune system of the host and leaves the infected vulnerable to otherwise defensible diseases. AIDS spreads rapidly than most other diseases in recent history which make this a global pandemic since first case back in 1981. An increased understanding about the life cycle and functions of HIV leads to a better treatment of the patients. According to UNAIDS 2016 report, 36.7 million people are still living with this disease. Also 1.8 million [1.6 million–2.1 million] people became newly infected and 1 million [0.8–1.2 million] people died from AIDS-related illnesses in 2016. An in-depth understanding of the virus life cycle and mechanism of binding of drugs to the target proteins of HIV is the best way to fight the disease.

1.2 Types of HIV

There are two major type of human immunodeficiency viruses. HIV-1, the one which discovered first is the most widespread type. The second one HIV 2 is a different type of retrovirus originating from different primate species [1]. HIV 2 is more than 55% genetically different from HIV 1, however, both are showing similar transmission mode in body and infection with any of these strands can leads to AIDS[2]. Although the majority of HIV-2 infections have been found in West Africa, the number of diagnosed cases in India is increasing. Compare to HIV 1, HIV 2 shows lower incidence rate and transmission risk as the viral load tends to remain lower for longer time which make it less infectious compared to HIV 1 [3].

Both the HIV 1 and HIV 2 are further divided into subtypes. There are at least three distinct HIV 1 groups: M (main), O (outlier), and N (non-M/non-O). About 90% of HIV-1 infections are classified as group M and these are distributed worldwide. Group O infections are endemic to several west central African countries and represent 1 to 5% of all HIV-1 infection in those areas. Group N has only been identified in a small number of individuals. Depending on the geographical pattern group M is further divided into more than ten subtypes including A, B, C, D, F, G, H etc. Among this subtype A and C are most common world wide and B is common in UK. Additionally, different subtypes can combine genetic material to form a hybrid virus, known as a 'circulating recombinant form' (CRFs), of which at least twenty have been identified. Similarly there are at least eight different subtypes of HIV-2 among which A and C are common.

1.3 HIV Genome, Structure and Life Cycle

The genome of HIV is encoded in positive-sense single-stranded RNA and replicates through double-stranded DNA intermediates. The two identical RNA strands enclosed in the core consists of about 9,800 nucleotides.

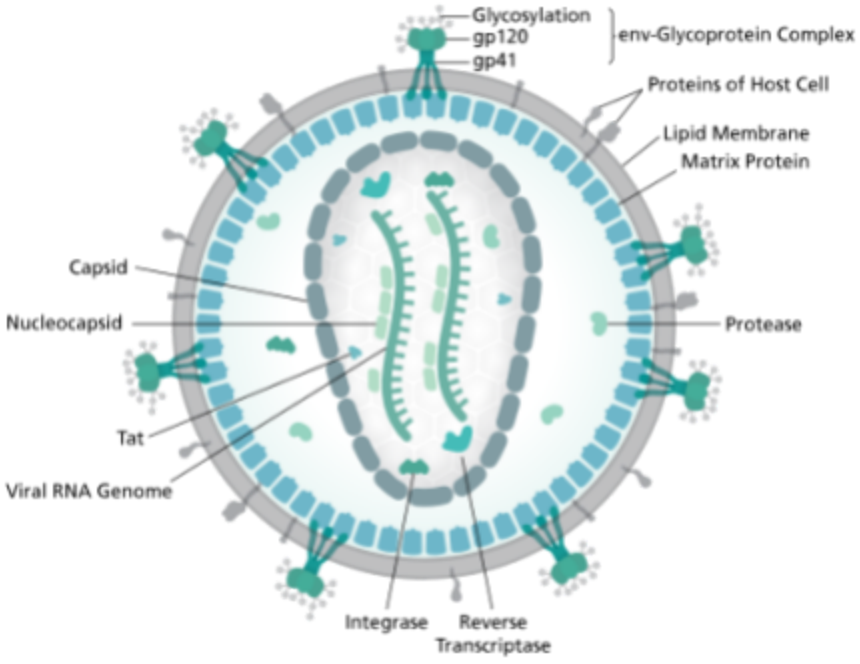


Fig.1. Cross Section of Virion [90]

These nucleotides encode for all the structural (gag, pol, and env) and nonstructural accessory genes (tat, rev, nef, vif, vpr, and vpu). In HIV 2 vpu is replaced by vpx which seems to be one reason for its less pathogenicity [4]. Both ends of the DNA genome is guarded with LTR (Long Term Repeat) sequences. Reading from the 5' end, the 5' LTR is coding for the promoter for the transcription of the viral gene. Then comes the gag gene which is coding for outer core membrane (MA, p17), the capsid protein (CA, p24), the nucleocapsid (NC, p7) and a smaller, nucleic acid-stabilizing protein. The gag reading frame is followed by which is coding for the enzymes protease (PR, p12), reverse transcriptase (RT, p51) and RNase H (p15) or RT plus RNase H (together p66) and integrase (IN, p32). Then comes the env reading frame which is coding for the two envelope

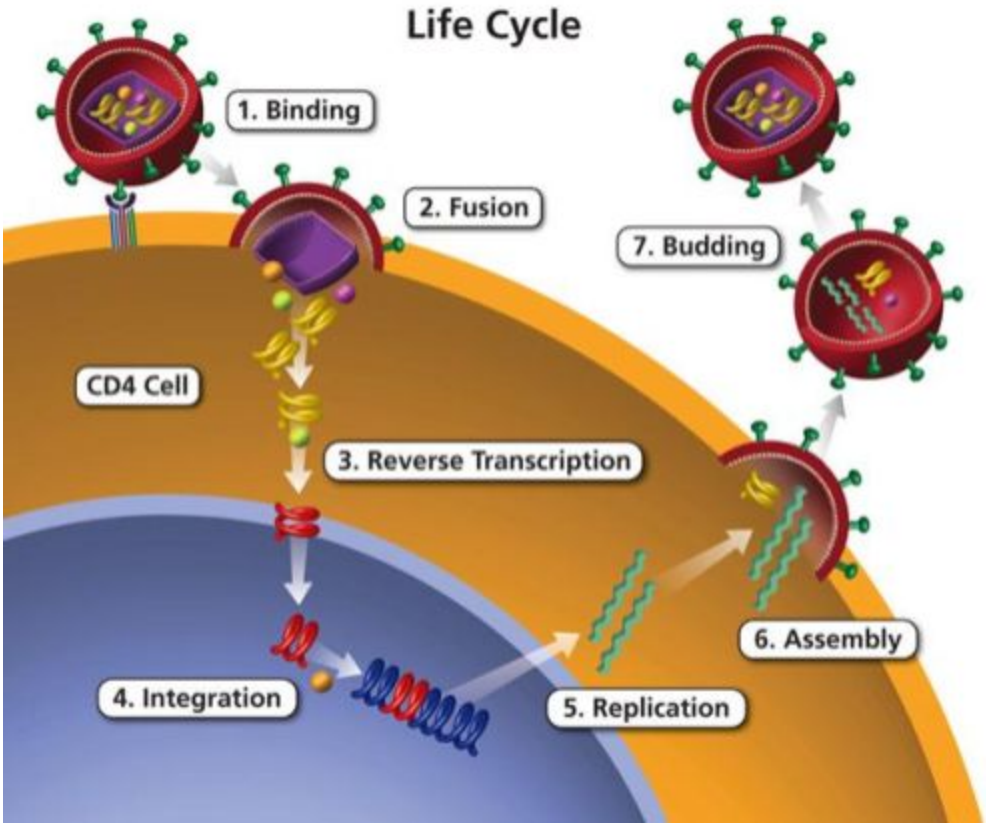


Fig.2. Showing different stages of HIV life cycle [91]

glycoproteins gp120 (surface protein, SU) and gp41 (transmembrane protein, TM) is derived. The HIV genome also codes for different regulatory proteins like Tat (transactivator protein), Rev (RNA splicing-regulator), Nef (negative regulating factor), Vif (viral infectivity factor), Vpr (virus protein r) and

Vpu (virus protein unique). The Tat and Rev are necessary for the starting of the replication while Nif, Vif, Vpr, and Vpu have a role in budding and pathogenesis also.

The structure of HIV is shown in Fig.1. The diploid ssRNA complexed with viral proteins p6 and p7 forms a stable nucleocapsid which is surrounding the RNA to prevent the genome being digest by the host cell

nucleases. Along with the RNA, viral enzymes like transcriptase and integrase are also enclosed safely inside the capsid. The second level of protection is by the capsid composed of the viral p24 protein. The sphere of viral encoded(p17) matrix protein is held within the capsid, and HIV Protease is placed in between the capsid and the matrix for its proper functioning in the maturation of the virus. The outer layer of the virion is a lipid bilayer originating from the plasma membrane of the host cell. The outer layer is completely covered with trimmers of the viral glycoprotein gp120 and gp41 these will receive the response to the external stimulus. The virus will bind to those cells which have the receptors complementary to gp120/gp41.

The different stages of HIV life cycle are the binding, fusion, reverse transcription. Integration, Replication, Assembly and Budding [5] (fig.2). Interrupting any of these stages will lead to the end of HIV progression. The detailed study of HIV life cycle leads to the discovery of different drugs that will interrupt different stages of HIV life cycle. The newly entered viral particle or the budded out matured viral particle expressing the gp120, and gp41 glycoproteins will recognize the complementary receptor pockets in CD4, which is present in the surface of the primary T-helper cells monocytes, macrophages and dendritic cells all belongs to the immunological cells. The HIV can infect only these cells which are having this receptor cells. The CD4 receptor will undergo an irreversible conformational change after binding with the gp120. This conformational change helps them to bind with the chemokine coreceptors CCR5 or CXCR4 [6]. This leads to the gp41-mediated fusion of the virus envelope to the plasma membrane of the cells, this whole process is known as fusion. When the fusion is completed, the reverse transcriptase enzymes start to convert the RNA genome into a linear dsDNA intermediate through the process known as reverse transcription. The completion of reverse transcription follows the cleaving of the 3' end of the DNA genome by the enzyme integrase. This leaves the hydroxyl end to expose to the surrounding. This nucleoprotein complex (linear DNA complexed with other viral proteins) will actively transport through the nuclear membrane and comes in contact with host DNA in the nucleus. Then the 3' hydroxyl group will attack the phosphodiester bond in the target DNA and forms a new bond between the host and viral DNA. This process is catalyzed by the enzyme integrase, and the process is known as the integration. The integrated host genome is known as the

provirus. The HIV uses the cellular transcription factors and to transcribe the proviruses. In the first step, it will transcribe a small amount of full-length viral RNA. This RNA will slice up to form mRNA for Tat and Rev. Both of this have a crucial role in the gene regulation of the HIV. The Tat plays the role of a gene-specific elongation factor by binding at the stable loop region of the mRNA known as Translating Responsive element (TAR). The transcription elongation rate will increases up to 30,000 times by the binding of Tat [7]. The shuttling protein Rev will export the unspliced HIV transcripts into the cytoplasm, the process is known as replication. The different transcripts are then translated to different cell organelles depending on the product. The Env is synthesized in the endoplasmic reticulum, while the Gag and Gag-Pol polyproteins are synthesized on cytoplasmic ribosomes. Followed by the production the gag and pol will associate with viral RNA molecule and transported to the cytoplasm where the Env is also present. This assembly of the viral particle is known as the assembly. The budding of the immature virion will take place in the plasma membrane. During the budding or just after it the protease cleaves the Gag and Pol polyproteins to form the mature virion particles.

The HIV uses reverse transcriptase to make the double-stranded DNA intermediate from its single-stranded RNA genome. However, this transcriptase doesn't have any 3'-5' exonuclease activity. Therefore, they can't distinguish between several of the different amino acid bases [8]. This leads to the production of different variants of viral particle compare to the one that infects the patients. The fidelity of the HIV reverse transcriptase is very poor; this makes an error rate of 1 in 2000-7000 nucleotides [9]. This rate compared with the large size of HIV virion genome leads to an average of 3 mutations in newly synthesized dsDNA intermediary. This notorious behavior of reverse transcriptase leads to a highly heterogeneous mixture of viral particles in the patient's body. This is the main reason for the drug resistance in HIV.

1.4 HIV Pathogenesis

CD4+ cells are the primary target of the HIV infection. This includes TH lymphocytes, macrophage, and dendritic cells. From these cells, the infection will start spread to the lymph nodes and spleens from where it is infecting the activated TH lymphocytes. The condition in these lymphoid is very suitable for the replication of the virus, and the replication rate will increase rapidly. This leads to the production of about ten offsprings in each day. There is a transient consumption of fringe CD4+ and related high blood-plasma heap o HIV during this intense stage.. Accordingly, an 'acute phase response' is propelled by the host's

immune system, where CD8 + cytotoxic T lymphocytes (CTL) particular to HIV antigens are clonally-expanded, also, killing antibodies are created. This humoral reaction brings about circling viral burdens dropping to imperceptible levels, however the tainted T H lymphocytes contain the HIV provirus coordinated into their genome thus go about as a repository, delivering virions. In the long run, the TH lymphocytes are crushed by the infection or by CTLs that remember them as contaminated. Gradually the levels of CD4 + cells are drained, which seriously immuno-bargains the host. Once the levels of CD4 + cells dip under a limit level, side effects due to opportunistic infections, for example, Kaposi's Sarcoma Herpesvirus are displayed in the host. This is alluded to as the beginning of AIDS, and quickly leads to death.

1.5 HIV-1 Antiretroviral Drug Therapy

The most critical progress in the therapeutic administration of HIV-1 contamination has been the treatment of patients with antiviral medications, which can smother HIV-1 replication to imperceptible levels. The disclosure of HIV-1 as the causative operator of AIDS together with the proper understanding of HIV life cycle leads to the progress. To date, there are 24 Food and Drug Administration (FDA)- approved drugs accessible for treatment of HIV-1 diseases. These medications are conveyed into six particular classes in view of their molecular mechanism and resistance profiles: (1) nucleoside-analogue reverse transcriptase inhibitors (NNRTIs), (2) non– nucleoside reverse transcriptase inhibitors (NNRTIs), (3) integrase inhibitors, (4) protease inhibitors (PIs), (5) fusion inhibitors, and (6) coreceptor antagonists [10].

The HIV-1 life cycle presents numerous potential open doors for therapeutic intervention; just a couple have been exploited. The initial phase in the HIV-1 replication cycle, viral passage, is the target for several classes of antiretroviral operators: attachment inhibitors, chemokine receptor antagonists, and fusion inhibitors. The HIV-1 envelope gp120/gp41 has an affinity for the CD4 receptor and guides HIV-1 to CD4+ cells [11]. Gp120 and CD4 are targets for little particle and antibody-based inhibitors BMS-378806 and TNX-355, every one of which has demonstrated some clinical guarantee, albeit nor is affirmed for use in HIV-1 patients [12,13]. BMS-378806 binds to a pocket on gp120 essential for restricting CD4 and modifies the adaptation of the envelope protein with the end goal that it can't perceive CD4 [14]. TNX-355 hinders HIV-1 envelope docking, however, does not repress CD4 work in immunological settings [12].

Reverse transcriptase (RT) was one of the main HIV-1 compound to be exploited for antiretroviral drug discovery. RT is a multifunctional protein, which is required to change over the single-stranded HIV-1 viral RNA into twofold stranded DNA. RT is the target for two particular classes of antiretroviral operators: the NRTIs, which are analogs of native nucleoside substrates, and the NNRTIs, which binds to a noncatalytic allosteric pocket on the protein. Among the FDA approved drugs 12 are targeting RT.

Integrase enzymes coordinates three processes helping for the virus DNA to integrate with the host DNA. it include endonucleolytic processing of the 3' end of the viral DNA strand transfer and joining with the host DNA. In an HIV-1 all theses processes happens in a stepwise manner, in which strand transfer is the rate limiting process. As the strand transfer takes lot time most of the newly designed Integrase Inhibitors (INI) are targeting this process.

The integration is followed by transcription in HIV-1. The transcript elongation requires the binding of HIV-1 regulatory protein Tat to the HIV-1 RNA element (TAR) [15]. This process is unique in the case of HIV-1 and believes to be one of the main target area. Many drugs are designed for inhibiting this process. However, none of them are sufficiently potent [16,17].

In the contest of HIV life cycle the last class of approved drugs are the HIV protease inhibitor (PIs). In HIV protease are used to cleave the immature viral particle to the mature one [18]. Once this process is inhibited only immature viral particle will be present and they won't be infectious. PIs are among the most potent agents developed to date. Till present there are nine FDA approved PIs. These drugs are These drugs are saquinavir (SQV), ritonavir (RTV), indinavir (IDV), nelfinavir (NFV), amprenavir (APV), lopinavir (LPV), atazanavir (AZV), tipranavir (TPV), and darunavir (DRV). However, the effectiveness of these inhibitors are limited by the occurrence of drug resistant mutations in the target enzyme caused by the high replication rate of HIV-1 and lack of a proofreading mechanism in its reverse transcriptase (RT). In this project we are studying about TMC-126, which is also a PI which is presently in the clinical stage.

1.6 HIV 1 Protease

This PR cleaves gag and pol nonfunctional polypeptides into functional protein particle which is a crucial step in the maturation of infectious HIV particles [19]. Inhibiting this enzyme through any drug can block the production of mature virus particles and further the infection of HIV.

The structure of HIV 1 PR (PR1) and HIV 2 PR (PR2) are more or less similar. There have about 50% similar sequences in PR1 and PR2 [20,21]. PR1 and PR2 have different specificity for there substrate and inhibitors[22]. PR1 is a homodimer made of two monomers; each consists of 99 amino acids. The binding cavity of PR comprises of active triads (Asp-Thr-Gly) one from each monomer and they located symmetrically [23] in the loop region. D25(D25') residue in the active site is highly conservative and known to be the active residue; they seem to locate in the same plane and make strong interaction with the substrate. This active site

region is lined by hydrophobic residues which form close van der Waals interaction with substrate or inhibitors [24]. An extended β sheet from each monomer (glycine-rich loop-residues from 46 to 56) gated the binding site, which opens and closes to allows the substrate or inhibitor to enter to the binding cavity [25,26]. Dynamics of flap region has a specific role in the binding of drugs; it is shown that the mutation is affecting the equilibrium between semi-open and closed conformation [27], which seems to be one aspect of drug resistance due to mutation. Also, flap dynamics are involved in the enzymatic mechanism [27]. HIV PR uses an activated water molecule to break substrate scissile bond. As with any other aspartic protease, HIV PR also uses D25/D25' for activating the water molecule. Still, the complete mechanism of the catalytic action of the PR is unknown although different mechanisms are proposed.

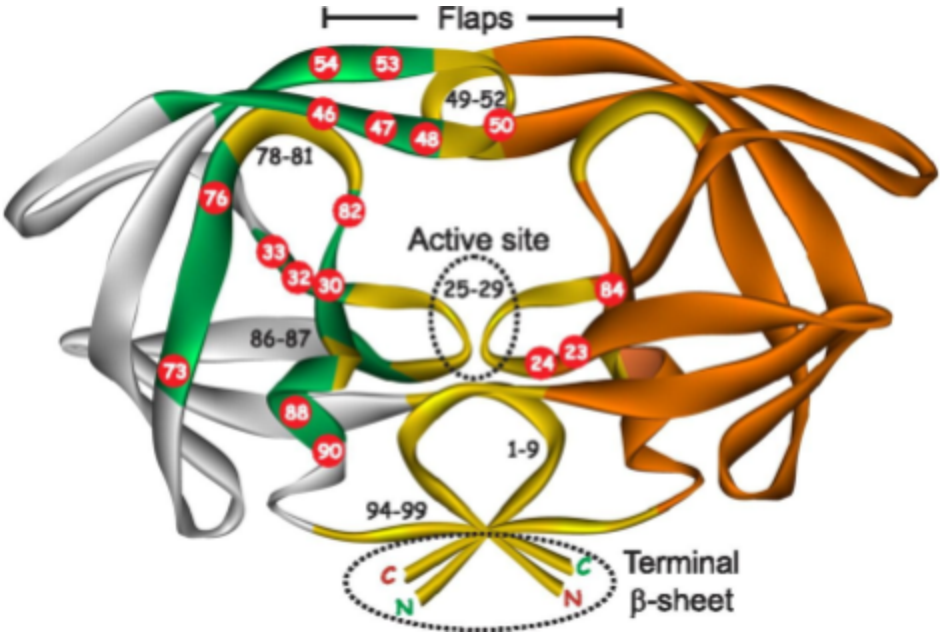


Fig.3. Structure of HIV 1 Protease, showing different regions in the protease [92]

The protonation state of D25/D25' is of particular importance as it is directly contacted with substrate or inhibitor. These Aspartic acids can be in an unprotonated, di-protonated or mono-protonated state [28]. Also depending on the position of hydrogen atom (on OD1 or OD2) mono protonation can be of four type [29]. We can't get any details of the protonation state from X-ray crystallographic structure or NMR, so we have to determine the protonation state through other mechanisms. Different suggestions are there for the factors depending on the protonation state of aspartic acids, which includes the proximity of the two aspartic acids, the environment of aspartic acid, etc. [30,31]-[32].

1.7 HIV 1 PR Drug resistance

The high rate of HIV infection, combined with the high transformation rate that happens amid each cycle of contamination, guarantees that patients have a mind boggling and different blend of viral quasispecies, each contrasting by at least one mutation. If any of these mutation can give a favorable contribution to the infection for example, if it decreases the affinity of some drugs without much affecting the substrate, this cause the resistance and the relating quasispecies will surpass the others,following a straightforward Darwinian principle of evolution. The rate of this procedure relies upon the level of the specific favorable position gave by the mutation, the commonness of the mutant inside the infection populace, and the level of medication at the site of HIV replication. Now and again, substitutions of single amino acids can create abnormal amounts of resistance. Since minority viral quasispecies conveying any single mutation is accepted to exist even before treatment, the rise of these single mutants can happen in a matter of weeks. For different drugs, just low-level resistance can be incited by single mutations. In these cases, higher drug resistance will be due to multi mutations.

In HIV 1 protease, mutation is shown by residues directly interacting with inhibitor, which is present in the binding cavity (primary mutation) or far away from the inhibitor binding site (secondary mutation) [33]. The mutations in the residues like 25−32,47−53, and 80−84 leads to primary mutation, as these residues are forming direct interaction with the substrate/inhibitors and all other mutations are secondary [34]. In many cases, the secondary mutations are compensatory mutations to reduce the dangerous effect of the primary mutation on the binding substrate to the protease [35]. In protease the regions like dimerization interface and flaps have their own on roles in the binding of drugs, so most of the secondary mutations are present in these regions[36]. These mutations decrease the affinity of drugs without affecting that much on the substrate by decreasing the nature and number of contacts with the residues. In some cases single mutation can cause a

high level of resistance towards inhibitors [37], among this, mutations in conservative hydrophobic residues (like V32I, I50V,etc.) are common, and these mutations are showing resistance towards many of the FDA approved drugs. In some agents, the single mutation will make an only low level of resistance, or gradual accumulation of additional mutation leads to multi-mutation in the protease which creates the total resistance. In this project we are focusing on different kind of mutations like V32I (binding site, primary mutation), M46L (flap region.secondary mutation), I50V (flap region, primary mutation) and MDR20 (multi-mutation).

Chapter 2

Molecular Mechanics

Molecular dynamics simulation is one of the most vital computational tool used to understand the structure and dynamical properties of proteins at atomic scale. There are wide varieties of applications for molecular dynamics simulation starting from the study of protein folding, protein-ligand binding, enzyme reactions, etc. Crucial to such simulations is the representation of energy of the protein as a function of its atomic coordinates. The state of low energy is considered as the most stable and densely populated one. The forces acting on each molecule will be treated as the gradient of this potential, which is the reason such potential functions are likewise regularly alluded to as "force fields."

For smaller chemical systems in the gas phase, it is possible to use quantum mechanical calculation to build up such force fields. However, in biological systems we have to treat macromolecules which are solvated also. This consists of atoms in the oder 10^5 where quantum mechanical calculations are not feasible. Empirical calculations can achieve this goal. In this simple equations will represent the interactions between atoms and by treating atomistic model for the calculations this method reduces the computational cost by several times. Most importantly, by using proper optimistic parameters in the mathematical modelling it can achieve the accuracy of the quantum mechanical model. In general the total potential energy of a chemical system $V(r)_{Total}$ can be separated into terms for the internal $V(r)_{int}$ and terms for external interactions $V(r)_{ext}$ as follows

$$V(r)_{Total} = V(r)_{int} + V(r)_{ext}$$

$$V(r)_{int} = \sum_{bond} k_b(b - b_0)^2 + \sum_{angle} k_\theta(\theta - \theta_0)^2 + \sum_{dihedral} k_\phi[1 + cos(n\phi + \square)] \qquad 1$$

$$V(r)_{ext} = \sum_{non\ bonded} [\frac{q_i q_j}{r_{ij}} + \frac{A_{ij}}{r_{ij}^{12}} - \frac{C_{ij}}{r_{ij}^{6}}]$$

From the Fig.4 the three summations are $V(r)_{int}$ are over bonds (1-2 interactions), angles (1-3 interactions), and torsions (1-4 interactions). The $V(r)_{ext}$ excludes 1-2 and 1-3 interactions and often uses separate parameters for 1-4 interactions as compared with those used for atoms separated by more than three covalent bonds. It describes electrostatics that use partial charges q_i on each atom that interact via Coulomb's law. The combination of dispersion and exchange repulsion forces are represented by a Lennard-Jones 6-12 potential; this is often called the ''van der Waals'' term. This equation is only a simplest equation which reproduce the atomistic potential energy. Currently there are different force fields which are using for the protein simulations, we are going to discuss three among them in detail. For our work we are using AMBER force field.

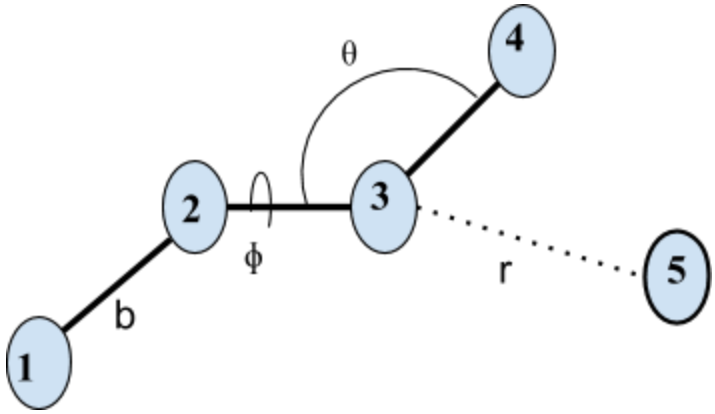


Fig.4. Schematic view of force field interactions. Covalent bonds are indicated by solid lines, nonbonded interactions by a light, dashed line

2.1 The AMBER force Field

AMBER is an acronym for Assisted Model Building with Energy Refinement. It is a family of force fields for molecular dynamics of the biomolecule. Different type of parameters is used for making the force fields. The parameters used to describe peptide, nucleic acid, and protein interactions are begins with"ff". GAFF

(General Amber Force Field) is used to describe small molecules like drugs and GLYCAM for simulating carbohydrates. It is written in Fortran 90 and C, also supporting most major Unix-like systems and compilers. Different programmes like antechamber, LEaP, MM-PBSA, SANDER, pmemd, etc. are used in Amber for running simulation and calculations.

In the 1980s enough knowledge had aggregated from the prior parametrization for a new generation of force fields. In the earlier efforts, they didn't take the hydrogen atoms explicitly for their calculation due to lack of computational power. The importance of hydrogen bond led many to take polar hydrogen bond explicitly. Still, they consider the hydrogen atoms bond to carbon as one entity. A force field at this level was created by Kollman group [38] and incorporated with Amber Molecular Mechanic package. In this one atomic charge are calculated at the Hartree-Fock STO-3G level, ESP (Electrostatic Potential) method. The Van der Waals interaction is treated with different data available at that time. The force constants are taken from the crystal structure, and their normal mode frequencies are matched with many peptide fragments. In this force field the k_{Φ} values are closely coupled to the nonbonded potentials used and are hardly transferable from one force field to another.

Different problems of only polar hydrogen atom models and improvements in the computational speed led many research group to think about an all-atom model. Initially, in this approach, they included the aromatic compounds using different optimization method, analogs problems in the ribose and deoxyribose in the nucleic acid gives momentum to the all-atom model idea. In 1986 Kollman and group published the work containing all-atom model [39]. In this and previous work, the parametrization is taken from the gas phase simulation. In the early 1990s, a new force field was introduced by the same group which is known as ff94 force fields [39]. Compare to the previous force fields; they made a complete effort to explicitly describe the algorithm by which the parameters were derived so that consistent extensions could be made to molecules other than proteins. The fulfillment of this effort comes true through the introduction of the antechamber, which completely automatizes the parametrization for amber force fields.

The recognition of the importance of solvent-solvent and solvent-solute interaction led another improvement in the force fields. Also, the determination of atomic charges that mimic the electrostatic potential outside the molecule helps for the making a better force field.
Earlier work had established that fitting charges to the potentials at the Hartree-Fock 6-31G * level tended to overestimate bond-dipoles (compared with observed gas phase values) by amounts comparable to that in empirical water models such as SPC/E or TIP3P; such ‘‘over-polarization’’ is an expected consequence of

electronic polarization in liquids. Hence, the use of fitted charges at the HF/6-31G * level appeared to offer a general procedure for quickly developing charges for all 20 amino acids in a way that would be roughly consistent with the water models that were expected to be used. Tests of this idea, with liquid-state simulations of amides and simple hydrocarbons, gave encouraging results. Once this charge and internal bonding parameters are determined the Lennard-Jones parameters can be established easily. One innovative idea emerges is on the case of hydrogen bonding by recognising the importance of electronegativity of the atoms they are bonded [40].

The final parameter to fit the force field is the soft torsional energy profile. Many research groups take completely empirical procedure for this fitting. Some account of the longer-range effects was provided in subsequent parameterizations, referred to as ff96 and ff99 [41] in which the and potentials were fit to tetrapeptide as well as dipeptide quantum mechanical conformational energies. The ff94 was extensively used for a long time. Then the limitations like over stabilization of α-helices were reported, and this leads to the new force fields. Improvements in the parameters of dihedral angles lead to the ff99SB [42]version of the force fields. This effort made improvements in the protein secondary structure stability and dynamics. However, the weaknesses in side chain rotamer and backbone secondary structure preferences have been identified. A complete refit to the dihedral angle of all amino acids, considering different protonation states of ionizable side chains leads to ff14SB [43] which is showing more accurate results compared to all previous versions of amber force fields

2.2 The CHARMM Force Fields

CHARMM (Chemistry at HARvard using Molecular Mechanics) is the most widely used force field for simulating molecules of biological interest. The CHARMM force fields for proteins include united-atom (sometimes termed extended atom) CHARMM19 all-atom CHARMM22 and its dihedral potential corrected variant CHARMM22/CMAP. In 2009 a general force field for drug-like molecules also introduced it covers a wide range of chemical groups present in biomolecules and drug-like molecules, including a large number of heterocyclic scaffolds.

Similar to Amber the CHARMM was also developed in the 1980s without any explicit hydrogen atoms. Later, in 1985 the old version improves to CHARMM19 by including hydrogen atoms bonded to nitrogen and oxygen are explicitly represented, while hydrogens bonded to carbon or sulfur are treated as part of an

extended atom. The parameterization in this model fitted to the quantum mechanical calculations at the HF/6-31G level

Unlike Amber, CHARMM19 force field did well in both gas phase simulation and solvated simulation. Also, the CHARMM19 values have often been used in conjunction with a distance-dependent dielectric constant as a rough continuum solvation model. Refining of parameters of the force field to get a good balance of interaction energy in solvated medium leads to CHARM22 [44]. The key approach from CHARMM19 was carried over by deriving charge models primarily from fits to solute-water dimer energetics (now calculated at the HF/6-31G * level).

As in Amber, Lennard-Jones parameters are refined to reproduce densities and heats of vaporization of liquids as well as unit cell parameters and heats of sublimation for crystals by making small variations in the parameters. The searching for the best-optimized parameters for the torsional energies leads to CHARMM27 [45]. Strictly speaking, CHARMM27 is the force field developed for nucleic acids and lipids. Improvements in the backbone scalar couplings across hydrogen bonds, residual dipolar couplings (RDCs) and relaxation order parameter, as well as scalar couplings, leads to CHARMM36 which can also be used for protein simulations.

2.3 The OPLS Force Fields

The third important force field that also developed during the same time is the OPLS (Optimized Potentials for Liquid Simulations). This potential was initially developed for simulating liquid properties. In its first stage, it is used for water and 40 other amino acids and placed a strong emphasis on deriving nonbonded interactions by comparison to liquid-state thermodynamics. These force fields reproduce the experimental values of densities and heat of vapor pressure with good approximation. These early models (now called OPLS-UA) treated hydrogens bonded to aliphatic carbons as part of an extended atom but represented all other hydrogens explicitly.

The first approach for the protein force field treated only polar type hydrogen atoms explicitly and taken the atomic parameters from the Amber 84 [46]. This was called the AMBER/OPLS force field, and for some time was reasonably popular. As with Amber and CHARMM, an all-atom version (OPLS-AA) was developed later, with much the same philosophy for the derivation of charges and van der Waals parameters

from simulations on pure liquids. Torsional parameters were developed consistently by fits to HF/6-31G * energy profiles, along with some recent modifications, especially for charged side chains. Bond stretching and angle bending terms were standardized and were largely taken from the 1986 Amber all-atom force field. The parameter choices were intended to be ‘‘functional group friendly,’’ so that they could be easily transferred to other molecules with similar chemical groupings.

2.4 Other Force Fields

There are several other force fields developed during the same period and later which have their importance in specific problems. GROMOS force field was developed along with the programming package of the same name. Different all-atom force fields like CEDAR and GROMACS also derived largely from the GROMOS itself. By aiming largely on the drug-like molecules, Halgren developed a new force field called MMFF (Merck Molecular Force Field). The DISCOVER force field has seen use primarily in conjunction with the commercial INSIGHT modeling package. Levitt's group developed a new force field named as ENCAD (ENergy Calculation And Dynamics) by incorporating different potential data they studied over the years. This has been ud=sed for studying the folding and unfolding of the proteins. This force field is unique compared to all other forcefields due to the usage of group-based approach compare to Atom-based approach in all other force fields. They also excluded the neighbor short-range electrostatic interaction and also used pairwise nonbonded potentials shifted to zero energy at short range, and specifically parameterized to reflect these small cutoff distances.

Chapter 3

Free Energy Simulations

3.1 Thermodynamics

Laws of physics are fundamental and every process happening in nature should obey these laws, biological organisms and systems are no different. So, in macroscopic level dynamics of the biological system should follow the two fundamental laws of thermodynamics (now we are treating zeroth law and third law as they are not significant in biological contest).

1. Conservation of total Energy

The first law of thermodynamics is just the conservation of total energy. It states that whatever happens or the in spite of all the exchange of energy during a biological process the total energy before and after the process will be same. If the system releases some energy, it will turn to some form of energy of the surrounding.

$$\Delta U = Q + W + \mu N \qquad 3.1$$

U is the total internal energy of the system; Q is the heat transfer and W is the work done.

2. Spontaneous means increased entropy

The Second law of thermodynamics states that A reaction can only occur spontaneously if it results in a net increase of entropy. Entropy is a measure of the disorder of a system; when a system becomes more disordered, its entropy increases. The change in entropy of a reaction where one mole of compound A is converted into one mole of compound B is given by:

$$\Delta S = K_B ln(\frac{P_B}{P_A}) \qquad 3.2$$

S in the entropy of the system; K_B is the boltzmann constant; P_B and P_A are the probability of occurrence of each state A and B. As can be seen, ΔS is directly proportional to the natural logarithm of P_B and P_A, so a positive change in entropy results from the probability of state B occurring being higher than the probability of state A occurring.

The entropy of a system can decrease during a spontaneous reaction (where the probability of the products occurring is lower than the probability of the reactants) as long as the subsequent increase in entropy of its surroundings results in a net increase. Therefore the formation of highly ordered structures seen in biological processes, which have a negative entropy, can occur spontaneously due to a larger increase in the entropy of its surroundings caused by the release of heat energy from the reaction. This is written mathematically as:

$$\sum\{\Delta S_{system} + \Delta S_{surrounding}\} > 0 \qquad 3.3$$

Of these two laws, the second law is of considerable importance in biochemistry, as it states whether a process will occur spontaneously. However, the problem with using entropy as an indicator of whether a biochemical reaction will occur spontaneously is that it is not easily measurable, especially as it requires knowledge of ΔS surroundings, which, in the case of a cellular reaction, is everything in the universe outside of the cell. For this reason, a composite thermodynamic function called Gibb's free energy was defined which combines equations from the first and second laws of thermodynamics without requiring consideration of any thermodynamic properties of the surroundings. Gibb's free energy of a system is defined as

$$G = H - TS \quad (3.4)$$

$$\Delta G = \Delta H - T\Delta S \quad (3.5)$$

G is the Gibb's free energy; H is the enthalpy of the system; T is the absolute temperature and s is the entropy of the system. We can determine the spontaneity of the process from the change in Gibb's free energy of the system. Interestingly no terms are corresponding to the surrounding in the given equation. This is because of Gibb's free energy itself include the effect of the surrounding.

$$-\Delta G = -\Delta H + T\Delta S \quad (3.6)$$

$$\frac{-\Delta G}{T} = \frac{-\Delta H}{T} + \Delta S_{system} \quad (3.7)$$

From the definition of Enthalpy of the system

$$H = U + PV \quad (3.8)$$

H is the enthalpy of the system; U is the internal energy; P is the pressure and V is the volume. Enthalpy is the thermal energy transferred from the system to the surrounding. If "h" is the thermal energy gained by the surrounding then,

$$\Delta H = -h \quad (3.9)$$

Using this idea in equation 3.7, the change in enthalpy is equal to the thermal energy transferred to the surrounding and these thermal energy will increase the number of possible configuration that the surrounding can take which leads to an increase in the entropy of the system. Then equation 3.7 can be rewrite as

$$\frac{-\Delta G}{T} = \Delta S_{surrounding} + \Delta S_{system} \quad (3.10)$$

That means measuring the Gibbs free energy of a system is the measure of total entropy change of the the universe. As mentioned earlier, change in Gibbs free energy can measure the the spontaneity of a process as

• A positive ΔG means that the reaction cannot occur spontaneously as it causes a net decrease in the universe's entropy, and therefore requires an input of free energy in order for the reaction to occur.

• A ΔG value of zero means that the system is at equilibrium as the reaction proceeds at equal rates in the forward and reverse directions.

• A negative ΔG means that the reaction causes a net positive entropy change in the universe and so the reaction occurs spontaneously.

A highly-negative ΔG simply indicates that the reaction results in a greater entropic disorder in the universe. However, it does not mean that it occurs at a faster rate. Also putting 3.1 and 3.6 together and assuming there is no wok other than PΔV

$$\Delta G = \mu \Delta N \qquad 3.11$$

Through Gibb's free energy we are measuring the change in the chemical potential of the system.

As free energy is giving the informations about the spontaneity of the system free energy is considered to be the most important parameter in studying the thermodynamic properties of a system. The Liouville's theorem which leads to the ergodic hypothesis made a logic background for using long time molecular dynamic simulation of an ensemble of the system for studying the complete phase of the system. In physics and thermodynamics, the ergodic hypothesis says that, over long periods of time, the time spent by a system in some region of the phase space of microstates with the same energy is proportional to the volume of this region, i.e., that all accessible microstates are equiprobable over a long period of time. After doing the simulation for a sufficiently long time, the Boltzmann distribution of the macrostate properties will provide different macroscopic values of the system. From statistical physics, the Helmholtz free energy of a system (A), which is equivalent to the Gibb's free energy for systems with constant N V T rather than N P T , can be written as

$$A = -\frac{1}{\beta} ln\{Z_{NVT}(q,p)\} \qquad 3.12$$

A is the Helmholtz free energy; $\beta = \frac{1}{K_b T}$; Z is the partition function (p and q as it is in phase space). From this equation we can moves to

$$A = -\frac{1}{\beta} ln < e^{H(p,q)} > \qquad 3.13$$

This equation states that the free energy of a system is equal to the time average of the exponential of its Hamiltonian. This means the contribution from higher energy state will be more. However, in molecular

dynamics simulation we are simulating the stable equilibrium states only which is the lowest energy state. Avoiding the higher energy states leads to a poor calculation of the free energy. This problem is solved by using different computational techniques for calculating the free energy based on the thermodynamic cycle of the process. The techniques range from computationally-demanding, yet highly accurate, methods such as Free Energy Perturbation (FEP) [47] and Thermodynamic Integration (TI) [48] through to more heuristic techniques such as Molecular Mechanics Poisson-Boltzmann Solvation Area (MMPBSA) [49]-[50].

In the computationally expensive methods like FEP and TI they use thermodynamic cycle perturbation approach. For example, consider we have to calculate the difference in the binding of two ligands to a particular receptor.

$$R + L_1 \rightarrow R.L_1 \quad (\Delta G_1) \tag{3.14}$$

$$R + L_2 \rightarrow R.L_2 \quad (\Delta G_2) \tag{3.15}$$

In the direct method, we have to do two simulations by bringing the ligands from far away to the final configuration by treating all the solvated molecules. However, by using the thermodynamic cycle approach we can split this equation as Fig.5. As the free energy of a system is a state function the the total change in free energy in a thermodynamic cycle will be zero. Therefore

$$\Delta G_2 - \Delta G_1 = \Delta G_4 - \Delta G_3 \tag{3.16}$$

Computationally we can calculate both ΔG_4 and ΔG_3 from which we can get the difference in the binding energy of both ligands with receptor.

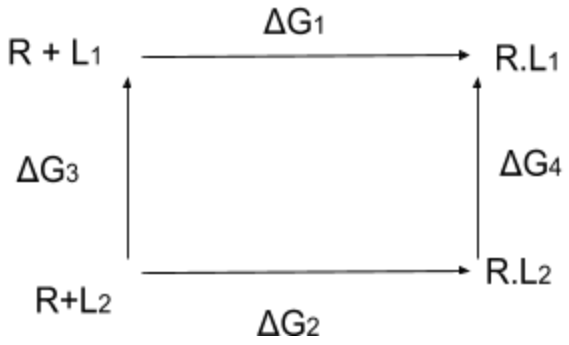


Fig.5: Conversion of Equations 3.14 and 3.15 into a thermodynamic cycle

However, it can also be adapted to calculate the absolute difference in free energy of binding in solution

(Fig.6).

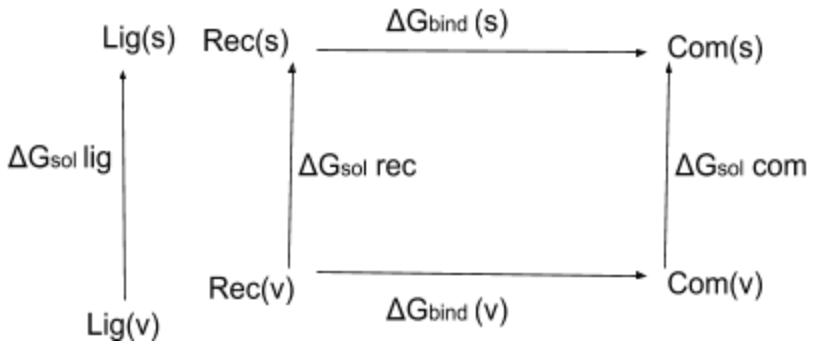


Fig.6. Thermodynamic cycle to calculate the change in free energy upon ligand binding (ΔG bind (s)).

Here, (s) denotes 'solvated'; (v) denotes 'in vacuum'; ΔG_{sol} denotes change in the free energy upon solvation for ligand (lig), receptor (rec), and complex (com). The principles underlying thermodynamic cycles allow the $\Delta G_{bind}(s)$, which cannot be calculated through standard molecular dynamics, to be indirectly determined through Equation 3.17

$$\Delta G_{bind}(s) = \Delta G_{bind}(v) + [\Delta G_{sol}^{com} - (\Delta G_{sol}^{lig} + \Delta G_{sol}^{rec})] \qquad 3.17$$

The methods like MM-PBSA used to calculate all the four binding energies in 3.17.

3.2 MM PBSA

As mentioned MMPBSA is using equation 3.17 for calculating the binding free energy of the solvated complex. In 3.17 ΔG bind (v)is the free energy of binding of the ligand with the receptor in the vacuum. It is calculated as the difference in the free energy of the system after and before binding as shown

$$\Delta G_{bind}(v) = \Delta G_{com}(v) - [\Delta G_{rec}(v) + \Delta G_{lig}(v)] \qquad 3.18$$

The free energy of ligand/receptor or complex involves all the interaction which hold them together in the specific stable state. These interactions are the electrostatic (ele), Van der Waals (VdW) and other internal interactions mainly due to the stress and strain due to the alignment of the bonds.

$$G_x(v) \sim U_x(v) = U_{ele} + U_{VdW} + U_{int} \qquad 3.19$$

Here x refers for ligand, receptor and complex. Then

$$\Delta G_{bind}(v) \simeq \Delta U_{bind} = \Delta U_{ele} + \Delta U_{VdW} + \Delta U_{int} \qquad 3.20$$

Once we got the vacuum binding energy of the system, the next term in equation 3.17 is the solvation free energy of the system. This part of the binding free energy consists of two parts one is the non-polar contribution and other is the polar contribution.

$$\Delta G_{solv} = \Delta G_{pol} + \Delta G_{np} \qquad 3.21$$

The non-polar contribution can be divided into two parts as the contribution from the Van der Waals interaction of the atoms and the other is the free energy change of forming a cavity in the solvent in which the molecule is placed. Due to the short-range distances over which the van der Waals forces act, and the major component of the free energy change associated with cavity formation arising in the first layer of solvent molecules, the non-polar contribution can be calculated from the solvent-accessible surface area (SASA) of the molecule:

$$\Delta G_{np}^{sol} = \Delta G^{VdW} + \Delta G^{cav}$$

$$= \gamma(SASA) + b \qquad 3.22$$

where γ and b are empirically-derived constants that reflect the surface-tension of the solvent and an off-set value respectively.

When the receptor is solvated in some medium the molecules of the solvent will interact with the molecules of the receptor. If water is its molecules will interact electrostatically to the hydrophilic residues in the receptor, and the polar part of the solvation energy account for the energy utilized by the ligand molecule to displace these water molecules. Compare to the non-polar contribution this is a bit complicated to calculate the polar contribution. Solvation interactions, especially solvent-mediated dielectric screening and Debye-Hückel screening, are essential determinants of the structure and function of proteins and nucleic acids [51]. Ideally, one would like to provide a detailed description of solvation through explicit simulation of a large number of solvent molecules and ions. This approach is frequently used in molecular dynamics simulations of solution systems. In many applications, however, the solute is the focus of interest, and the detailed properties of the solvent are not of central importance. In such cases, a simplified representation of solvation, based on an approximation of the mean-force potential for the solvation interactions, can be employed to accelerate the computation. Among the different implicit solvent methods Poisson-Boltzmann

(PB) method and the generalised-Born (GB) approach is of particular importance.

The Poisson-Boltzmann (PB) method is a most widely used implicit water model for calculating the solvent-solute electrostatic interaction. [51].This method reproduces the results of many explicit solvent models and also agrees with experimental results. In this method, a solvent is treated as a continuous medium and the solutes as point particle forming an electrostatic field around. Then assuming the solute is forming different layers on the solvent the potential between these layers will be treated by Poisson's equation and the distribution of the charged layers will be treated with Boltzmann distribution leads to Poisson-Boltzmann equation. Considering other screening effects and other parameters the total potential can be calculated using the given equation [52].

$$\nabla.[\epsilon(r)\nabla\Phi(r)] = -4\rho(r) - 4\lambda(r)\sum_i z_i c_i exp[\frac{z_i\Phi(r)}{K_B T}] \qquad 3.23$$

where ε(r) is the dielectric constant, φ (r) is the electrostatic potential, ρ(r) is the solute charge, λ (r) is the Stern layer masking function, z_i is the charge of ion type i, c_i is the bulk number density of ion type i far from the solute, k_B is the Boltzmann constant, and T is the temperature; the summation is over all different ion types. For a system we can solve this equation by using different numerical methods.

However, in molecular dynamics applications, the computational cost associated with the PB analysis is very high as we have to solve the PB equation whenever there is a conformational change in the molecules. Amber developers have pursued an alternative approach, the analytic generalized Born (GB) method, to obtain a reasonable, computationally efficient estimate to be used in molecular dynamics simulations. The methodology has become popular, especially in molecular dynamics applications [53,54], due to its relative simplicity and computational efficiency, compared to the more standard numerical solution of the Poisson-Boltzmann equation. Within Amber GB models, each atom in a molecule is represented as a sphere of radius R_i with a charge q_i at its center; the interior of the atom is assumed to be filled uniformly with a material of dielectric constant 1. The molecule is surrounded by a solvent of a high dielectric ε (80 for water at 300 K). The GB model approximates ΔG_{ele} by an analytical formula [55].

$$\Delta G_{ele} \simeq -\frac{1}{2}\sum_{i,j}\frac{q_i q_j}{f_{GB}(r_{ij},R_i,R_j)}(1 - \frac{exp(-kf_{GB})}{\varepsilon}) \qquad 3.24$$

where r_{ij} is the distance between atoms i and j , the R_i are the so-called effective Born radii, and f_{GB} is a certain smooth function of its arguments. The electrostatic screening effects of (monovalent) salt are

incorporated via the Debye-Huckel screening parameter κ. The effective Born radius of an atom reflects the degree of its burial inside the molecule: for an isolated ion, it is equal to its van der Waals (VDW) radius ρ_i . Then one obtains the particularly simple form:

$$\Delta G_{ele} \simeq -\frac{q_i^2}{2\rho_i}(1-\frac{1}{\varepsilon}) \qquad 3.25$$

If the GB method is implemented with the SASA method of calculating the non-polar solvation free energy, and the molecular mechanics method of calculating the internal energies of the components, the methodology is called MMGBSA. If, instead of the GB method, the PB method is implemented to calculate the electrostatic component of the solvation free energy, the methodology is called MMPBSA.

While the MMPBSA and MMGBSA methodologies include the change in configurational entropy of the solvent with the ΔG_{sol}^{cav} term of the non-polar contribution to the solvation free energy, they do not include the change in free energy due to the decrease in configurational entropy of the ligand and receptor upon complex formation. When a ligand binds to a receptor, both components undergo a decrease in configurational entropy as the non-bonded forces between the two restrict their degrees of freedom.This change in configurational entropy upon complex formation can be subdivided into three components [56]:

$$\Delta S_{conf} = \Delta S_{tra} + \Delta S_{rot} + \Delta S_{vib} \qquad 3.26$$

$$\Delta S_{tra} = -Rln(\delta x_1 \delta x_2 \delta x_3/1660)$$

$$\simeq -n_{tra}Rln(\delta x/11.8)$$

$$\Delta S_{rot} = -Rln(\delta\alpha_1 \delta\alpha_2 \delta\alpha_3)/8\pi^2 \qquad 3.27$$

$$\simeq -n_{rot}Rln(\delta\alpha/4.3)$$

$$\Delta S_{vib} = R[\frac{x}{e^{(x-1)}} - ln(1-e^{-x})$$

Where $x = \frac{hc\omega}{KT}$ and ω is the frequency of the normal mode; $\delta x_1, \delta x_2, \delta x_3$ are the r.m.s. amplitudes, in Angstrom units, of the movements in principal directions and δx is their mean. Similarly α is the angle.

3.3 Free Energy Perturbation Method (FEP)

FEP is another technique using in molecular dynamics or Monte Carlo method for calculating the free energy difference of the processes. This method is using equations 3.14 /3.15 for calculating the free energy difference between the two systems. In this method, one ligand is gradually perturbed into another using some couplings. In MD we are not sampling the higher energy space, which causes the low convergence of the value of free energy (equation 3.13). In FEP we are solving this problem by doing a little trick in the equation for the free energy which leads to Zwanzig equation (Equation 3.28). The Helmholtz free energy change corresponds to the change of ligand A to ligand B is given by

$$A_A - A_B = -\frac{1}{\beta} < e^{-\beta(E_A - E_B)} >_A \qquad 3.28$$

The triangular brackets denote an average over a simulation run for state A. Compare to equation 3.13 equation 3.28 is compare likely to compare due to the negative on the exponential. Also now the sampling is over coordinates for A, which determines the exponential of energy difference between A and B. The new problem arising is what happens if the sampling over A is not ergodic for B. For example, consider the case of ethanol in water. Ethanol will have a water molecule bonded to its lone pair. To swap the trans to the gauche orientation of the hydroxyl group on ethanol presumably we will whack the hydrogen of the hydroxyl group of the ethanol close to the water. This won't be an ergodic distribution for the water. For solving this problem in FEP, we were couple one molecule (A) to another molecule (B) using equations like 3.29 and gradually transform one to other by changing the coupling constant. This will ensure that the sampling space is ergodic for both the molecules.

$$E(\lambda) = \lambda E_B + (1 - \lambda) E_A \qquad 3.29$$

Where λ is the coupling constant. By considering the coupling and changing with small fraction is each stage of simulation we can rewrite the equation 3.28 as

$$< A >_B - < A >_A = -\frac{1}{\beta} \sum_{\lambda=0}^{1} ln < e^{-\beta(E_{\lambda+d\lambda} - E_\lambda)} >_\lambda \qquad 3.30$$

In this we are changing λ in each step and calculate the free energy change and will sum up over all the different value of the λ to get the net free energy change due to changing the molecule from one to another. The predictions of FEP is in good agreement with the experimental values. However, computationally the

method is costly.

3.4 Thermodynamic Integration Method (TI)

Thermodynamic integration method is a similar technique similar to FEP used to calculate the free energy difference between two states A and B whose potential has different dependence on the coordinates. This method is formally identical to the FEP what we are doing here is we are calculating the derivative of the potential energy with respect to L and calculate an ensemble average with respect L at a specific value of L and integrate over all value of L leads to the free energy change (equation 3.31) [57].

$$\Delta A = \int d\lambda < \frac{dE}{d\lambda} >_{\lambda} \qquad 3.31$$

λ is the coupling constant as defined in the case of FEP. We can easily reach equation 3.31 by considering equation 3.28 and 3.29.

Chapter 4

Molecular Dynamic Simulation

An MD simulation is a technique to produce a dynamical trajectory for a system composed of N particles by integrating Newton's equations of motion. We need a set of initial conditions (positions and velocities of each particle), a good model to represent the forces acting between the particles (either from electronic structure calculations or using the empirical force fields presented in the previous section), and to define the boundary conditions to be employed. Then we need to solve the classical equation of motion:

$$m_i \frac{d^2 r_i}{dt^2} = f_i = - \frac{\partial U(r_1, r_2 ... r_N)}{\partial r_i} \qquad 4.1$$

Where $U(r_1, r_2 ... r_N)$ is the potential energy depending on the coordinates of the N particles. This is a system of N coupled second-order nonlinear differential equations that cannot be solved exactly, so equation 4.1 has to be solved numerically step by step using an appropriate integration algorithm.

4.1 Initial Conditions

To start the MD, we need to the initial position and velocity of the system. For a crystallographic structure, we will get the initial position from the crystallographic file (X-crystallographic file in the PDB format is available). If the crystallographic structure is not available, we have to make the initial coordinated by using different software. In this project, we use PBD files PDB id 2i4u for HIV 1 and PDB id 3s45 for HIV 2 for the initial position. The velocity of each particle is attributed randomly from a Maxwellian distribution centered on the desired temperature, and then they are adjusted to zero the angular momentum and the center of mass velocity of the total system.

4.2 Boundary Conditions

In a realistic model of the solution, a large number of solvent should be involved along with the solute. However, simply putting the solute in a solvent won't work properly.We could use rigid walls, then the surface effects would blur the real bulk physics. This is because the fraction of atoms near the walls is proportional to $N^{-1/3}$. In this model, a large number of solvent molecules will interact with the solute, and some will be

in the bulk of the solution, and many should be in the boundary of solvent and vacuum. This is not a proper model, to avoid the boiling off of solvent molecules and to reduce the number of solvents to reproduce the properties of the bulk we have to use the Periodic boundary condition(PBC) in our simulation. In this method, the particles being simulated are enclosed in a box which is then replicated in all three dimensions to give a

periodic array (Fig.7). In this periodic array, a particle at position r represents an infinite set of particles at:

$$r + ax + by + cz \; ; \quad -\infty < a, b, c < \infty \qquad 4.2$$

a,b,c corresponds to the length of the three-dimensional box. During the simulation, only one of this particle is represented, and the effect of all particle is reproduced using image particles. The particle in one box can interact with other particles of the same box as well as with the image particles also. Also, the particle that is

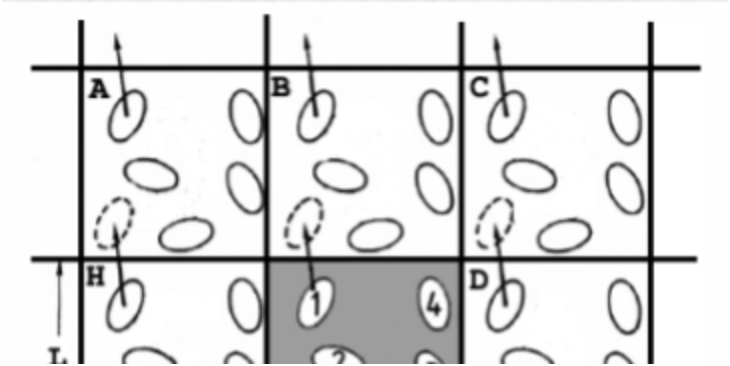


Fig.7 A two-dimensional array of boxes. As molecule 1 moves from the central box into box B it is replaced by it's image which moves from box F into the central box. This movement is replicated across all the boxes

leaving through one end of the box will re-enter through the other. This ensures the dynamical equilibrium and the consistency of the total number of particles throughout the simulation.

Using PBC in the case of a crystal structure is reasonable and will reproduce the real system also. However, in the case of disordered systems, we should be much careful in using the PBC because we are introducing an artificial periodicity. If the size of the simulation box is sufficiently large, the effects are normally not important. However, we need to be careful when considering any property that depends on long-range correlations as the only allowed fluctuations are those having wavelengths that are compatible with the box size. It will be problematic to simulate any fluctuation beyond the size of the box length. In the case of macromolecules like DNA, it is recommended to use a minimum layer of 10 Å of water [58].

Considering interaction with nearest images along with PBC is not a nice way to treat the problem, as the time for calculating the interaction force is proportional to the N, and the potential along the surface of a cube is not a constant. In order to solve the problem, it is better to consider the interaction of particle within a sphere of radius R_c. R_c should be less than L/2 because otherwise, some interactions between a particle and more than two images would occur and it has to be large enough so that the potential can be safely neglected beyond R_c. This is called spherical truncation. It must also be noted that this truncation should be applied only to short-ranged interactions like Lennard-Jones interaction. Long-range forces (typically electrostatic interactions) should never be truncated. For treating long-range interactions, we have to use other methods.

4.3 Long Range Interactions

Long-range interactions play a major role in molecular dynamics (MD) simulations. Also, periodic boundary conditions (PBC) is used in most simulations to approach bulk systems, which further complicates the calculation. Let's look at the direct sum method first. If we are calculating the total electrostatic interaction directly for N atoms/ions in the vacuum, at locations $r_1, r_2, ..., r_N$, and possessing point charges $q_1, q_2 \ldots, q_N$, respectively. The total Coulomb interaction energy is given by the following expression:

$$U = \frac{1}{4\pi\varepsilon_0} \sum_{i,j} \frac{q_i q_j}{|r_{ij}|} \qquad 4.3$$

However, in our case system is satisfying PBD and from 4.2 we can re-write 4.3 as

$$U = \frac{1}{8\pi\varepsilon_0} \sum_{n} \sum_{i=1}^{N} \sum_{j=1, j\neq i, if\ n=0}^{N} \frac{q_i q_j}{|r_{ij}+nL|} \qquad 4.4$$

This summation converges very slowly, and the series is infinite one as we are summing over n. This equation only considers single cell, and it is crazy to use this equation if the number of image cells is large, as it is almost impossible for the single cell itself.

A simple alternative is to use the reaction field method, consisting in considering all the molecules beyond a certain cutoff as a dielectric continuum. For a given atom, all the interactions with the particles inside the cutoff sphere are explicitly considered, and the charge distribution inside that sphere polarizes the dielectric continuum, so this one produces an additional electrical field into the cavity. This reaction field is given by the equation:

$$E_{rf} = \frac{1}{4\pi\varepsilon_0} \frac{2(\varepsilon_{rf}-1)}{2\varepsilon_{rf}+1} \frac{1}{R_c^3} \sum_{j\epsilon R} I_j \qquad 4.5$$

Where ε_0 is the dielectric constant of the vacuum, ε_{rf} is the dielectric constant of the system, R_c is the cutoff radius which normally coincides with the cutoff employed for the Lennard-Jones interactions, and I_j is the dipole moment of molecule j. However, compare to this method Ewald Sum method or Particle Mesh Ewald (PME) is a much accurate method of calculating the long-range interactions.

Ewald sum is a faster method to compute energies or forces. In Ewald summation method we divide electrostatic interactions into short-range and long-range components. The short-range component quickly decays to zero in real space, thus it can be truncated using a switching function, as was the case for the Lennard-Jones potential. The long-range component is more difficult to deal with, though it can be calculated efficiently using some tricks. The charge on the system can be written as the sum of delta Dirac function as shown in equation 4.6.

$$\rho_i(r) = \rho_i^s(r) + \rho_i^L(r)$$

$$\rho_i^s(r) = q_i\delta(r - r_i) - q_i G(r - r_i) \qquad 4.6$$

$$\rho_i^L(r) = q_i G(r - r_i)$$

L for long range; s for short range

Once we solve this charge distribution for potential using Poisson's equations, we can see that the contribution from the long-range terms are not converging in the real space and then we will treat them in the Fourier space. The Fourier terms can be represented as

$$\frac{1}{r} = \frac{f(r)}{r} - \frac{1-f(r)}{r} \qquad 4.7$$

Different functions are used for f(r). In the case of Ewald sum, we have to specify the cut off range for both real and reciprocal space/Fourier space. There isn't a straightforward way to obtain an optimal set of values, which often leads to errors. Another inconvenience is that it is computationally expensive. An optimized Ewald sum scales with the number of atoms as $O(N^{3/2})$, while for large systems the computation of the short-range potential can scale as O(N). There are several variants of the Ewald sums allowing scalings of O(N log N). Most import among them is the Particle Mesh Ewald Sum method in which we are using Fast Fourier Transform algorithm to calculate the reciprocal sum.

4.4 Integration algorithms

We already said that equation 4.1 could only be solved numerically. So we need to discretize the trajectory and use an integrator to advance over small time steps:

$$r_i(t_0) \rightarrow r_i(t_0 + \Delta t) \rightarrow r_i(t_0 + 2\Delta t) \rightarrow \ldots r_i(t_0 + n\Delta t) \qquad 4.8$$

As this is the most time-consuming step in the simulation, the algorithms which use more than one cycle of evaluation force per cycle is not efficient. Also, it should give accurate and stable result in consistent with the conservation of momentum and energy of the system. Also, the phase space volume of the system should be conserved and the algorithm satisfying this condition is known as symplectic algorithms. The immediately apparent solution would be to use a simple Taylor expansion, so:

$$r_i(t_0 + \Delta t) = r_i(t_0) + \frac{dr_i(t_0)}{dt}\Delta t + \frac{1}{2}\frac{d^2 r_i(t_0)}{dt^2}\Delta t^2 + O(\Delta t^3) \qquad 4.9$$

However, this algorithm is unstable and inaccurate. A better solution was proposed by Verlet. If we sum the Taylor expansions for $+\Delta t$ and $-\Delta t$, the terms in Δt, Δt^3 , etc. cancel and we obtain:

$$r_i(t_0 + \Delta t) = -r_i(t_0 - \Delta t) + 2r_i(t_0) + a_i(t_0)\Delta t^2 + O(\Delta t^4) \quad 4.10$$

The velocities are not used in the algorithm, but they can be obtained as:

$$v_i(t_0) = \frac{1}{2\Delta t}[r_i(t_0 + \Delta t) - r_i(t_0 - \Delta t) \quad 4.11$$

The intrinsic error of the algorithm or local truncation error is therefore of $O(\Delta t^4)$. Two equivalent algorithms producing the same trajectory as the Verlet integrator are the leap-frog algorithm:

$$r_i(t_0 + \Delta t) = r_i(t_0) + v_i(t_0 + \frac{\Delta t}{2})\Delta t \quad 4.12$$

$$v_i(t_0 + \frac{\Delta t}{2}) = v_i(t_0 - \frac{\Delta t}{2}) + a_i(t_0)\Delta t \quad 4.13$$

And the velocity verlet algorithm;

$$r_i(t_0 + \Delta t) = r_i(t_0) + v_i(t_0)\Delta t + \frac{1}{2}a_i(t_0)\Delta t^2 \quad 4.14$$

$$v_i(t_0 + \Delta t) = v_i(t_0) + \frac{1}{2}[a_i(t_0) + a_i(t_0 + \Delta t)]\Delta t \quad 4.15$$

Those three algorithms are very simple, but nevertheless they are efficient, stable and reasonably

accurate. Furthermore they are also time-reversible and symplectic, so they constitute a good choice as integrators for an MD simulation.

4.5 Constraints to the System

Throughout the simulation different constraints are used which help to make the ensemble of interest and also help to study desirable properties of the system.

4.5.1 Hydrogen constraints

MD simulations determine the future position of each atom by determining its position, velocity, and acceleration, and using these to calculate its position at a future point in time. However, the length of time over which the equations of motion are integrated must be relatively small because as the atomic positions change, the forces experienced by the atoms summarily change. If small enough timesteps are not used then the change in forces experienced by each atom is not correctly captured, and subsequently, the simulation will not be realistic. The length of time over which the equations of motion are integrated is therefore governed by the fastest oscillating atoms, as these will experience the change in forces acting on them most

rapidly. In biomolecular systems, these are the hydrogen atoms, which oscillate at a frequency that requires timesteps of 1 femtosecond. However, the flexibility of covalent hydrogen bonds is often less important than the low-frequency motions of the biomolecular system over much longer time periods. Therefore covalent bonds to hydrogen atoms are kept rigid to allow the timestep to increase to 2 femtoseconds, which allows observation of large-scale motions over longer time periods for the same computational cost. In our simulation,SHAKE algorithm [59] is used to constraint hydrogen bonds.

4.5.2 Different Types of Ensembles

The simulation process based on equation 4.1 will keep the total number of particle and the volume of the system constant along with the total energy of the system. This means we are working in an NVE system. But integration error, fluctuation in forces and inappropriate cut off leads to drifts in the total energy of the system. Not only this to compare our result with the experimental results it is better to keep constant pressure and constant temperature (lab condition) during the simulation. This can be done using one of the available thermostats and barostats.

Velocity scaling method is the easiest method for controlling the temperature of the system. This is based on the statistical idea that the temperature of the system is directly dependent on the velocity of the system, so the velocity of the molecules will be scaled using some functions of temperature ($\sqrt{\frac{T_B}{T(t)}}$, T_B is the desired temperature) such that in each stage this will bring back to the system to the desired temperature. However, this method won't make a true NVT system, or it is not a realistic approach for a thermostat. A better approach is the thermostat proposed by Berendsen. In this, we are coupling the system to a heat bath by making some modification in the equation of motion. This method allows to achieve rapidly the desired temperature, and it is very flexible. However, it does not allow to sample the correct canonical ensemble, and it can affect the system dynamics. Another widely known approach to temperature coupling is based on the generalized Langevin equation of motion [60].

$$m\frac{dv_i(t)}{dt} = F_i(t) - m_i\gamma_i v_i(t) + R_i(t) \qquad 4.16$$

where $v_i(t)$ is the velocity of particle i with mass m_i at time t, $F_i(t)$ is the systematic force, γ_i the friction coefficient and $R_i(t)$ is a stochastic force. In Eq 4.16, a constant temperature consistent with the canonical (NVT) ensemble is maintained by balancing the dissipative effect of the frictional terms with a stochastic force due to thermal noise. This approach create a proper NVT condition and allows several aspects of dynamics to be studied as a function of the friction rate. The most rigorous method to control the temperature

is to use the Nosé-Hoover thermostat. In this case, the heat bath becomes an integral part of the system by adding an artificial variable with an associated effective mass, so we have an extended Lagrangian. This method samples the correct canonical ensemble, and it does not seem to produce severe effects on the dynamics, even if some care is always needed.

The approaches to control the pressure are similar to those employed for the temperature. In this case, the instantaneous virial takes the role of the temperature and the pressure is controlled by scaling the system volume. One such method is the Berendsen barostat. In this, the system is weakly coupled to an external bath using the principle of least local perturbation. Similar to the temperature coupling, an extra term is added to the equations of motion that affect a pressure change.

$$(\frac{dP}{dt})_{bath} = \frac{P_0 - P}{T_p} \qquad 4.17$$

Whare T_P is the time constant for the coupling; P_0 is the desired pressure. An extra term is added to the equation of motion as

$$\frac{dx}{dt} = v + \alpha x \qquad 4.18$$

α is the coupling constant. Considering the effects on volume and and pressure

$$\alpha = -\beta \frac{(P_0 - P)}{3T_P} \qquad 4.19$$

β is the isothermal compressibility. This represents a represents a proportional scaling of coordinates. In each stage the coordinates are rescaled such that pressure remains constant.

4.6 Solvating the System (Water Models)

Water is the most abundant solvent in nature. It plays a vital role in several chemical and biological processes. All the biological process in taken place in the body fluid, which mainly consists of water. To consider this effect explicit water models are used in the computational studies. Implicit water models are also available which is known as the continuum model. The explicit models are determined from quantum mechanics, molecular mechanics, experimental results, and these combinations. To imitate a specific nature of molecules, many types of models have been developed. Broadly these classification is based on the number of interaction sites used in the model, the flexibility of the bonds and the polarization effect.

The simplest water model is the rigid model based on the non-bonded interactions. The bonded interactions

are treated implicitly by using different holonomic constraints. In this model, electrostatic interactions are treated by using Coulomb's law and Lenard-Jones potential is used for the Van der Waals interactions.TIP3P and TIP4P are good examples of the rigid model.

On the basis of interactions sites, different models like 2-site,3-site,4-site,5-site, and 6-site are using. Among this 3-site and 4-site are essential. 3-site models achieve high computational efficiency, and it is widely using MD simulations. Three-site models have three interaction points corresponding to the three atoms of the water molecule. Each site has a point charge, and the site corresponding to the oxygen atom also has the Lennard-Jones parameters. Most of the models use a rigid geometry matching that of actual water molecules. Most of the commonly used water models come under this category, which are TIP3P, SPC, SPC/E, etc. The four-site models have four interaction points by adding one dummy atom near of the oxygen along the bisector of the HOH angle of the three-site models. The dummy atom only has a negative charge. This model improves the electrostatic distribution around the water molecule. TIP4P is the commonly using 4-site water model.

The model SPC/E falls into the category of water model having an additional polarization parameter. In this model, there is an additional term in the Hamiltonian of the system which describes the polarization effect. Due to this addition, this model is giving better values of the density and the diffusion constant compared to SPC.

Another modification to the SPC model leads to the flexible SPC model. In this model, the O–H stretching is made anharmonic, and thus the dynamical behavior is well described. This is one of the most accurate three-center water models without taking into account the polarization. In molecular dynamics simulations it gives the correct density and dielectric permittivity of water

4.7 Minimisation Algorithms

Before starting the actual simulation process, we have to ensure that the system is in the stable equilibrium state. In computation minimization algorithms are doing this job.

4.7.1 Newton Raphson Method

This is a direct method to find the root of a function f, i.e., the points x_0 where f(x)=0. In this method, we initially start with a point x_1 and find the tangent of the function at the point x1, look for the intersection of the tangent to the x-axis and take this point as x_2 and repeat the process until we reach the minimum value.

By using taylor expansion in the first order and assuming that f is continuously differentiable

$$f(x_0 + h) = f(x_0) + hf'(x_0) \qquad 4.20$$

where h is small variation from the minimum value f' is the derivative of f with respect to x. From this we can get equation of the iteration ,i.e,

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} \qquad 4.21$$

However, this method will fail if the initial point is taken far from the actual minimum state. So for actual simulation this method is not that good.

4.7.2 Steepest Descent Method

The way it works is we start with an initial guess of the solution, and we take the gradient of the function at that point. We step the solution in the negative direction of the gradient, and we repeat the process. The algorithm will eventually converge where the gradient is zero (which correspond to a local minimum).

Let's say we are trying to find the solution to the minimum of some function f(x). Given some initial value x_0 for x, we can change its value in many directions (proportional to the dimension of x: with only one dimension, we can make it higher or lower). To figure out what is the best direction to minimize f, we take the gradient ∇f of it (the derivative along every dimension of x). Intuitively, the gradient will give the slope of the curve at that x and its direction will point to an increase in the function. So we change x in the opposite direction to lower the function value:

$$x_{n+1} = x_n - \lambda \nabla f(x_n) \qquad 4.22$$

The $\lambda>0$ is a small number that forces the algorithm to make small jumps. That keeps the algorithm stable and its optimal value depends on the function. Given stable conditions (a certain choice of λ), it is guaranteed that $f(x)_{n+1} \leq f(x)_n$.

4.7.3 Conjugate Gradient Method

In the conjugate gradient method, the first steps for the search of minimum points are similar as that of the steepest descent in which the system will be directed towards the largest gradient. However, to avoid the

oscillatory nature of the steepest descent in the near minimum stages the in conjugate gradient method mixes a little of the previous direction in the next steps. This will help in decreasing the time lag in the steepest descent method.

Chapter 5

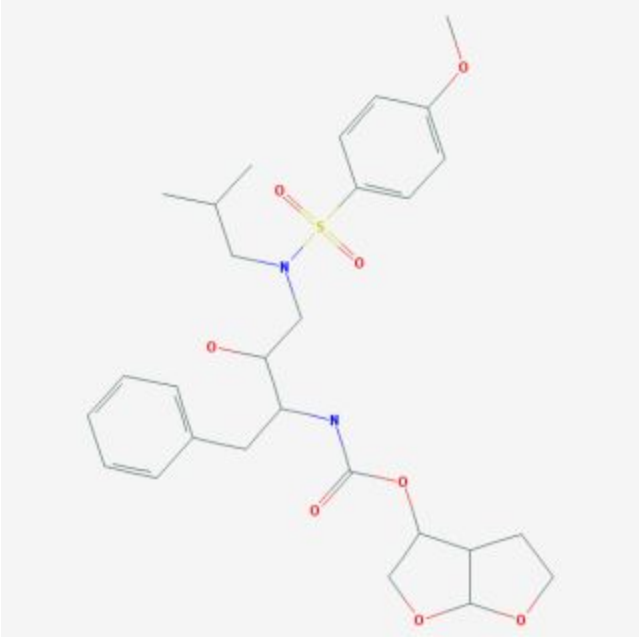
Materials and Methods

Published works discuss the high affinity of TMC-126 towards HIV 1 PR[29,61]-[62]. Different properties of TMC-126 is highlighted as the reason for the high affinity of the drug. A dynamical study based on thermodynamical properties of the binding of TMC-126 with HIV 1 PR will give an insight for designing new highly potent drug for HIV. In this project, we are doing the same by running an MD simulation on WT HIV 1 PR, five of its mutant variants and HIV 2 PR. The trajectories are analyzed using MMPBSA which gives a sound knowledge about the binding affinity of the drug and how mutation is affecting the binding affinity.

As we discussed in chapter one, the HIV 1 PR is the most potent drug target to interfere HIV life cycle. Inhibiting HIV 1 PR cause the production of immature viral particles and help to control the widespread effect of HIV. Structure and mechanism-based studies on HIV 1 PR lead to the discovery of HIV 1 inhibitors among which nine drugs are approved by FDA, and many are in the clinical stage. The nine FDA approved drugs are saquinavir (SQV), ritonavir (RTV), indinavir (IDV), nelfinavir (NFV), amprenavir (APV), lopinavir (LPV), atazanavir (AZV), tipranavir (TPV), and darunavir (DRV). These drugs are effective in inhibiting the progression of HIV 1 PR and leads to the production of immature viral particles. But the reverse transcription in the HIV is highly prone to errors which leads to mutations in the PR[63] and the inhibitors designed for the WT PR is not that effective in the mutant variants. In all affected persons the production of mutant variants decreases the effectiveness of the drug. This shows the importance of dynamical study of HIV 1 PR and its mutant variants.

5.1 TMC-126

In the late 1990s, the designs and syntheses of a new class of PIs have faced with the following significant challenges: 1) improvement of potency and pharmacokinetic properties which can substantially reduce therapeutic doses, maximize effectiveness eliminate the substantial "peptide-like" character and minimize side effects; 2) design of inhibitors that can effectively combat drug resistance; and 3)cost-effective synthesis of PIs to make these drugs readily accessible to third world countries, where the epidemic continues to worsen. Recognition of these problems leads to research effort to design new drugs which can resolve this problem. The efforts from Arun K Ghosh and group guide to the discovery of some highly potent non-peptide inhibitors including APV and TMC-126. One of the crucial elements in this design is to incorporate a stereochemically defined and conformationally constrained cyclic ether that will replace peptide bonds, mimic the biological mode of action, and make maximum interactions in the active site including hydrogen bonding with the protein backbone. The idea of incorporating cyclic ethers is from the observation that a number of naturally occurring biologically active motifs comprise cyclic ether as one of their epitopes. The addition of cyclic ether group leads to the discovery of APV. The high potency of APV against the WT and different mutant variants encourages the work. Based on the structure of APV, a lead



optimization program initially produced a series of **Fig.7.** Chemical structure of TMC-126 [94]

bis-tetrahydrofuranyl compounds of which TMC-126 was studied as the prototype [64]. The further increasing the efficiency of drug targeted on making maximum hydrogen bond and the hydrophobic interaction, which leads to the synthesize a new class of cyclic, fused bis-THF urethane-based HIV protease

inhibitors. Improvements in these drugs create the TMC-126, a prototype of DRV another FDA approved drug.

From the structure of TMC-126, both oxygen atoms of the bis-THF are within the hydrogen bond distance to D29 and D30 NH. Also, the 4-methoxyl oxygen atom is within hydrogen bond distance to D29' and D30' NH. These two factors are concluded as the reason for the high affinity of TMC-126 with HIV 1 PR [65]. compare to APV the entropy contribution to the binding is more favorable in the case of TMC-126 [66]. Upon selection of HIV-1 in the presence of TMC-126, mutants carrying a novel active site mutation A28S appeared along with L10F, M46I, I50V, A71Vand N88V [61]. This observation lead us to study about the A28S mutation which is not present in the case of other FDA approved drugs. In this project, we are also studying some other important mutation like V32I;mutation present in the cavity region and present in almost all the FDA approved drugs, M46L; present in the flap elbow region and common in other drugs, I50V; in the flap tip region, and MDR20 consist of twenty mutation. Also we are discussing the binding of the drug with another major subtype HIV 2 PR.

5.2 Input Files

As we discussed in chapter two, there are different force fields available for protein simulation. In our project, we are using AMBER17 package and using ff14SB force field for protein and GAFF2 for ligand simulation. The initial coordinates for our simulation are obtained from the X-ray crystallographic structure of HIV 1 PR complexed with TMC-126 (PDB id 2i4u). Hatada.M and co-workers have determined the crystallographic structure of PR1 complexed with TMC-126 with resolution 1.5Å in 2006 [67]. Mutant variants were created from the same crystallographic structure by changing the corresponding backbone atoms of the residue by hand and adding sidechain atoms using leap module of AMBER16 [68]. As the crystallographic structure of HIV 2 PR-TMC-126 was not available, we prepared the HIV 2 complex structure by keeping the same orientation of TMC-126 in 2i4u and replacing the receptor part with the help of CHIMERA visualizer (PDB id 5upj, HIV 2 complexed with some other drug) instead of using molecular docking. Published work with similar drugs shows that monoprotonated state of D25/D25' is more desirable for PR1, so for this project we consider the only monoprotonated state of catalytic residues. Protonation state is determined using Propka [69]. All the crystal water molecule is kept in the starting structure. Then the atomic partial charges of ligands are determined using AM1-BCC model [70,71] using antechamber [72] module of Amber [73]. AM1-BCC was used to calculate the partial charge of the ligand by adding Bond Charge Correction to the electron structure computed using Austin Model 1 [74] , which is a semi-empirical

quantum mechanical method for molecular electronic structure calculation. This will emulate HF/6-31G* potential on the surface of the molecule [70].

The complex structure was then solvated using TIP3P water model with truncated octahedron periodic box, extending at least 10 Å from the complex. An appropriate number of Cl-/Na+ atoms were added to neutralize the system. High-frequency bonds (hydrogen bonds) were constrained using SHAKE [75] algorithm which allows extending the time scale to 2fs. The Langevin thermostat [76] was used to keep the temperature at 300K with a collision frequency of 2 ps -1 . The particle-mesh Ewald (PME) [77] scheme was used to treat electrostatic interactions. The no bonded cut off was set to 10 Å.

5.3 Simulation Protocol

5.3.1 Minimisation

All the atoms of the complex is restrained to its initial position and we start the minimization procedure. Initially (first 500 steps) complex is minimized using steepest descent algorithm, and then we use the conjugate gradient algorithm for the minimization (next 500 steps). Then we remove the restrictions on the position of atoms and again optimize the complex using steepest descent algorithm (first 100 steps) followed by conjugate gradient algorithm (next 900 steps).

5.3.2 Equilibration

After minimization, in order to equilibrate the system at 300K without much drift in the structure we did a 50ps constant volume MD simulation using 2 Kcal mol -1 Å -2 restrain on the complex. This MD simulation is followed by another 50ps MD simulation at constant pressure (1 atm) keeping the same restrain on the complex using Berendsen's barostat for density equilibration. Then, the complex was equilibrated for 1ns without any restrains.

After all this equilibration the system was taken for a 100ns constant pressure MD simulation. During the simulation, the coordinates of the system were saved after every 10ps leads to a total of 10000 configuration for each simulation.

5.4 Methods used for Structural Analysis

5.4.1 Root Mean Square Deviation (RMSD)

Root Mean Square Deviation (RMSD) is a statistical measure of the similarity between two sets of values. In the field of biology, these values are commonly the atomic coordinates of homologous proteins. The proteins first need to be structurally aligned through the method of least-squares, which rotates one protein around its geometric center to match its orientation to the other protein. Then the protein is translated such that the sum of the distances between homologous pairs of atoms between the proteins is minimized. Once aligned, the RMSD is performed as

$$RMSD = \sqrt{\frac{1}{N}\sum_{i=1}^{N} d_i^2} \qquad 5.1$$

where N is the number of superimposed atoms in each set of values, and d_i is the Euclidean distance between the i^{th} pair of superimposed values.

Most commonly only the C_α atoms from the backbones of each protein are considered. This is because proteins that do not share exact sequence similarity do not necessarily share the same number of side-chain atoms. However, all amino acids share the same core backbone atoms (NH-CH-CO2), so regardless of the protein's genotype, as long as the RMSD is performed over homologous subsets of the proteins' structures, it will calculate a measure of structural similarity. Furthermore, as the C α atoms are located in the structural backbone of the protein, they represent a good indication of the tertiary- or quaternary-structure of the proteins.

5.4.2 Root Mean Square Deviation (RMSF) and B-factor

The variability in the conformation of trajectories can be monitored by calculating the root mean square fluctuations (RMSF) for individual atoms. The main difference between RMSD and RMSF is that RMSF is calculated over time. RMSF for a specific number of structures is defined as a root mean average square distance between an atom and its average position in a given set of structures and characterizes local changes along the protein chain. The RMSF captures, for each atom, the fluctuation about its average position. This gives insight into the flexibility of regions of the protein and corresponds to the crystallographic B-factors (temperature factors) and thus is a measure of the deviation between the positions of particle i and some reference position. The RMSF of a residue is computed as

$$\mathrm{RMSF} = [\sum_{i}^{N} < |R_i - < R_i >^2|^2 >]^{1/2} \qquad 5.2$$

<> indicates the time average and the corresponding B-factor is defined as

$$B = \frac{8}{3N}\pi^2(RMSF)^2 \qquad 5.3$$

5.4.3. Dynamic Correlation Matrix

The dynamic correlation matrix is defined as

$$C_{ij} = \frac{\langle x_i x_j \rangle}{\sqrt{\langle x_i^2 \rangle \langle x_j^2 \rangle}} \qquad 5.4$$

x_i and x_j are difference vectors between i^{th} and j^{th} C_α atom, respectively, and their average positions in the molecule-fixed frame. Correlations measured between fluctuating C_α carbon atoms would provide insight about protein function and pairwise correlation of residues. As the residue clusters are likely to involve inter-domain communication, which can not be seen in the interatomic interactions. This correlation may cause to the overall functioning of protein structure.

The measure of correlation between the fluctuations ΔX_i and ΔX_j of i^{th} and j^{th} C_α can be assessed by finding the projection of one on the other. The average of this projection indicates the strength of correlation. The positive average value of the projection indicates that the residues are correlated and the negative value indicates that the residues are anticorrelated. In the case of correlated residues they are moving in the same direction and the distance between these residues are not changing during the simulation. The anti-correlated motion indicates that the residues are moving in the opposite direction and distance between them will change during the simulation.

Chapter 6

Results and Discussion

6.1 Structure Stability

The Root Mean Square Deviation (RMSD) of the backbone atoms from the X-ray crystallographic structure of WT PR1, five of its mutant variants, and PR2 TMC-126 is shown in fig.8. The mean values of RMSD

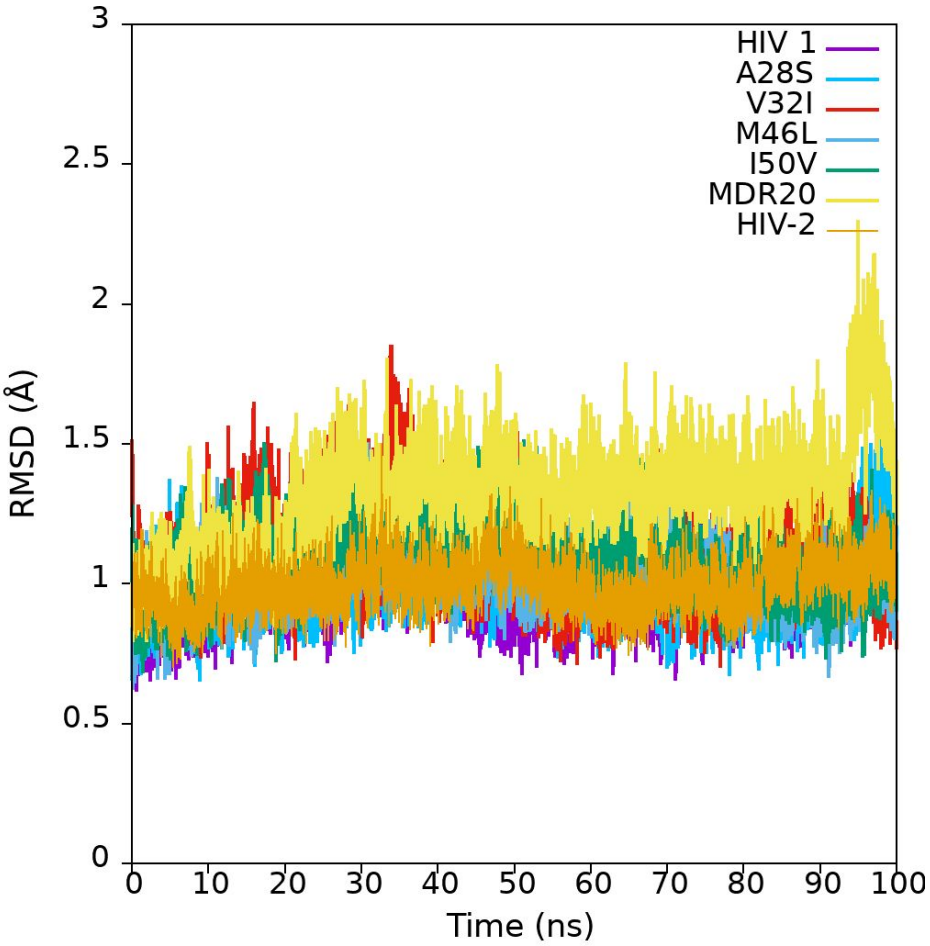


Fig.8. Time evolution of root-mean-square deviations (RMSD) of backbone atoms relative to their initial configurations for WT PR1, five of its mutant variant and PR2 complexed with TMC-126

(Table.1) are of few angstroms (<1.28 Å) only which indicates that the structure attains stability in the early stages of simulation as desired. There is a small fluctuation in RMSD of MDR20 around 90ns (Fig.8 yellow), So all the free energy calculations are taken from 50-90ns for MDR20 which form the stable region

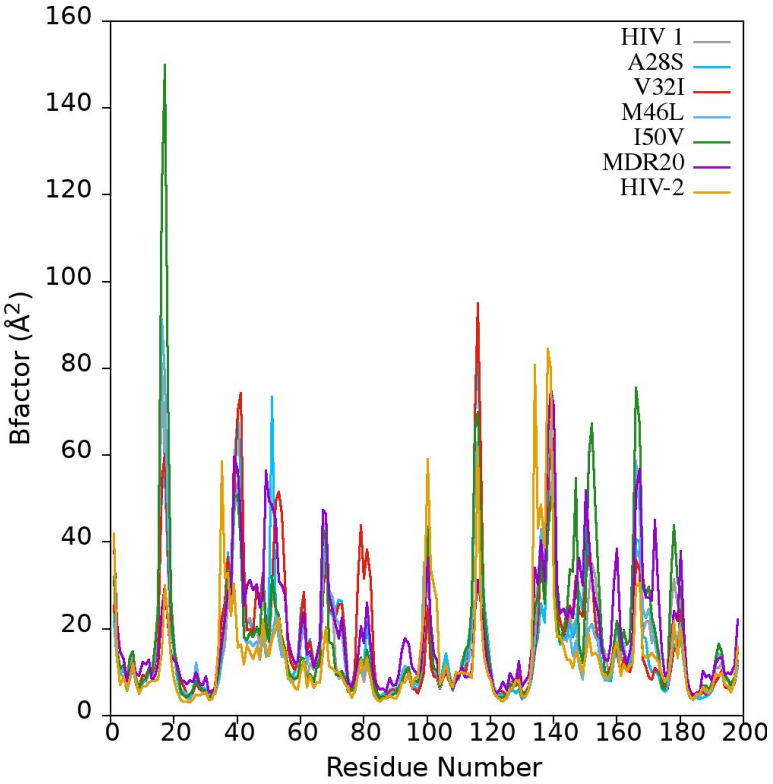
Variants	RMSD (Å)	Std of RMSD
WT (HIV 1)	0.97	0.11
A28S	1.00	0.12
V32I	1.10	0.16
M46L	1.02	0.11

I50V	1.09	0.13
MDR20	1.28	0.19
HIV 2	0.99	0.09

Table.1. Mean variation in the RMSD of back bone residue.

The B-factors of individual residues of TMC-126 complexed with PR1, five of its mutant variant and PR2 are

shown in fig.9. The higher values of fluctuation are shown by the side chain loop residues (In particular residues around G17 in I50V variant shows a fluctuation of 149.96Å). Other fluctuating residues are located



in

Fig.9. B-factor for TMC-126 complexed with WT PR1, five of its mutant variant and PR2 WT.Residues 1–99 and 100–198 correspond to residues 1–99 and 1 0 -99 0 , respectively

the flap regions; similar results are reported in earlier studies of PR1[78] and also for PR2[79]. In MDR20 flap region residues seems to be more dynamic. Except for MDR20 comparatively small B-factors are observed for catalytic dyads (3.22-5.02Å) for D25 and (3.40-5.57) for D25 ' . A slightly higher value is

observed for MDR20 8.26Å for D25 and 7.87Å for D25'. These small values are expected for D25/D25' as the catalytic function of these residues presumably requires a well-defined stable three-dimensional structure.

6.2 Dynamic Correlation and Flap Dynamics

To further understand the difference in the internal dynamics induced due to mutation cross-correlation matrices of Cα atom fluctuation after equilibrium of MD simulation were calculated. Overall inhibitor binding induces anti-correlated (blue in Fig.10) or non-correlated (green in Fig.10) motion than correlated motion. Complete correlation is shown by the diagonal part only, interaction with the same residue(black in Fig.10). But high correlation is found in residues near to 15, 50, 66 and residues from 20 to 28 with residues from 75 to 83 in both the chains (yellow in Fig.10). Correlated motion between residues from 20 to 28 with residues from 75 to 83 is in complete agreement with the low value of B-factor shown in Fig.9 of this region, showing the stability of cavity region. Highly anti-correlated motion is shown by flap region residues (red in Fig.10, box A/A') as expected. In WT PR1 the residues on the flap of first monomer (residues from 41-57) are anti-correlated with residues in the flap elbow region (residues from 32'-37') of the other monomer (region 1/1' of A/A' in Fig.10). Also, residues from 41 to 57(flap region) are anti-correlated with residues on the flap of the other monomer (residues from 48' to 56') (region 2/2' of A/A' in Fig.10). These residues in the first monomer are also anti-correlated with residues around 77' (region 3/3' of A/A' in Fig.10). This anti-correlation may be showing the flexibility of flap region. But flap tip region of one monomer (residues near 50) is correlated with flap tip region of the other monomer (residues near 50') (the yellow spot near 2/2' region in A/A'). This may be showing the closed state of the flap, which can also be concluded from the small value of 50-50' distance in Fig.11.

The dynamic correlation map of A28S is showing more anti correlated region compare to WT. Anti-correlation is mostly shown to the residues around 60/60 ' and with similar side residues. Anti-correlation in the flap region also increased. In V32I anti-correlation in A/A' is decreasing compared to WT. Also, the correlation in the flap

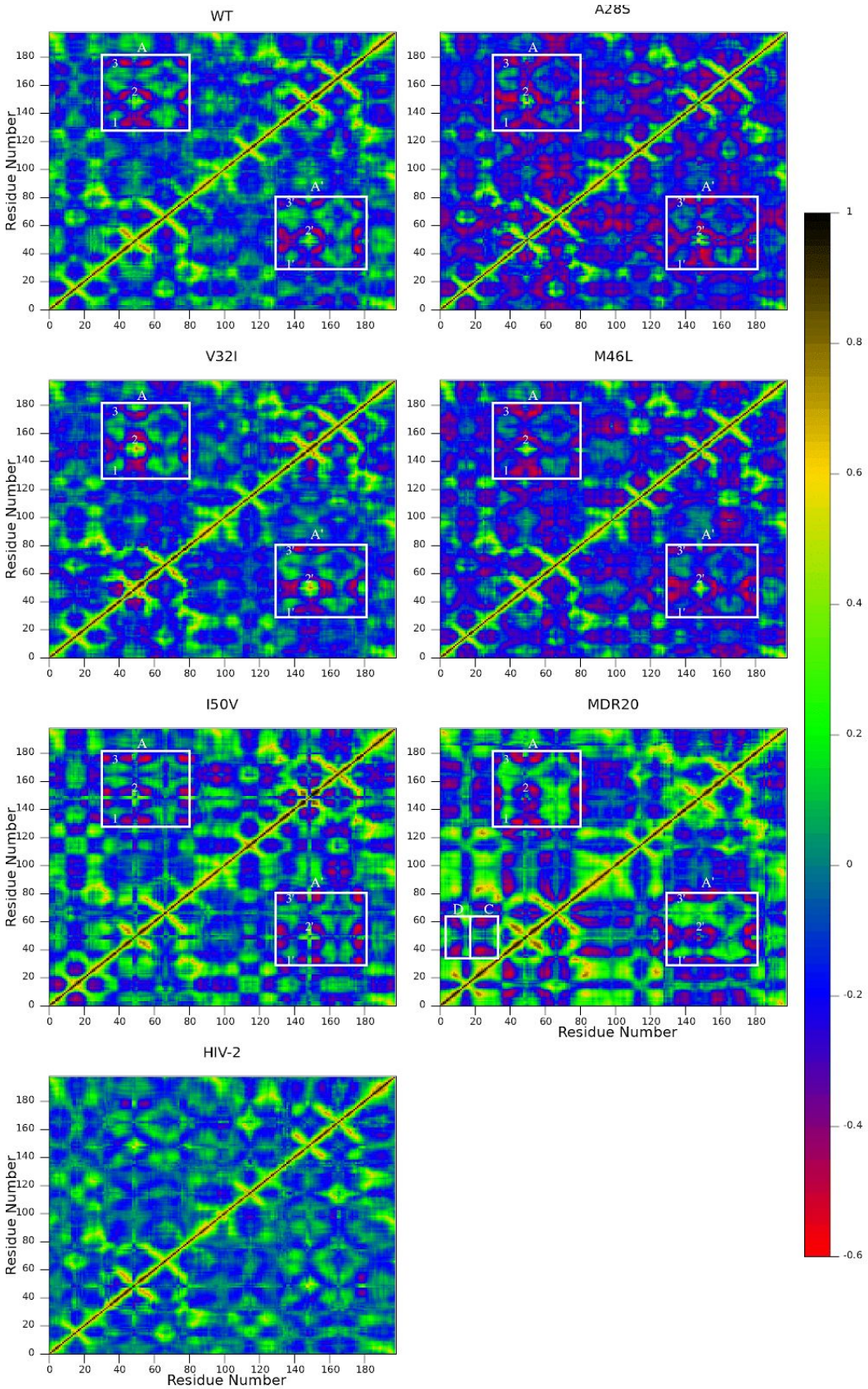
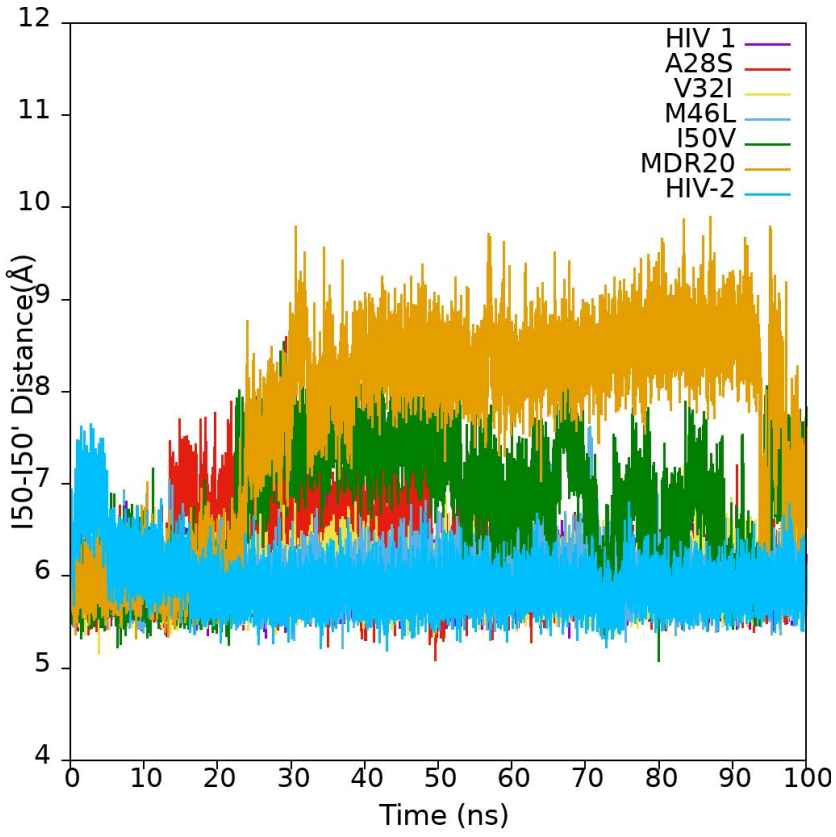


Fig.10. Dynamic correlation map for each residues of WT PR1, five of its mutant variants and PR2.

residues and decreases the correlation between the flap tip region (A/A' region of I50V in Fig.10). This may be showing the high flexibility of the flap region, and increased probability of flap region to undergo opening and

closing. From Fig.11 also we can conclude that the flaps are more opened in I50V variant in agreement with previous observation. In the multi-mutation also there is a similar observation as I50V for flap region. MDR20

tips (yellow spot) is increasing, similar the distance between 50-50 ' residues are decreasing (Table.2). This may be due to the reduced flexibility of the flap region induced by the mutation. Similar nature is showing for M46L mutation also. Mutation in the flap tip (I50V) increases the anti-correlation between flap and flap elbow is also showing a higher value of the distance between 50-50' in Table.2 There is a net increase in the anti-correlation in MDR20, between residues in the same monomers. In particular residues from 23 to 31 is anticorrelated with residues from 55 to 61 (region C of MDR20 in Fig.11. This may be due to D30N mutation as negatively charged D is replaced by polar N or may be due to Q58E as polar Q is replaced by negatively charged E. Similarly, the presence of new anticorrelation in residues from 7 to 10 with residues from 55 to 60 (region D in of MDR20 in Fig.10) may be due to Q7K or Q58E. In PR2 anti-correlation in the



flap-region is

Fig.11. The distance between the 50th and 50'th residue of PR1, five of its mutant variant and PR2 throughout the simulation (showing the opening and closing of flaps.

completely decreasing (Fig.9). This may be due to all the residues from 32 to 38 is different for PR1 and PR2, also except the flap tip residue, most of the residues in flap region are also different in PR1 and PR2.

6.3 Binding Free Energies

The different contribution to the binding free energy is separately studied using MM-PBSA analysis for understanding the binding mechanism of the protease, and the results are given in Table.2 and Fig.12. The binding free energy associated with the binding of TMC-126 with PR1 WT,A28S,V32I,M46L,I50V, MDR20 and PR2 in Kcal/mol are -11.29, +0.05, -5.32, -10.97, -9.39, -10.49, -8.25 respectively. Compare to one published work[80] our value of PR1 WT -TMC126 binding free energy is very less, but we are more close to experimental value[81] compare to them. The slight positive binding value in A28S is in agreement with published work[82]. Compare to the WT PR1 binding affinity is decreases in the order M46L>MDR20>I50V>V32I>A28S for mutant variants. TMC-126 is severely affected by the A28S and V32I mutation. Also, the TMC-126 is showing less affinity towards PR2.

Variants	ΔE_{vdW}	ΔE_{ele}	ΔG_{np}	ΔG_{pol}	ΔG_{solv}	$\Delta G_{pol+ele}$	$-T\Delta S$	ΔH	ΔG_{bind}
WT	-63.63 (0.06)	-57.08 (0.13)	-5.87 (0.00)	83.23 (0.08)	77.36 (0.08)	26.15 (0.15)	32.07 (1.16)	-43.36 (0.08)	-11.29 (1.16)
A28S	-60.90 (0.06)	-48.28 (0.18)	-5.99 (0.00)	84.02 (0.12)	78.02 (0.12)	35.74 (0.21)	31.21 (1.21)	-31.16 (0.12)	+0.05 (1.27)
V32I	-60.00 (0.07	-52.55 (0.12)	-5.99 (0.00)	83.17 (0.09)	77.18 (0.09)	31.62 (0.15)	30.05 (0.64)	-35.37 (0.09)	-5.32 (0.65)
M46L	-62.51 (0.06)	-53.22 (0.14)	-5.88 (0.00	81.72 (0.10)	75.84 (0.09)	28.50 (0.17)	28.93 (0.45)	-39.90 (0.08)	-10.97 (0.47)
I50V	-63.56 (0.06)	-58.50 (0.11)	-6.06 (0.00)	87.88 (0.07)	81.81 (0.07)	29.38 (0.13)	30.86 (0.53)	-40.25 (0.08)	-9.39 (0.54)
MDR20	-65.77 (0.05)	-45.41 (0.07)	-6.46 (0.00)	74.71 (0.08)	68.25 (0.07)	29.30 (0.10	32.44 (0.58)	-42.93 (0.07)	-10.49 (0.58)
HIV 2	-64.42 (0.06)	-39.44 (0.12	-5.84 (0.00)	68.84 (0.08)	63.00 (0.08)	29.40 (0.15	32.61 (0.59)	-40.86 (0.08)	-8.25 (0.59)

Table.2. Binding free energy and its components of TMC-126 complexed with PR1,five of its mutant variants and PR2 (in Kcal/mol)

Standard errors of the mean are provided in parentheses. $\Delta G_{solv} = \Delta G_{np} + \Delta G_{pol}$; $\Delta G_{pol,elec} = \Delta G_{pol} + \Delta E_{elec}$

The contribution favoring the binding is the Van der Waals (-65.77 to -60.00), Electrostatic interaction (-58.50 to -39.44) and nonpolar interaction (-6.46 to -5.84). The contributions from polar interaction (68.84 to 87.88) and configurational entropy (32.61 to 28.93) terms are opposing the binding.

The Van der Waals contribution to the free energy of binding is more or less same in all the variants. The intermolecular electrostatic interaction includes the contribution mainly from the hydrogen bonds, and from the water-mediated hydrogen bond. There has only one water-mediated hydrogen bond in all the variants except for V32I and A28S variants, which has two water bridge. So the hydrogen bond formation mainly decides the electrostatic contribution. From the fig.13 and fig.14. The creation of more number of hydrogen bond leads to more favorable electrostatic contribution in I50V and WT. Also, less number of hydrogen bonds (only 1) leads to the low value of electrostatic interaction in PR2. A28S, M46L and MDR20 variants have a similar amount

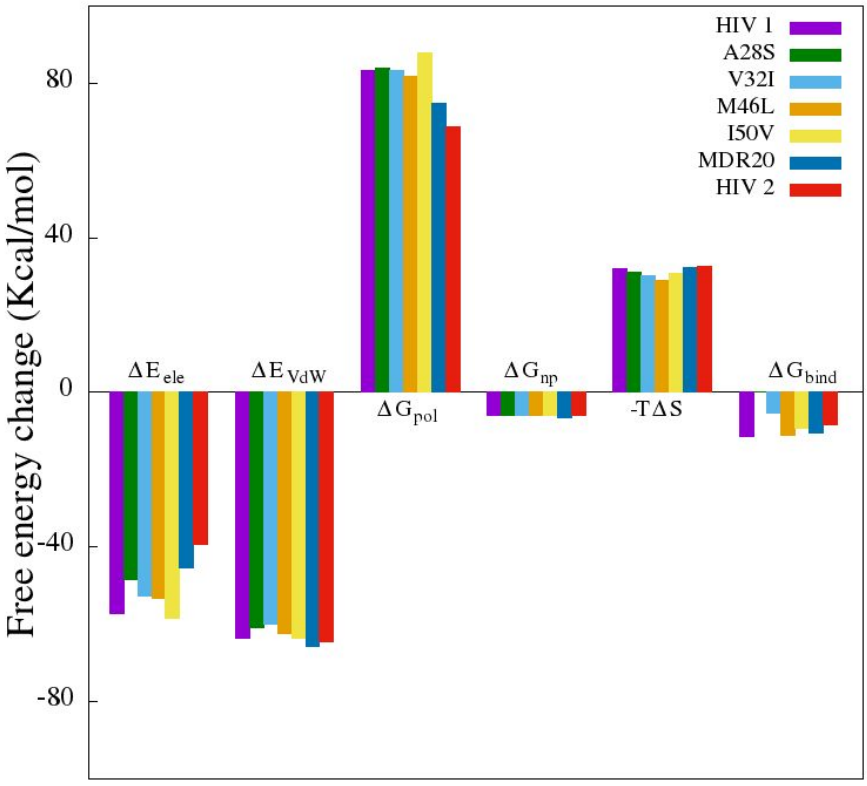


Fig.12. Binding free energy and its components of TMC-126 complexed with PR1,four of its mutant variant and PR2

of hydrogen bond (Fig.5). But in MDR20 the second hydrogen bond is forming with I50 with an average distance of 2.9Å, but in A28S and M46L the second bond is forming with D25 ' (OD2) with an average

distance of 2.7Å which may be the reason for the higher value of electrostatic interaction in M46L and A28S compare to MDR20. The frequency of occurrence of hydrogen bond is more in M46L compared to A28S, as expected the electrostatic interaction energy is more to M46L compare to A28S. V32I is showing less occurrence of hydrogen bond compared to A28S but it have higher electrostatic interaction energy (Table.3) may be because of the more frequent occurrence of two water mediated hydrogen bond compare to two water mediated hydrogen bonds of A28S. Also, the donor hydrogen atom for the second hydrogen bond in MDR20 is bonded with the nitrogen atom (NH-O hydrogen bond) which has less binding energy compared to the hydrogen bond forming with OH-O which is present in V32I. These may be the reason for higher favorable electrostatic contribution in V32I and A28S. In published works[83,84] it is mentioned that formation of the

hydrogen bond with D29/D29’ and D30/D30’ can be the reason for the higher potency of TMC-126. In our 100ns simulation, only one among the oxygen atom in the bis-THF moiety is forming the hydrogen bond, that

is with D29' (42.64% occupancy), and 4-mexothyl oxygen is not making any hydrogen bond also. But there is a favorable electrostatic interaction from D29 (-2.97 Kcal/mol) and D30 (-4.40 Kcal/mol) and the net electrostatic interaction from D29' is opposing as though it is forming the hydrogen bond.

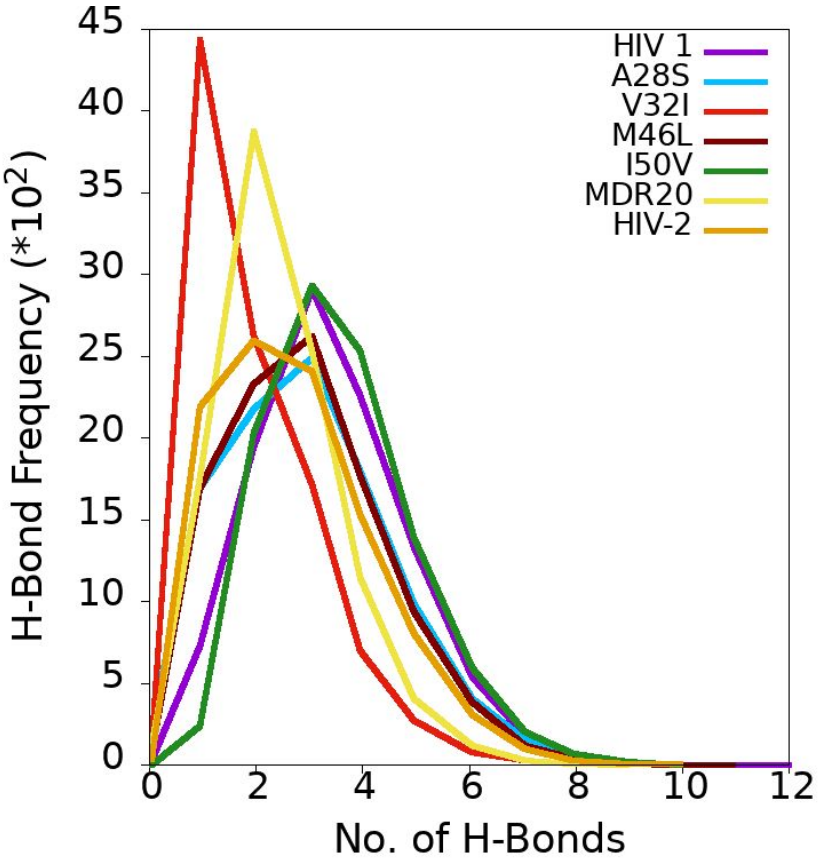


Fig.13. Frequency of occurrence of hydrogen bond during 100ns simulation of TMC-126 complexed with PR1, five of its mutant variants and PR2

The polar nature of PR1 and PR2 is reflected in the higher positive value of the polar interaction. In this system, the polar contribution is due to the work done by the inhibitor to displace the water molecule attracted to the hydrophilic residues of the receptor which is also electrostatic in nature. The term ΔG pol,ele is showing the sum of both polar and electrostatic contribution which is positive in all cases, this indicates that polar terms overcompensate the favorable contribution from the electrostatic interaction as in many other systems[85,86]. So the binding process is mainly driven by Van der Waals interaction only.During the formation of macromolecules, there will be a reduction in the degrees of freedom of the individual molecules. This fact, in general, oppose the formation of macromolecules[86],[87]. This loss of translational, rotational and vibrational degrees of freedom of the individual compounds are reflected in the entropic contribution to the binding free energy. The higher positive value of entropic contribution may be due to the higher number of rotatable bonds (13 in number) which increases the degree of freedom of the ligand in its unbound state, which is constrained during binding.

6.3.1 A28S Mutation

This mutation is present in the catalytic region close to the catalytic triads. The 28th residue is conservative in all the subtypes of HIV, but it is reported that TMC-126 can't inhibit the variants having the A28S mutation which is in full agreement with our result. The direct effect on the drug through VdW interaction is very less but in total VdW interaction is decreased by 2.73 Kcal/mol mainly due to the residues like G27 (close residue to the mutation) G49' (fig.18). There is a drastic decrease in the electrostatic interaction due to this mutation which seems to be the main reason for decreased inhibition in this variant. There is an increased favorable electrostatic contribution from residues like D29' and K45' (an increase of 1.26 and 1.86 Kcal/mol), but it can't compensate the decreased contribution from S28' (mutated residue) and R87' (decrease of 2.42 and 2.52), etc. (fig.19). In the case of 28' residue, the decrease in the electrostatic interaction is not compensated in polar solvation interaction.

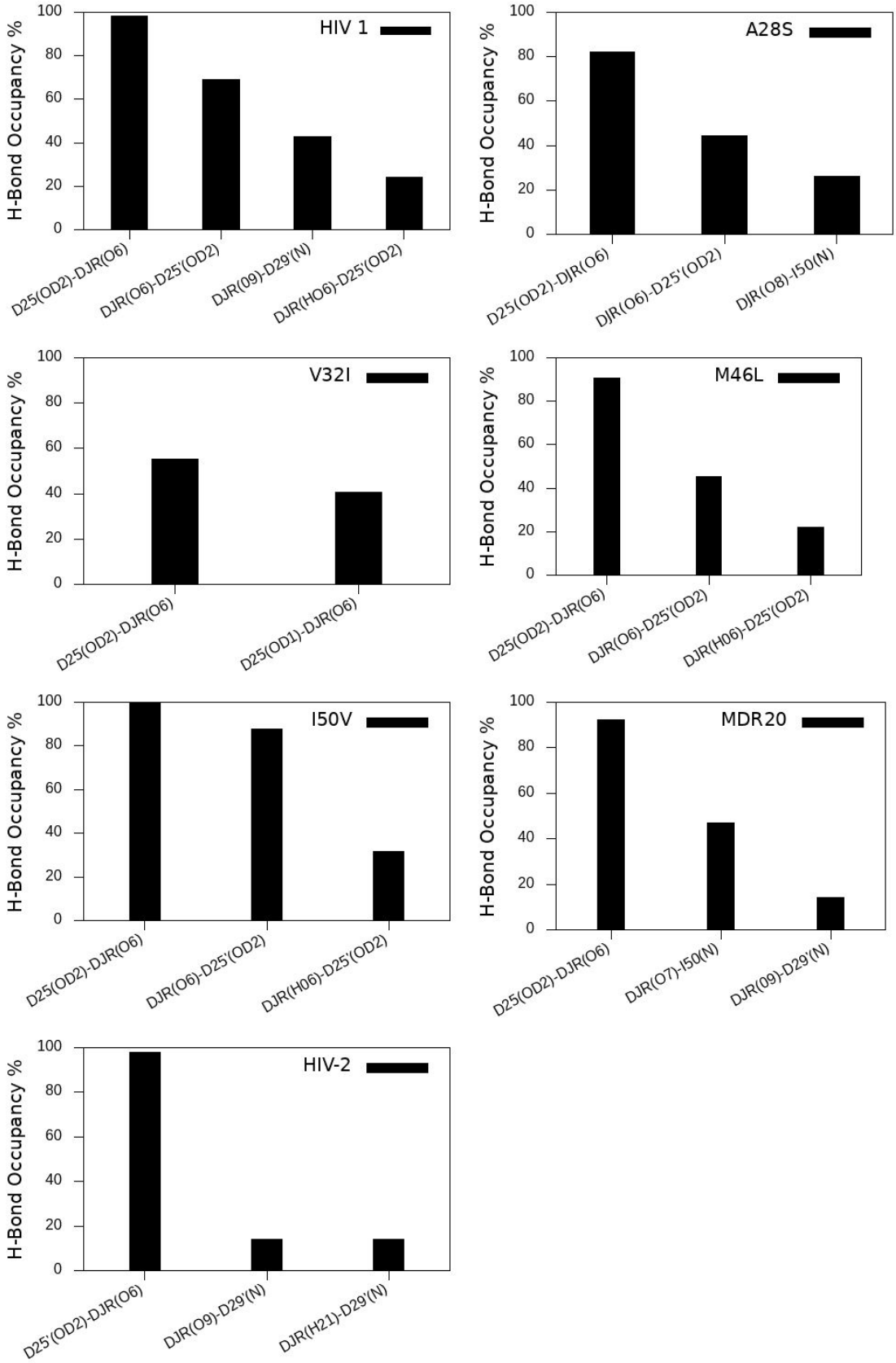


Fig.14. Hydrogen bond occupancy plot for PR1,five of its mutant variant and PR2 with TMC-126

6.3.2 V32I Mutation

V32I mutation is an active site mutation, which can directly contribute to the drug resistance through unfavorable interaction with the drug. As isoleucine is more prominent compared to valine[88] , the direct interaction may mainly through VdW. It is clear from Table.2 that there is a drastic decrease in the affinity of TMC-126 due to this mutation. V32I mutation directly increases the binding affinity by 0.11 Kcal/mol (V32 to I32) and 0.48 Kcal/mol (V32' to I32') through vdW interaction, but residues like D25/D25' also increases the binding through vdW more than this [Fig.18]. So direct effect of the mutation is less compared to the indirect effect. Net vdW interaction is decreasing by a factor of 3.63 kcal/mol due to decrease in the vdW interaction at residues like G27', A28' ,D29', R8 and D30' [Fig.18]. The decrease in favorable contribution from VdW is may due to the increase in the cavity volume (467.25 in WT to 548.78 in V32I) which is known as one reason for drug resistance [89]. Electrostatic interaction also decreases by a factor of 4.53 Kcal/mol due to this mutation, which mainly due to the breaking of a hydrogen bond with catalytic residue D25' (fig14.) (contribution of D25' decreases by a factor of 2.71 Kcal/mol, 59.8% of the decrease). But there is a favorable decrease in the electrostatic repulsion from D29' and D30', but this can't compensate the unfavorable contribution from D25', A28' etc. [Fig.19]. There has a net favorable contribution from the entropic and hydrophobic interaction. In total, the catalytic residue D25' is contributing favorably to V32I [Fig.15] (unfavorable contribution from electrostatic interaction (+2.48) i equally compensated by a favorable contribution from polar solvation energy (-2.72) by increasing the binding energy by an amount of 0.85 Kcal/mol. Similarly, D25 is also contributing favorably to binding in V32I. But the net decrease in the binding energy is mainly due to unfavorable contribution from A28' [fig.15] (decreasing the binding energy by an amount of 1.84 Kcal/mol, 30.8% of the total decrease) (Table.4.). In A28' unfavorable contribution from electrostatic is not compensated by polar solvation energy.

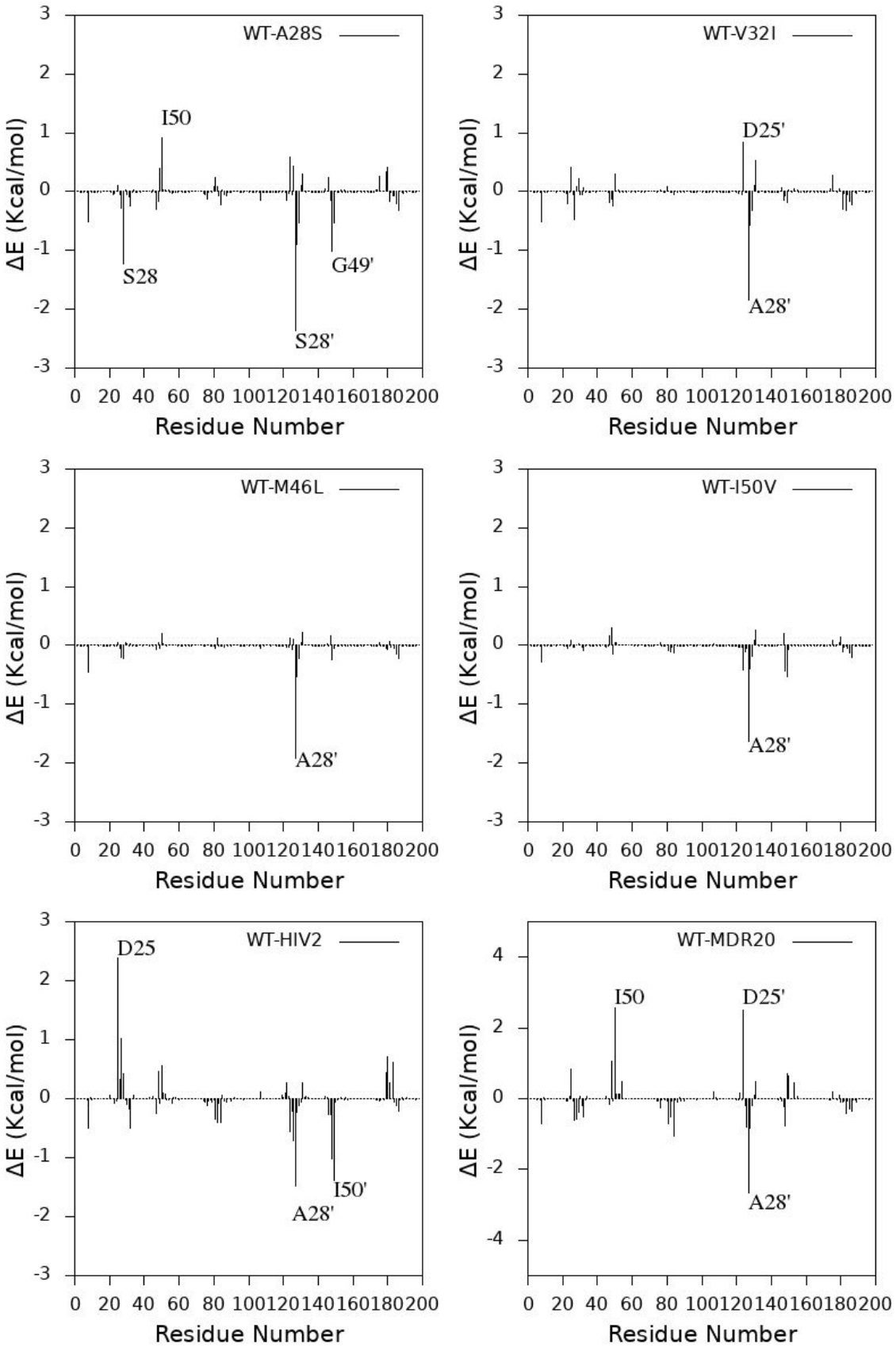


Fig.15. Difference in the total binding energy of PR2 and five mutant variants of PR 1 compare to the WT structure of HIV 1 $\{\Delta G_{Bind}(WT) - \Delta G_{Bind}(Mutant)\}$

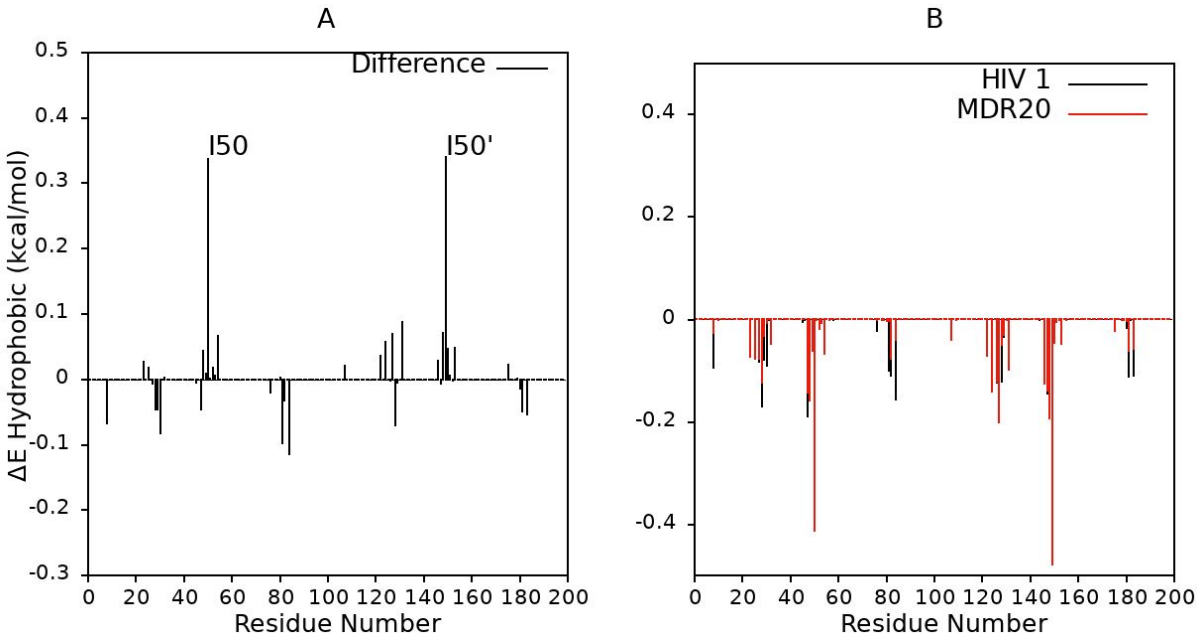


Fig.16. A. Difference in the hydrophobic energy of WT PR1 and MDR20 mutant variant (WT-MDR20) B. Hydrophobic energy contribution from WT(black) and MDR20 mutant variant (Red).

6.3.3 M46L Mutation

This mutation is located in the flap region. This can't make any direct effect on the drug. Compare to other mutation TMC-126 can overcome this mutation. Due to mutation, residues like I50, D25' etc. contributes more favorable to the VdW interaction while residues like G27,G27', A28' [Fig.18] lost their favorable interaction [Fig.12], which leads to a loss of 1.12 Kcal/mol in VdW energy contribution. Electrostatically residues like D29' and K45' are giving a more favorable contribution, but the loss of electrostatic energy in A28' R87' and R8 residues are higher [Fig.19.] which leads to a loss of 3.86 Kcal/cal. The polar solvation compensates all the favorable and unfavorable contribution from electrostatic interaction except for A28' which leads this to be the most affected residue due to mutation (fig.15) (decreasing the total binding energy by an amount of 1.92 Kcal/mol) (Table.4).

6.3.4 I50V Mutation

I50 is located at the tip of the flap region on each PR monomer. V50' residue is decreasing the vdW contribution by 0.44Kcal/mol compared to I50 in WT. This can be explained as Isoleucine is bigger

compared to Valine. But this decrease in VdW is compensated by increased contribution from other residues like G48 and I47 which leads to a negligible effect on net vdW interaction. In total electrostatic interaction is making favorable effects on mutation. But there is a decrease in the electrostatic contribution from residues like A28', R8 and R87' which is overcoming by the favorable contribution from many residues mainly both of the catalytic residues (D25/D25') [Fig.19]. Similar to M46L mutation most of the favorable and unfavorable electrostatic contribution are compensated by polar solvation energy, but this compensation is very less for A28' which is making it the most affected residue due to mutation [fig.15]. Both the hydrophobic and entropic contributions are also favoring the binding of TMC-126 with the mutated receptor.

6.3.5 MDR20 Mutation

It consists of 20 mutations in different part of PR1. Compared to other mutations that we have studied, this mutation shows higher fluctuations in the contributions from various residues for all the different interactions compare to the WT. Residues like I50, I50', G48 (flap region residues; mutations in flap region are I54L and I47V) D25 and D25' (catalytic residues) are contributing more favorably through VdW interaction (above 0.5 Kcal/mol) after mutation [Fig.18]. There is a relative decrease in the VdW contribution in residues, such as V84, N30', N30 (all mutated), A28 and A28'[Fig.18]. A decrease in vdW for residues N30/N30' may be due to D30N mutation. N is larger compared to D, then also VdW interaction is decreasing may be drug is moving away from the residue through electrostatic repulsion as D is charged. But in total VdW interaction is contributing more favorably to the binding than in WT. The decrease in the electrostatic contributions to the binding free energy seems to be the main reason for drug resistance to this mutation. There is more than 2Kcal/mol decrease in the electrostatic contribution from the residues like R8, D29, N30, G48, A28'and R87' [Fig.19]. The decrease in R8 may be due to Q7K mutation as K has an additional positive charge. In A28',

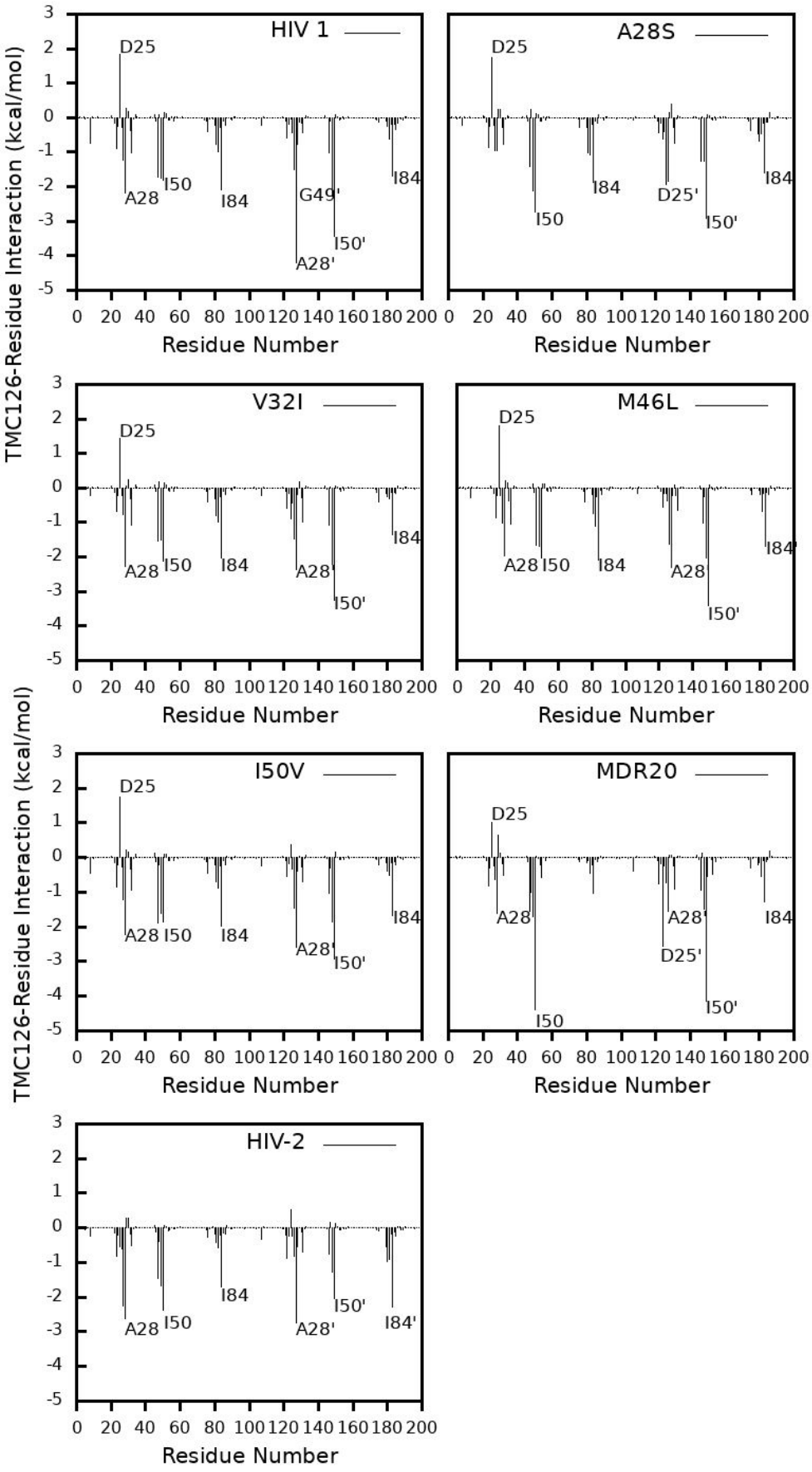


Fig.17. Interaction free energy of TMC-126 with each of the residues in the receptor of PR1, five of its mutant variants and HIV 2

D29 and N30 decrease in electrostatic contribution are may be due to D30N mutation, as in WT 4-mexothyl oxygen may be interacting with NH of D30. Similarly reduced electrostatic interaction in R87' may be due to N88D mutation as D has an extra charge compare to N. Residues like L45, I50, R87, R8', D29' and N30' are contributing favorably through electrostatic [Fig.13] which can also explain by similar arguments. As in all other mutation cases, electrostatic interactions are compensated by polar contribution except for A28'. In this mutation, there is a formation of additional hydrophobic interactions in I50 and I50' which is favoring the binding [Fig.16.]. In total both I50 and D25' is contributing more favorably to the binding after mutation but A28' is contributing unfavorably to the binding [Fig.15.]. Single V32I mutation is causing considerable resistance in the binding of TMC-126, but due to other mutation resistance of V32I is somehow overcoming in MDR20.

6.3.5 Comparison between PR1 and PR2

PR1 and PR2 shares about 50% of sequence identity and TMC-126 is a drug designed for PR1. There is a decrease of 3.04 Kcal/mol difference in the total binding energy (Table.2.). In PR2 also TMC-126 is binding firmly with the same residues as in PR1 (Fig.17). There is a small increase in the VdW contribution in PR2 compare to PR1, the main improvement in the VdW contribution is from the catalytic residue (D25) in which the VdW contribution is increased more than 1.5 Kcal/mol [Fig.12]. But there is a considerable decrease in the favorable electrostatic contribution (17.64 Kcal/mol). This decrease may be due to the breakdown of the hydrogen bond between the drug and D25 (there is a net decrease of 13.84 Kcal/mol favorable electrostatic contribution only for D25). In PR2 D25' contributing more favorably towards the electrostatic interaction by strengthening the hydrogen bond (net increase of 10.27 Kcal/mol) [Fig.19]. Similar to all other mutation most of the electrostatic interactions are compensated in the polar solvation energy except for few residues. From Fig.6. Residues like D25 A28' and I50' are the residues which are having higher effects due to mutation. From Fig.17, it is observed that TMC-126 is interacting more strongly with residues such as A28/A28', I47/I47', I50/I50', and I84/I84'. Contribution from 50th residue is not that affecting due to I50V mutation, similarly the case of 84th residue (I84V included in MDR20) and 47th residue (I47V mutation included in MDR20). A28/A28' is a conservative residue in all the subtypes of HIV and this residue is mostly affected due to other mutation. The mutation in this residue is making the drug unbind to the receptor.

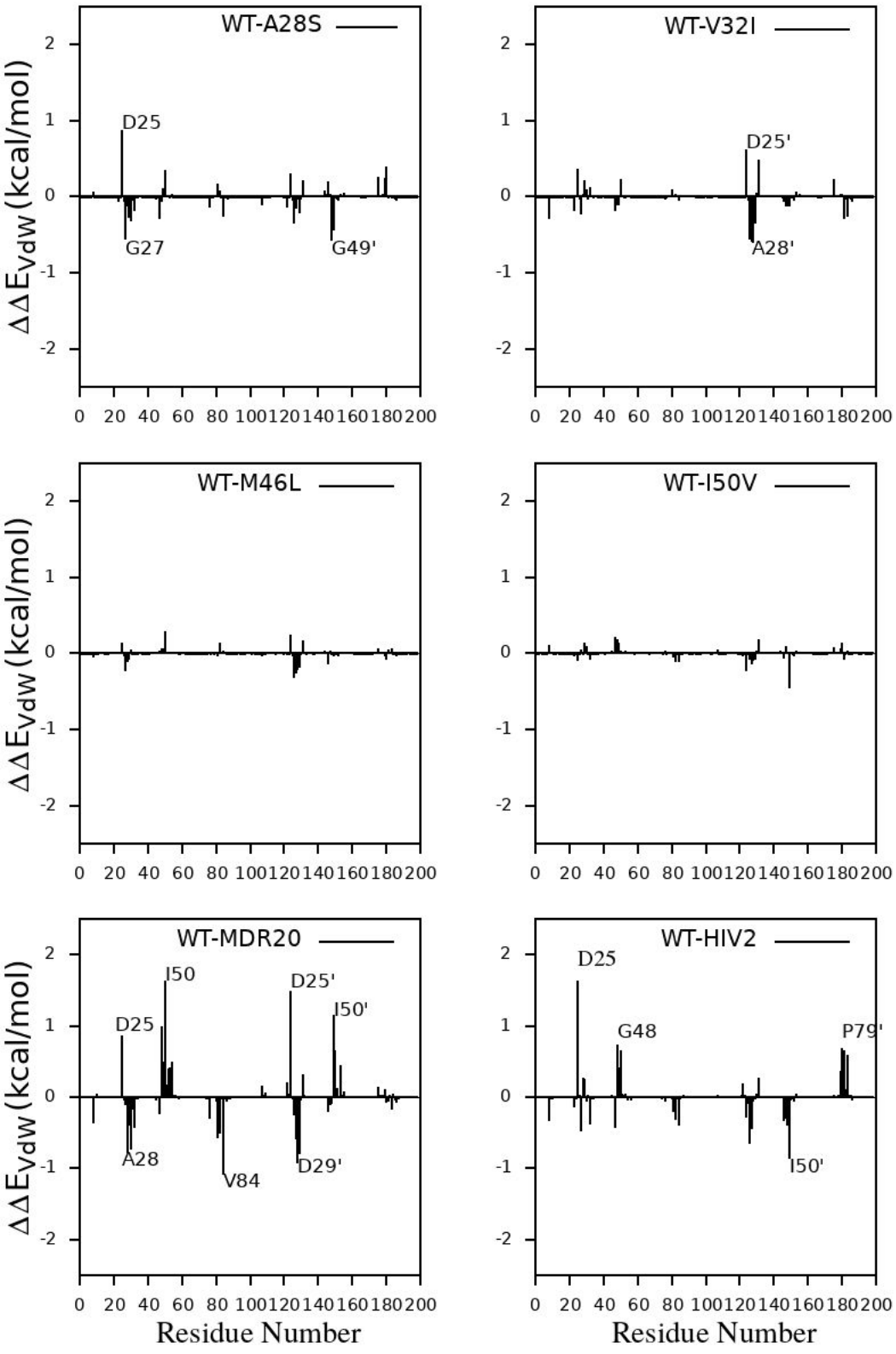


Fig.18. Difference in the VdW contribution to the free energy of different mutant variants of PR1 compare to its WT structure (VdW contribution from WT – VdW contribution from mutant variant)

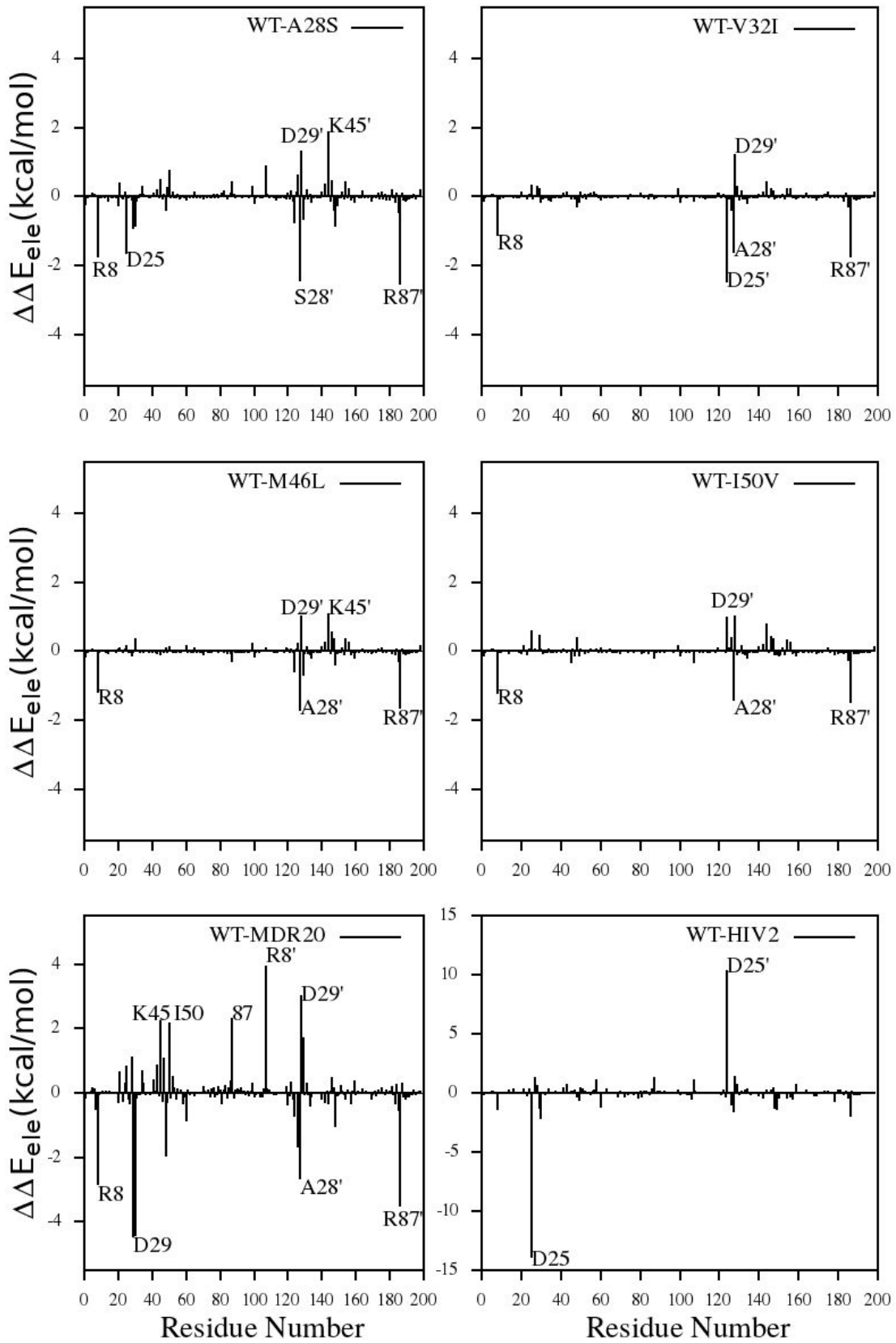


Fig.18. Difference in the electrostatic contribution to the free energy of different mutant variants of PR1 compare to its WT structure (Electrostatic contribution from WT – Electrostatic contribution from mutant variant)

Chapter 7

Conclusions

In this work, we have investigated the mechanism of binding of the inhibitor TMC-126 with the wild-type and five different mutant variants of HIV-1 protease. We have also studied the effectiveness of this inhibitor against HIV-2 protease for which currently no drug is available. For this purpose, we have conducted atomistic molecular dynamics simulations for 100 ns and later, employed Molecular Mechanics -Poisson-Boltzmann Surface Area (MM-PBSA) scheme to elucidate how different molecular forces compete with each other resulting in the complex formation. Over all, our investigation suggests that for all cases, the intermolecular electrostatic and van der Waals interactions as well as nonpolar component of the solvation free energy favor the binding, while the configurational entropy and polar solvation free energy disfavor the complex formation. Furthermore, we notice that the Coulombic interactions is overcompensated by the polar solvation free energy. This means that the net electrostatic interaction energy is unfavorable to the complex formation. Therefore, in all cases, the binding is mainly driven by the van der Waals interactions.

Our computational investigations predict that the drug will lose its potency against all five mutant variants. Among all the mutations we have studied, A28S and V32I are highly resistant against TMC-126 mainly due to decrease in electrostatic interactions. There is also a decrease in van der Waals interactions resulting in drug resistance. In cases of other mutations, such as M46L, I50V and MDR20 drug resistance occurs mainly due to unfavorable change in electrostatic interactions. The inhibitor is found to lose its potency significantly against HIV-2 protease due to unfavorable shift in intermolecular electrostatic interactions. This underscores the critical need of designing new drugs for combating HIV-2.

Over all, this investigation is very useful for designing new drugs. Our study suggests that a more potent drug against HIV protease can be obtained by maximizing the van der Waals interactions rather than increasing the number of hydrogen bonds between the ligand and protein.

In the future, we shall evaluate the potency of various currently used drugs against commonly found mutations in the protease. We shall also extend the current study to different subtypes of HIV. In the end, we hope to provide new insights into the mechanism of binding of drugs to HIV-1/HIV-2 protease which can be helpful in designing new drugs.

Reference

1. Hahn BH, Shaw GM, De Cock KM, Sharp PM. AIDS as a zoonosis: scientific and public health implications. Science. 2000;287: 607–614.

2. Jaffar S, Grant AD, Whitworth J, Smith PG, Whittle H. The natural history of HIV-1 and HIV-2 infections in adults in Africa: a literature review. Bull World Health Organ. 2004;82: 462–469.

3. Martinez-Steele E, Awasana AA, Corrah T, Sabally S, van der Sande M, Jaye A, et al. Is HIV-2-induced AIDS different from HIV-1-associated AIDS? Data from a West African clinic. AIDS. 2007;21: 317–324.

4. Vicenzi E, Poli G. Novel factors interfering with human immunodeficiency virus-type 1 replication in vivo and in vitro. Tissue Antigens. 2013;81: 61–71.

5. Simon V, Ho DD, Karim QA. HIV/AIDS epidemiology, pathogenesis, prevention, and treatment. Lancet. 2006;368: 489–504.

6. Eckert DM, Kim PS. Mechanisms of Viral Membrane Fusion and Its Inhibition. Annu Rev Biochem. 2001;70: 777–810.

7. Feinberg MB, Baltimore D, Frankel AD. The role of Tat in the human immunodeficiency virus life cycle indicates a primary effect on transcriptional elongation. Proc Natl Acad Sci U S A. 1991;88: 4045–4049.

8. Hanes JW, Johnson KA. Exonuclease removal of dideoxycytidine (zalcitabine) by the human mitochondrial DNA polymerase. Antimicrob Agents Chemother. 2008;52: 253–258.

9. Roberts J, Bebenek K, Kunkel T. The accuracy of reverse transcriptase from HIV-1. Science. 1988;242: 1171–1173.

10. Arts EJ, Hazuda DJ. HIV-1 antiretroviral drug therapy. Cold Spring Harb Perspect Med. 2012;2: a007161.

11. Dalgleish AG, Beverley PC, Clapham PR, Crawford DH, Greaves MF, Weiss RA. The CD4 (T4) antigen is an essential component of the receptor for the AIDS retrovirus. Nature. 1984;312: 763–767.

12. Reimann KA, Lin W, Bixler S, Browning B, Ehrenfels BN, Lucci J, et al. A humanized form of a CD4-specific monoclonal antibody exhibits decreased antigenicity and prolonged plasma half-life in rhesus monkeys while retaining its unique biological and antiviral properties. AIDS Res Hum Retroviruses. 1997;13: 933–943.

13. Kuritzkes DR, Jacobson J, Powderly WG, Godofsky E, DeJesus E, Haas F, et al. Antiretroviral activity of the anti-CD4 monoclonal antibody TNX-355 in patients infected with HIV type 1. J Infect Dis. 2004;189: 286–291.

14. Lin P-F, Blair W, Wang T, Spicer T, Guo Q, Zhou N, et al. A small molecule HIV-1 inhibitor that targets the HIV-1 envelope and inhibits CD4 receptor binding. Proc Natl Acad Sci U S A. 2003;100:

11013–11018.

15. Karn J, Stoltzfus CM. Transcriptional and posttranscriptional regulation of HIV-1 gene expression. Cold Spring Harb Perspect Med. 2012;2: a006916.

16. Cupelli LA, Hsu MC. The human immunodeficiency virus type 1 Tat antagonist, Ro 5-3335, predominantly inhibits transcription initiation from the viral promoter. J Virol. 1995;69: 2640–2643.

17. Hwang S, Tamilarasu N, Kibler K, Cao H, Ali A, Ping Y-H, et al. Discovery of a small molecule Tat-trans-activation-responsive RNA antagonist that potently inhibits human immunodeficiency virus-1 replication. J Biol Chem. 2003;278: 39092–39103.

18. Sundquist WI, Kräusslich H-G. HIV-1 assembly, budding, and maturation. Cold Spring Harb Perspect Med. 2012;2: a006924.

19. Navia MA, Fitzgerald PMD, Mckeever BM, Springer JP. THREE-DIMENSIONAL STRUCTURE OF ASPARTYL PROTEASE FROM HUMAN IMMUNODEFICIENCY VIRUS HIV-1 [Internet]. 1989. doi:10.2210/pdb2hvp/pdb

20. Gustchina A, Weber IT. Comparative analysis of the sequences and structures of HIV-1 and HIV-2 proteases. Proteins. 1991;10: 325–339.

21. Tie Y, Wang Y-F, Boross PI, Chiu T-Y, Ghosh AK, Tozser J, et al. Critical differences in HIV-1 and HIV-2 protease specificity for clinical inhibitors. Protein Sci. 2012;21: 339–350.

22. Bagossi P, Sperka T, Fehér A, Kádas J, Zahuczky G, Miklóssy G, et al. Amino acid preferences for a critical substrate binding subsite of retroviral proteases in type 1 cleavage sites. J Virol. 2005;79: 4213–4218.

23. Navia MA, Fitzgerald PMD, Mckeever BM, Springer JP. THREE-DIMENSIONAL STRUCTURE OF ASPARTYL PROTEASE FROM HUMAN IMMUNODEFICIENCY VIRUS HIV-1 [Internet]. 1989. doi:10.2210/pdb2hvp/pdb

24. Foulkes-Murzycki JE, Scott WRP, Schiffer CA. Hydrophobic Sliding: A Possible Mechanism for Drug Resistance in Human Immunodeficiency Virus Type 1 Protease. Structure. 2007;15: 225–233.

25. Hornak V, Okur A, Rizzo RC, Simmerling C. HIV-1 protease flaps spontaneously open and reclose in molecular dynamics simulations. Proc Natl Acad Sci U S A. 2006;103: 915–920.

26. Miller M, Jaskólski M, Rao JK, Leis J, Wlodawer A. Crystal structure of a retroviral protease proves relationship to aspartic protease family. Nature. 1989;337: 576–579.

27. Perryman AL, Lin J-H, Andrew McCammon J. HIV-1 protease molecular dynamics of a wild-type and of the V82F/I84V mutant: Possible contributions to drug resistance and a potential new target site for drugs. Protein Sci. 2004;13: 1434–1434.

28. Brik A, Wong C-H. HIV-1 Protease: Mechanism and Drug Discovery. ChemInform. 2003;34. doi:10.1002/chin.200318281

29. Hou T, Yu R. Molecular dynamics and free energy studies on the wild-type and double mutant HIV-1 protease complexed with amprenavir and two amprenavir-related inhibitors: mechanism for binding and drug resistance. J Med Chem. 2007;50: 1177–1188.

30. Smith R, Brereton IM, Chai RY, Kent SB. Ionization states of the catalytic residues in HIV-1 protease. Nat Struct Biol. 1996;3: 946–950.

31. Yamazaki T, Nicholson LK, Wingfield P, Stahl SJ, Kaufman JD, Eyermann CJ, et al. NMR and X-ray Evidence That the HIV Protease Catalytic Aspartyl Groups Are Protonated in the Complex Formed by the Protease and a Non-Peptide Cyclic Urea-Based Inhibitor. J Am Chem Soc. 1994;116: 10791–10792.

32. Blundell TL, Sewell BT, Jenkins JA, Tickle IJ. THE ACTIVE SITE OF ASPARTIC PROTEINASES [Internet]. 1982. doi:10.2210/pdb2ape/pdb

33. Ghosh AK, Osswald HL, Prato G. Recent Progress in the Development of HIV-1 Protease Inhibitors for the Treatment of HIV/AIDS. J Med Chem. 2016;59: 5172–5208.

34. Jakalian A, Jack DB, Bayly CI. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. J Comput Chem. 2002;23: 1623–1641.

35. Ohtaka H, Schön A, Freire E. Multidrug resistance to HIV-1 protease inhibition requires cooperative coupling between distal mutations. Biochemistry. 2003;42: 13659–13666.

36. Todd MJ, Luque I, Velázquez-Campoy A, Freire E. Thermodynamic basis of resistance to HIV-1 protease inhibition: calorimetric analysis of the V82F/I84V active site resistant mutant. Biochemistry. 2000;39: 11876–11883.

37. Coffin JM. HIV population dynamics in vivo: implications for genetic variation, pathogenesis, and therapy. Science. 1995;267: 483–489.

38. Weiner SJ, Kollman PA, Case DA, Singh UC, Ghio C, Alagona G, et al. ChemInform Abstract: A NEW FORCE FIELD FOR MOLECULAR MECHANICAL SIMULATION OF NUCLEIC ACIDS AND PROTEINS. Chemischer Informationsdienst. 1984;15. doi:10.1002/chin.198419080

39. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM, et al. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. J Am Chem Soc. 1995;117: 5179–5197.

40. Gough CA, Debolt SE, Kollman PA. Derivation of fluorine and hydrogen atom parameters using liquid simulations. J Comput Chem. 1992;13: 963–970.

41. Wang J, Cieplak P, Kollman PA. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? J Comput Chem. 2000;21: 1049.

42. Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. Proteins. 2006;65: 712–725.

43. Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. J Chem Theory Comput. 2015;11: 3696–3713.

44. MacKerell A Jr. Atomistic Models and Force Fields. Computational Biochemistry and Biophysics. 2001.

45. Foloppe N, MacKerell AD, Jr. All-atom empirical force field for nucleic acids: I. Parameter

optimization based on small molecule and condensed phase macromolecular target data. J Comput Chem. 2000;21: 86–104.

46. Jorgensen WL, Tirado-Rives J. The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. J Am Chem Soc. 1988;110: 1657–1666.

47. Kollman P. Free energy calculations: Applications to chemical and biochemical phenomena. Chem Rev. 1993;93: 2395–2417.

48. Lybrand TP, McCammon JA, Wipff G. Theoretical calculation of relative binding affinity in host-guest systems. Proc Natl Acad Sci U S A. 1986;83: 833–835.

49. Miller BR 3rd, McGee TD Jr, Swails JM, Homeyer N, Gohlke H, Roitberg AE. MMPBSA.py: An Efficient Program for End-State Free Energy Calculations. J Chem Theory Comput. 2012;8: 3314–3321.

50. Jayaram B, Sprous D, Young MA, Beveridge DL. Free Energy Analysis of the Conformational Preferences of A and B Forms of DNA in Solution. J Am Chem Soc. 1998;120: 10629–10633.

51. Honig B, Nicholls A. ChemInform Abstract: Classical Electrostatics in Biology and Chemistry. ChemInform. 2010;26: no–no.

52. Klapper I, Hagstrom R, Fine R, Sharp K, Honig B. Focusing of electric fields in the active site of Cu-Zn superoxide dismutase: effects of ionic strength and amino-acid modification. Proteins. 1986;1: 47–59.

53. Dominy BN, Brooks CL. Development of a Generalized Born Model Parametrization for Proteins and Nucleic Acids. J Phys Chem B. 1999;103: 3765–3773.

54. Onufriev A, Bashford D, Case DA. Exploring protein native states and large-scale conformational changes with a modified generalized born model. Proteins. 2004;55: 383–394.

55. Still WC, Clark Still W, Tempczyk A, Hawley RC, Hendrickson T. Semianalytical treatment of solvation for molecular mechanics and dynamics. J Am Chem Soc. 1990;112: 6127–6129.

56. Finkelstein AV, Janin J. The price of lost freedom: entropy of bimolecular complex formation. Protein Eng Des Sel. 1989;3: 1–3.

57. Lybrand TP, McCammon JA, Wipff G. Theoretical calculation of relative binding affinity in host-guest systems. Proc Natl Acad Sci U S A. 1986;83: 833–835.

58. de Souza ON, Ornstein RL. Effect of periodic box size on aqueous molecular dynamics simulation of a DNA dodecamer with particle-mesh Ewald method. Biophys J. 1997;72: 2395–2397.

59. Ryckaert J-P, Ciccotti G, Berendsen HJC. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. J Comput Phys. 1977;23: 327–341.

60. Paterlini MG, Ferguson DM. Constant temperature simulations using the Langevin equation with velocity Verlet integration. Chem Phys. 1998;236: 243–252.

61. Yoshimura K, Kato R, Kavlick MF, Nguyen A, Maroun V, Maeda K, et al. A Potent Human Immunodeficiency Virus Type 1 Protease Inhibitor, UIC-94003 (TMC-126), and Selection of a Novel

(A28S) Mutation in the Protease Active Site. J Virol. 2002;76: 1349–1358.

62. Ohtaka H, Velázquez-Campoy A, Xie D, Freire E. Overcoming drug resistance in HIV-1 chemotherapy: the binding thermodynamics of Amprenavir and TMC-126 to wild-type and drug-resistant mutants of the HIV-1 protease. Protein Sci. 2002;11: 1908–1916.

63. Roberts J, Bebenek K, Kunkel T. The accuracy of reverse transcriptase from HIV-1. Science. 1988;242: 1171–1173.

64. Ghosh AK, Ramu Sridhar P, Kumaragurubaran N, Koh Y, Weber IT, Mitsuya H. Bis-tetrahydrofuran: a privileged ligand for darunavir and a new generation of hiv protease inhibitors that combat drug resistance. ChemMedChem. 2006;1: 939–950.

65. Ghosh AK, Kincaid JF, Cho W, Walters DE, Krishnan K, Hussain KA, et al. Potent HIV protease inhibitors incorporating high-affinity P2-ligands and (R)-(hydroxyethylamino)sulfonamide isostere. Bioorg Med Chem Lett. 1998;8: 687–690.

66. Ohtaka H, Velázquez-Campoy A, Xie D, Freire E. Overcoming drug resistance in HIV-1 chemotherapy: The binding thermodynamics of Amprenavir and TMC-126 to wild-type and drug-resistant mutants of the HIV-1 protease. Protein Sci. 2002;11: 1908–1916.

67. Hatada M. HIV-1 protease with TMC-126 [Internet]. 2007. doi:10.2210/pdb2i4u/pdb

68. Case DA, Cheatham TE 3rd, Darden T, Gohlke H, Luo R, Merz KM Jr, et al. The Amber biomolecular simulation programs. J Comput Chem. 2005;26: 1668–1688.

69. Li H, Robertson AD, Jensen JH. Very fast empirical prediction and rationalization of protein pKa values. Proteins. 2005;61: 704–721.

70. Jakalian A, Jack DB, Bayly CI. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. J Comput Chem. 2002;23: 1623–1641.

71. Jakalian A, Bush BL, Jack DB, Bayly CI. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: I. Method. J Comput Chem. 2000;21: 132.

72. Wang J, Wang W, Kollman PA, Case DA. Automatic atom type and bond type perception in molecular mechanical calculations. J Mol Graph Model. 2006;25: 247–260.

73. Case DA, Cheatham TE 3rd, Darden T, Gohlke H, Luo R, Merz KM Jr, et al. The Amber biomolecular simulation programs. J Comput Chem. 2005;26: 1668–1688.

74. Dewar MJS, Zoebisch EG, Healy EF, Stewart JJP. Development and use of quantum mechanical molecular models. 76. AM1: a new general purpose quantum mechanical molecular model. [Erratum to document cited in CA103(2):11627f]. J Am Chem Soc. 1993;115: 5348–5348.

75. Ryckaert J-P, Ciccotti G, Berendsen HJC. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. J Comput Phys. 1977;23: 327–341.

76. Paterlini MG, Ferguson DM. Constant temperature simulations using the Langevin equation with velocity Verlet integration. Chem Phys. 1998;236: 243–252.

77. Darden T, York D, Pedersen L. Particle mesh Ewald: An N·log(N) method for Ewald sums in large

systems. J Chem Phys. 1993;98: 10089–10092.

78. Zhu Z, Schuster DI, Tuckerman ME. Molecular dynamics study of the connection between flap closing and binding of fullerene-based inhibitors of the HIV-1 protease. Biochemistry. 2003;42: 1326–1333.

79. Kar P, Knecht V. Origin of decrease in potency of darunavir and two related antiviral inhibitors against HIV-2 compared to HIV-1 protease. J Phys Chem B. 2012;116: 2605–2614.

80. Hou T, Yu R. Molecular dynamics and free energy studies on the wild-type and double mutant HIV-1 protease complexed with amprenavir and two amprenavir-related inhibitors: mechanism for binding and drug resistance. J Med Chem. 2007;50: 1177–1188.

81. Hou T, Yu R. Molecular dynamics and free energy studies on the wild-type and double mutant HIV-1 protease complexed with amprenavir and two amprenavir-related inhibitors: mechanism for binding and drug resistance. J Med Chem. 2007;50: 1177–1188.

82. Yoshimura K, Kato R, Kavlick MF, Nguyen A, Maroun V, Maeda K, et al. A Potent Human Immunodeficiency Virus Type 1 Protease Inhibitor, UIC-94003 (TMC-126), and Selection of a Novel (A28S) Mutation in the Protease Active Site. J Virol. 2002;76: 1349–1358.

83. Ghosh AK, Kincaid JF, Cho W, Walters DE, Krishnan K, Hussain KA, et al. Potent HIV protease inhibitors incorporating high-affinity P2-ligands and (R)-(hydroxyethylamino)sulfonamide isostere. Bioorg Med Chem Lett. 1998;8: 687–690.

84. Ohtaka H, Velázquez-Campoy A, Xie D, Freire E. Overcoming drug resistance in HIV-1 chemotherapy: The binding thermodynamics of Amprenavir and TMC-126 to wild-type and drug-resistant mutants of the HIV-1 protease. Protein Sci. 2002;11: 1908–1916.

85. Kuhn B, Kollman PA. Binding of a Diverse Set of Ligands to Avidin and Streptavidin: An Accurate Quantitative Prediction of Their Relative Affinities by a Combination of Molecular Mechanics and Continuum Solvent Models. J Med Chem. 2000;43: 3786–3791.

86. Kar P, Lipowsky R, Knecht V. Importance of polar solvation for cross-reactivity of antibody and its variants with steroids. J Phys Chem B. 2011;115: 7661–7669.

87. Knecht V. Model amyloid peptide B18 monomer and dimer studied by replica exchange molecular dynamics simulations. J Phys Chem B. 2010;114: 12701–12707.

88. Wu TD, Schiffer CA, Gonzales MJ, Taylor J, Kantor R, Chou S, et al. Mutation patterns and structural correlates in human immunodeficiency virus type 1 protease following different protease inhibitor treatments. J Virol. 2003;77: 4836–4847.

89. Clavel F, Hance AJ. HIV drug resistance. N Engl J Med. 2004;350: 1023–1035.

90. https://microbeonline.com/describe-structure-of-hiv-virus-t-u-2058/

91. https://aidsinfo.nih.gov/understanding-hiv-aids/glossary/1596/life-cycle

92. https://doi.org/10.1073/pnas.1102278108

93. http://ambermd.org/tutorials/basic/tutorial1/section5.htm

94. https://pubchem.ncbi.nlm.nih.gov/compound/200104#section=Top