Computational investigation of HIV-1 protease dynamics by comparing the effects of mutation, force fields and pressure

A THESIS SUBMITTED IN PARTIAL FULFILMENT
OF THE REQUIREMENTS FOR THE AWARD OF THE DEGREE OF

DOCTOR OF PHILOSOPHY

BY

BISWA RANJAN MEHER

ROLL NO. 04610604

DEPARTMENT OF BIOTECHNOLOGY

INDIAN INSTITUTE OF TECHNOLOGY GUWAHATI

GUWAHATI, INDIA

MARCH 2009

TH-772_04610604

Computational investigation of HIV-1 protease dynamics by comparing the effects of mutation, force fields and pressure

A THESIS SUBMITTED IN PARTIAL FULFILMENT
OF THE REQUIREMENTS FOR THE AWARD OF THE DEGREE OF

DOCTOR OF PHILOSOPHY

BY

BISWA RANJAN MEHER

ROLL NO. 04610604

DEPARTMENT OF BIOTECHNOLOGY

INDIAN INSTITUTE OF TECHNOLOGY GUWAHATI

GUWAHATI, INDIA

MARCH 2009

TH-772_04610604

Dedicated
to
My Bada Bapa & Parents

TH-772_04610604

INDIAN INSTITUTE OF TECHNOLOGY GUWAHATI

DEPARTMENT OF BIOTECHNOLOGY

STATEMENT

I do hereby declare that the matter embodied in this thesis is the result of investigations carried out by me in the Department of Biotechnology, Indian Institute of Technology Guwahati, India under the guidance of Dr. Rajaram Swaminathan, my present supervisor. In keeping with the general practice of reporting scientific observations, due acknowledgements have been made wherever the work described is based on the findings of other investigators.

Date: 2nd March, 2009 **Biswa Ranjan Meher**

IIT Guwahati

TH-772_04610604

INDIAN INSTITUTE OF TECHNOLOGY GUWAHATI

DEPARTMENT OF BIOTECHNOLOGY

Dr. Rajaram Swaminathan
Associate Professor
Department of Biotechnology,
IIT Guwahati, Assam-781039, INDIA

Phone: +91-361-258-2248
Fax: +91-361-258-2249
Email: rsw@iitg.ernet.in

CERTIFICATE

It is certified that the work described in this thesis entitled "***Computational investigation of HIV-1 protease dynamics by comparing the effects of mutation, force fields and pressure.***" by Biswa Ranjan Meher for the award of degree of Doctor of Philosophy is an authentic record of the results obtained from the research work carried out under my supervision in the Department of Biotechnology, Indian Institute of Technology Guwahati, India, and this work has not been submitted elsewhere for a degree.

Date: 2nd March, 2009

IIT Guwahati

Dr. Rajaram Swaminathan

(Supervisor)

TH-772_04610604

INDIAN INSTITUTE OF TECHNOLOGY, GUWAHATI

Department of Biotechnology

CERTIFICATE OF COURSE WORK

This is to certify that Biswa Ranjan Meher has satisfactorily completed all the courses required for the Ph.D degree program. These courses include

BT 602:	Basic Biotechnology
BT 605:	Gene Therapy
BT 606:	Food Biotechnology
BT 607:	Plant Biotechnology
BT 608:	Microbial Biotechnology
BT 610:	Frontiers in Biomolecular simulations

Biswa Ranjan Meher has successfully completed his Ph.D qualifying examination in December 2005.

Dr. Pranab Goswami
Head, Department of Biotechnology
I. I. T. Guwahati

Dr. S. S. Ghosh
Secretary, DPPC
I. I. T. Guwahati

TH-772_04610604

Acknowledgements

The results of the works reported in the thesis are based on interdependence rather than independence, which reflect where I have been escorted and supported by many people and that gives me a pleasant opportunity to express my gratitude to all of them. First and foremost, with a deepest sense of gratitude, I would like to express my sincere gratitude to my present supervisor Dr. Rajaram Swaminathan, for his valuable suggestions, encouragement and inspiration. Also it is with my deepest sense of appreciation that I express my heartiest acknowledgement to my former research supervisor Dr. Pradipta Bandyopadhyay for introducing me into the interdisciplinary topic “Computational Biology” and giving me the freedom to pursue my own interests. I earnestly thank him for his astute guidance, creative and scientific ideas which helped me to enhance my knowledge in this field and made my pursuit really exciting.

I would like to acknowledge my sincere gratitude to all my doctoral committee members, Dr. V. Dasu, Dr. Ashish Gupta and Dr. Biplab Bose for their insightful advices and valuable suggestions.

I would like to give special thanks to Dr. Utpal Bora and Dr. Sanjukta Patra for their constant inspiration and help. My sincere thanks are due to all other faculty members in the Department of Biotechnology for their help and encouragement and the non-teaching staffs of the Department for their technical support. I would like to take this opportunity to thank Department of Biotechnology for proving me the necessary computational facilities during the entire duration of my research tenure. I am thankful to the Institute, Indian Institute of Technology Guwahati for providing me with the state of the art infrastructure and facilities for advanced research. The financial support from All India Council for Technical Education (AICTE), New Delhi in the form of (National Doctoral Fellowship) NDF is duly acknowledged.

I would like to thank my research group members Venkat, Kausik, Satish, Ravi, Nividh for their timely help, constant support, and for the wonderful time we shared during this period.

I would like to extend my gratitude to my departmental friends Mahanty, Atul, Satish, Gopi, Siva, Kiran, Shampa, Preety, Priyanka, Prarthana, Bandana, Sharan, Nurul, Anita, Niranjan for their timely support and steady help.

TH-772_04610604

I would also like to offer my special thanks to my friends@JNU Prashant, Amit, Payal, Raghu, Sarbhasis and Smriti for their encouragement and all the help they extended during the time I spent over there.

I extend my sincerest thanks to my top friends@Siang Sandeep (Gadhe), Subash, Biswanath, Jugal (Alia), Smruti (Pabana), Sumanta (Sadhu), Laxmidhar (Budha), Ballav (Controll), Dr. Sahid Hussain (Dayal Baba), Dr. Biswa Ranjan Panda (Saare), Mahanty Bhai, Atul (Doc Saab), Satish (Bhaiya), Kiran, Dr. P. Gopinath and Dr. Siva (Mama) for their constant help, motivation, enthusiastic company and all the wonderful time we spent in various events.

How can I forget my "Utkal University (Vanivihar)", my friends over there? They are and will be always there in my soul. I must acknowledge my friends for their constant feeds.

I earnestly express my heartiest gratitude to my loving junior Seema, for her constant support, caring and help whenever needed and for being a continuous source of inspiration.

Finally, my Ph. D. endeavor could not be completed without the endless love, unending support, tolerance and blessings from my family. I wish to express my sincere gratitude to my parents, my brother, and especially to my Bada Bapa. They are the main soul and inspiration for each and every step that I achieve in my life.

Date: 2nd March, 2009. **Biswa Ranjan Meher**

TH-772_04610604

Abstract:

AIDS, one of the major epidemics of the human society has been the constant worry. UNAIDS projection shows the existence of millions of AIDS patients at the end of 2008 with high mortality rate and fresh infections in coming days. All the FDA approved drugs are getting resistant against the clever HIV because of its frequent mutations. Hence there is an urgent need of developing new drugs with greater potential. AIDS is mainly due to the infection of retrovirus HIV. It infects the CD4+ cells of immune system thereby reducing its number to reach an immune deficiency condition. The viral life cycle is controlled by the activities of its essential proteins like gp41, gp120, HIV-RT, HIV-IN, and HIV-pr. Each protein has substantial role on the viral life cycle to be continued. The present thesis focuses on the protein HIV-pr, which is important for the cleavage of Gag and Gag-Pol polyproteins to form the smaller structural and functional proteins. Structurally, the homodimeric aspartyl protease has 198 residues in both chains with the catalytic aspartate in the dimer interface. Glycine rich flexible flaps cap the active site of HIV-pr, which controls the accession of ligands and size of the active site. The conformation of the protein plays a pivotal role in ligand binding and the catalytic process, which is affected by the rapid point mutations and various physiological parameters.

In the present thesis, the conformational dynamics of HIV-pr is being studied in an atomistic detail by MD simulations. Both biological and technical aspects of the protein conformation and dynamics have been explored. With regards to the biological aspect, initially the effect of I47V mutation on the dynamics of HIV-pr and its effect on ligand (JE-2147) binding were investigated and was observed to have both direct and indirect effects on drug resistance. Also it gave the clues for increased flexibility of flaps in mutant compared to WT. We suggest the positioning of a larger group to the ligand at the P2’ position that may decrease the mutation assisted flexibility in flaps and increase the binding properties between the ligand-receptor to diminish the resistance. Furthermore, we checked the protein conformation and

dynamics under high pressure compared to normal condition. It was observed that protein under high pressure validate the general decrease in the protein's structural degrees of freedom and pressure plays a crucial role in reducing the structural variability in proteins. Coming to the technical aspects, we studied the effects of polarization and the difference in force fields on the protein conformation. It was found that, polarizations of force fields influence both the global and specific local motions of protein and solvent thereby increasing the rigidity in proteins. Also the water movements around different types of residues are marked to be different and are high for charged residues. Comparative study of the HIV-pr dynamics by multiple force fields shed light on the difficulty in modeling dynamics of proteins with flexible binding site and *in silico* drug design against flexible receptors. The complex dynamics of HIV-pr can be sensitive enough to the force field difference. Hence a careful examination with different simulation parameters is required to conclude regarding the biological functions drawn from MD simulation studies.

Altogether, these studies indicate that conformational dynamics of HIV-pr is sensitive enough to the simulation setups and force fields difference. Also mutation in some specific region has both direct and indirect effects on the conformation and dynamics of HIV-pr. The outcome of the thesis has noteworthy applications in understanding the molecular basis of drug resistance and designing new effective HIV-pr inhibitors.

TH-772_04610604

TABLE OF CONTENTS

CONTENTS

CONTENTS

LIST OF FIGURES

LIST OF FIGURES

LIST OF FIGURES

TH-772_04610604

LIST OF TABLES

TH-772_04610604

ABBREVIATIONS

μs	microsecond
AIDS	Acquired Immuno Deficiency Syndrome
AMBER	Assisted Molecular Building Energy Refinement
BD	Brownian Dynamics
BPTI	Bovine Pancreatic Trypsin Inhibitor
CA	Capsid protein
CCR5	Chemokine (C-C motif) receptor 5
CXCR4	CXC chemokine receptors 4
DNA	Deoxyribo nucleic acid
DSSP	Dictionary of Secondary Structure for Proteins
FDA	Food and Drug Administration
FFT	Fast Fourier Transforms
H-bond	Hydrogen bond
HIV	Human Immunodeficiency Virus
HIV-IN	HIV-1 Integrase
HIV-pr	HIV-1 Protease
HIV-RT	HIV-1 Reverse Transcriptase
MA	Matrix protein
MC	Monte Carlo
MD	Molecular Dynamics
ms	millisecond
NC	Nucleocapsid protein
NMR	Nuclear Magnetic Resonance
ns	nanosecond
PBC	Periodic Boundary Condition
PDB	Protein Data Bank
PME	Particle-Mesh Ewald
ps	picosecond

TH-772_04610604

ABBREVIATIONS

RDF	Radial Distribution Function
Rg	Radius of Gyration
RMSD	Root Mean Square Deviation
RMSF	Root Mean Square Fluctuation
RNA	Ribo Nucleic Acid
SAS	Solvent Accessible Surface
SPDBV	Swiss PDB Viewer
T_H-cell	Helper T cell
TIP3P	Transferable Intermolecular Potential 3-Point
VDW	van der Waals
VMD	Visual Molecular Dynamics
VPR	Viral Protein R
WT	Wild Type

TH-772_04610604

CHAPTER 1

Introduction

TH-772_04610604

1. Introduction:

The conformational dynamics of HIV-1 protease (HIV-pr) emphasizing on the drug resistance phenomenon is being studied in an atomistic detail in the present thesis along with a presentation of brief review of current research status in this area. The structural and mechanistic aspects of HIV-pr highlighting their respective importance in designing new effective HIV-pr inhibitors are also studied. In the recent years, the trend of drug resistance of HIV-pr has fascinated a great deal of attention. Over time, experimental and synthetic tools leads to the development of drugs against HIV-pr, but soon after, the drugs became resistant susceptible due to the rapid mutations. The presence of the drugs places strong selective pressure on the virus to evolve a resistant form of the proteins concerning the drugs applied. In the recent past, collections of *in silico* as well as *in vitro* studies were carried out to overwhelm the problem of drug resistance. However, a number of reports in literature have appeared, indicating a steady increase in the importance of the phenomenon with the increase in drug abuses. Food and Drug Administration (FDA) approved all anti-HIV-pr drugs, except few, are going in to vain by the clever HIV-pr enzyme by its conformational changes through frequent mutations.

1.1: AIDS and HIV: a brief outline

One of the most challenging problems to human health in 21st century is the evolution of AIDS (Acquired Immuno Deficiency Syndrome) through Human Immunodeficiency Virus (HIV) infection. It is estimated that 33 million people worldwide were living with HIV at the end of 2008. An estimated 2.7 million became newly infected with HIV and an estimated 2.0 million lost their lives to AIDS in 2007 [www.unaids.org]. UNAIDS projections indicate that an additional 45 million people will become affected in next decade unless the world succeeds in getting a protective. A widespread approach is required to prevent the current infections and the additional infections in the coming days. For an efficient treatment against AIDS, it is mandatory to understand how the virus works and which stages of its life cycle are the best targets for effective drug development. At the onset of the epidemic in the early 1980s, no existing drug was known to be useful against AIDS and completely new pharmaceutical agents had to be

TH-772_04610604

created. The rapid progress in understanding of the structure and life cycle of the HIV then led to exceptional development of drugs targeted to a variety of viral proteins.

1.1.1: HIV structure and life cycle:

A detailed explanation of the virus is shown in Figure 1.1 and a detailed description of the HIV life cycle is shown in Figure 1.2.

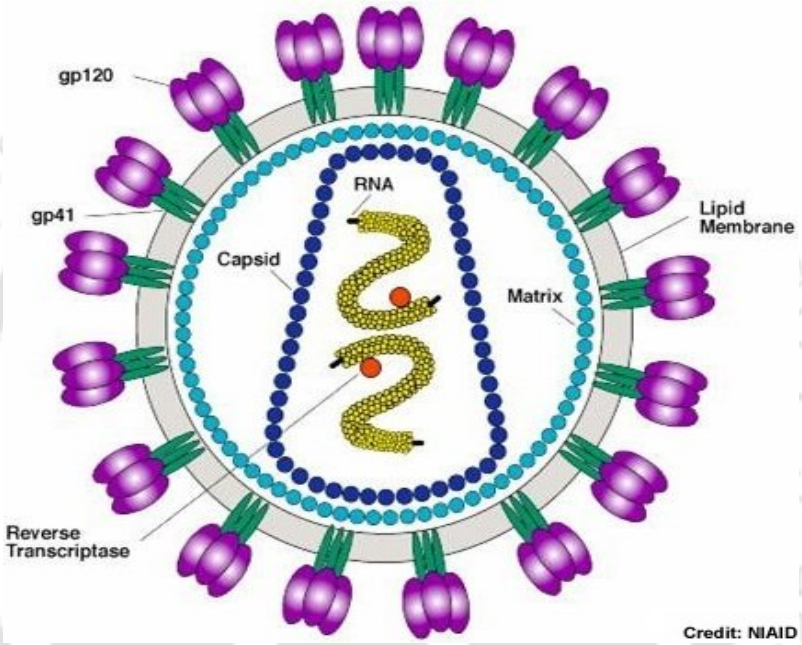


Figure 1.1: *A schematic structure of a Human Immunodeficiency Virus type 1(HIV-1). (Adapted from NIAID)*

HIV-1 is a lentivirus from the *Retroviridae* family, where the genetic information is stored on two single strands of Ribo nucleic acid (RNA) as compared to the other organisms which use Deoxyribo nucleic acid (DNA). It is a 100 to 120 µm enveloped virion [Knipe *et al.* 2001]. The outer covering of the virus is made up of a lipid bilayer membrane, where the two sets of antireceptors, gp120 and gp41 are intertwined. The RNA strands, enzymes like reverse transcriptase (RT), integrase (IN) and protease (PR) are protected by a shell of proteins called nucleocapsid that is engrossed in the cytoplasm.

TH-772_04610604

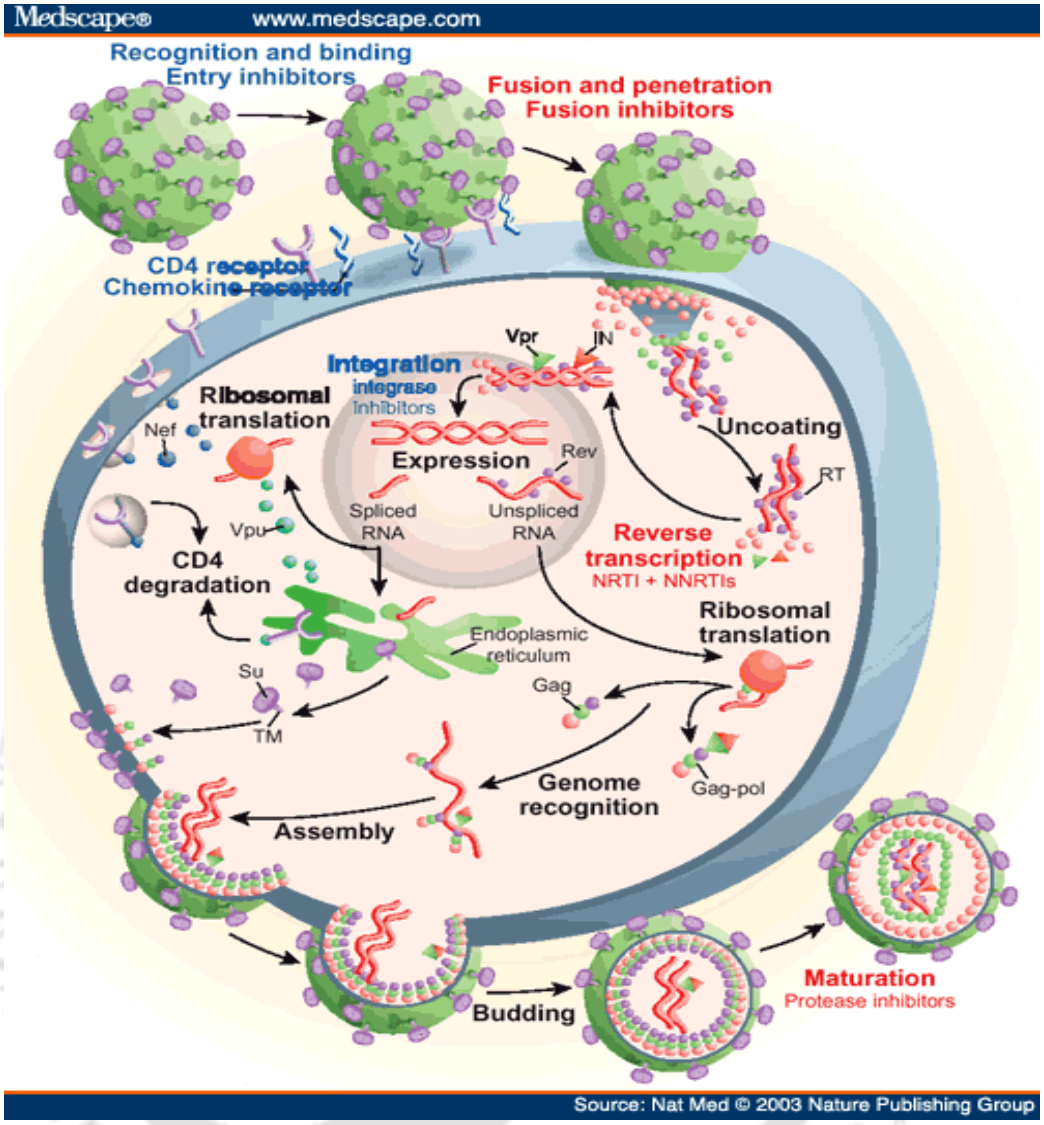


Figure 1.2: *A schematic representation of the HIV-life cycle. (Adapted from a review "Twenty years of therapy for HIV-1 infection. R. J. Pomerantz and D. L. Horn. Nature Medicine. Volume 9, Issue 7, July 2003, Pages 867-873)*

Viruses obligately need living cells for their replication. HIV usually targets the macrophages, dendritic cells and CD4+ cells like helper T-cells (T_H-cell). The T_H-cells are a crucial component of the immune system as they protect the body against many pathogenic infections. HIV's life cycle directly or indirectly causes a drastic fall in the number of T_H-cells in the body. In absence of enough T_H-cells, the body's immune system fails to defend itself against the external and internal infections.

TH-772_04610604

As HIV comes in contact with a T_H-cell, the antireceptors, gp41 and gp120 on its envelope attach themselves to the host cell receptors, CD4 and beta-chemokine (either CCR5 or CXCR4). Following the attachment, HIV penetration into the host cell occurs facilitating the injection of its genetic material. As the virus penetrates the cell membrane, the nucleocapsid embedding the viral genetic material disintegrates and the RNA is released into the cytoplasm. With help of the viral enzyme reverse transcriptase (HIV-RT), RNA is converted into DNA by the process of reverse transcription. Then the DNA is transported to the cell nucleus. Although the exact mechanism is still to be revealed, the viral protein R (VPR) is suspected to be the carrier of the preintegration complex to the nucleus. Then the enzyme HIV integrase (HIV-IN) fuses the viral DNA with the host's DNA by the process called integration. This integrated form of the virus is termed "provirus". When the T_H-cell is activated, the latent provirus hijacks the cellular machinery to produce the required viral proteins. DNA from the cell nucleus is transcribed to viral mRNA, which acts as the template for viral protein synthesis. This process is referred as translation. From the viral DNA (gag and gag-pol genes), two strands of mRNAs are constructed and transported out of the nucleus. One strand is translated into HIV-pr, RT, IN, and other structural proteins while the other strand becomes the genetic material for the progenies. HIV forms a large polypeptide that is cleaved by the HIV-pr to form small structural and functional units. These cleavage reactions occur in a tightly regulated order and with particular specificity [Pettit *et al.* 1998; Pettit *et al.* 2005], that is necessary for viral maturation and for subsequent infection, making the viral protease a superb target for antiretroviral therapy. The envelope proteins incorporate the core proteins, enzymes and RNA which assemble to form a new virion. The newly formed virus then pinches off the cell and starts to bud. Following assembly and budding, it then enters into circulation and infects a new T_H-cell to continue the cycle.

During HIV's life cycle, the T_H-cell is eventually killed. The low content of TH-cell makes the immune system weak and the body falls prone to a host of infections. Nevertheless, much of HIV's life cycle is still under mystery. Advanced research is required for proper understanding of the HIV infection mechanism. More research will

TH-772_04610604

enable scientists to win over HIV into giving up more secrets of how it survives and spreads in the body. In turn, it will also allow for designing of new drugs and vaccines against AIDS. Theoretically, it is possible to arrest each step of the HIV replication process through structure based drug designing. The high specificity of HIV-pr makes itself the key target of a class of drugs used for HIV treatment.

Currently the four HIV proteins, namely HIV-gp41, RT, IN and HIV-pr are the obvious targets for drug discovery. Of these four targets, HIV-pr acts at a later stage of infection by cleaving viral Gag and Gag-Pol polyproteins to yield mature infectious virions for continuation of the viral life cycle [Yong *et al.* 1998]. The knowledge of incomplete processing of these virions produces immature, non-infectious viral particles lead to intense research to achieve HIV-pr inhibitors [Swain *et al.* 1990, Thaisrivongs *et al.* 1996, Wang *et al.* 2000, Perryman *et al.* 2004]. A major challenge for designing drugs against HIV-pr requires a detailed structural and dynamical characteristic of HIV-pr. This enzyme was subject of theoretical [Collins *et al.* 1995, Chen *et al.* 1995, Tropsha *et al.* 1992, Shaffer *et al.* 1998, Weber *et al.* 1989, Hodge *et al.* 1998, Trylska *et al.* 1999, Harte *et al.* 1993, Geller *et al.* 1997, Holloway *et al.* 1995, Rao *et al.* 1992, Luo *et al.* 1998, Tawa *et al.* 1998, Okimoto *et al.* 2000, Harrison *et al.* 1994, Chatfield *et al.* 1995, Liu *et al.* 1996, Silva *et al.*1996, Okimoto *et al.*1999, Lee *et al.* 1996, Venturini *et al.* 1998] and experimental [Hong *et al.* 1998, Smith *et al.* 1996, Katoh *et al.* 1999, Baldwin *et al.* 1995, Ridky *et al.* 1998, Miller *et al.* 1989, Erickson *et al.* 1990, Lee *et al.* 1999, Poorman *et al.* 1991, Bagossi *et al.* 1996, Rose *et al.* 1996, Loeb *et al.* 1989, Polgár *et al.* 1994, Hyland *et al.* 1991a, Hyland *et al.* 1991b, Ishima *et al.* 1998, Keinnan *et al.* 2000, Wang *et al.* 1996a, Wang *et al.* 1996b, Freedberg *et al.* 1998, Yamazaki *et al.* 1994a, Yamazaki *et al.* 1994b, Grzesiek *et al.* 1994, Baca *et al.* 1993, Todd *et al.* 1998, Todd *et al.* 1999, Velasquez *et al.* 2000, Wlodawer *et al.*1993, Wlodawer *et al.* 1998] studies aimed at elucidating structural and functional aspects of its biochemistry. Also, in the recent past a number of reports in literature have appeared, indicating the importance of the structure, function and dynamics of HIV-pr. The structure of the HIV-pr gives many interesting directions in uncovering the mystery of designing effective drugs against it. Due to its crucial importance as target for anti-AIDS therapies, the free enzyme and the

TH-772_04610604

complexes with inhibitors have been the subject of a large number of X-ray crystallography [Baldwin *et al.* 1995, Miller *et al. Science* 1989b, Erickson *et al.* 1990, Rose *et al.* 1996, Wlodawer *et al.* 1989, Miller *et al. Nature* 1989a, Navia *et al.* 1989] and nuclear magnetic resonance (NMR) [Wang *et al.* 1996a & 1996b, Freedberg *et al.*1998, Yamazaki *et al.* 1994, Grzesiek *et al.* 1994] studies.

1.2. Drug resistance of HIV-1 Protease:

As mentioned earlier, HIV-pr is one of the most crucial enzymes in the life cycle of HIV and is known as the foremost causal agent for AIDS. The HIV encodes several proteins which are essential for the viral maturation process on a long chain of peptide called "polyprotein" with several proteins strung together. The polyproteins in its long form is not active unless it is being cleaved into several pieces by the HIV-pr. HIV-pr exclusively recognizes and cleaves at least nine non-homologous, asymmetric sites within the Gag and Gag-Pol polyprotein. Processing of these polyproteins occurs in a regulated order and at specific rates that are necessary for viral maturation. There is a time frame for the HIV-pr to cleave the polyprotein and to activate the proteins, before the polyprotein undergoes degradation by a process called autoproteolysis. HIV-pr recognizes its substrate sites based on structural homology [Jeyabalan *et al.* 2000; Jeyabalan *et al.* 2002] and helps the virus to produce the structural and functional units for the viral assembly and subsequent activity. Inhibition of the viral maturation process by the activity of HIV-pr is therefore very critical, with the consequence that lead to the immature stage of the virus. The importance of the HIV-pr is thus made itself the major target by antiretroviral therapies for curbing AIDS.

In the early days, there was a great deal of optimism that, computer-aided drug design with several other techniques would reform the way in which drugs could be developed against AIDS. Over time experimental and synthetic tools lead to the development of drugs, but soon after, the drugs became resistant or less efficient due to the rapid mutations of the proteins involved. The presence of the drugs places strong selective pressure on the virus to evolve a resistant form of the proteins with respect to the drugs

TH-772_04610604

applied. HIV-pr, the retroviral aspartyl protease falls under this category and exhibits the mutant resistant form by the interaction with drug molecules. This phenomenon is known as the drug resistance of HIV-pr. It is estimated that almost all the FDA (Food and Drug Administration) approved drugs, except few are resistant to HIV-pr, with no matter of adherence and combinations. A list of the FDA approved inhibitors for HIV-pr with their chemical structures is shown in the Figure 1.3:

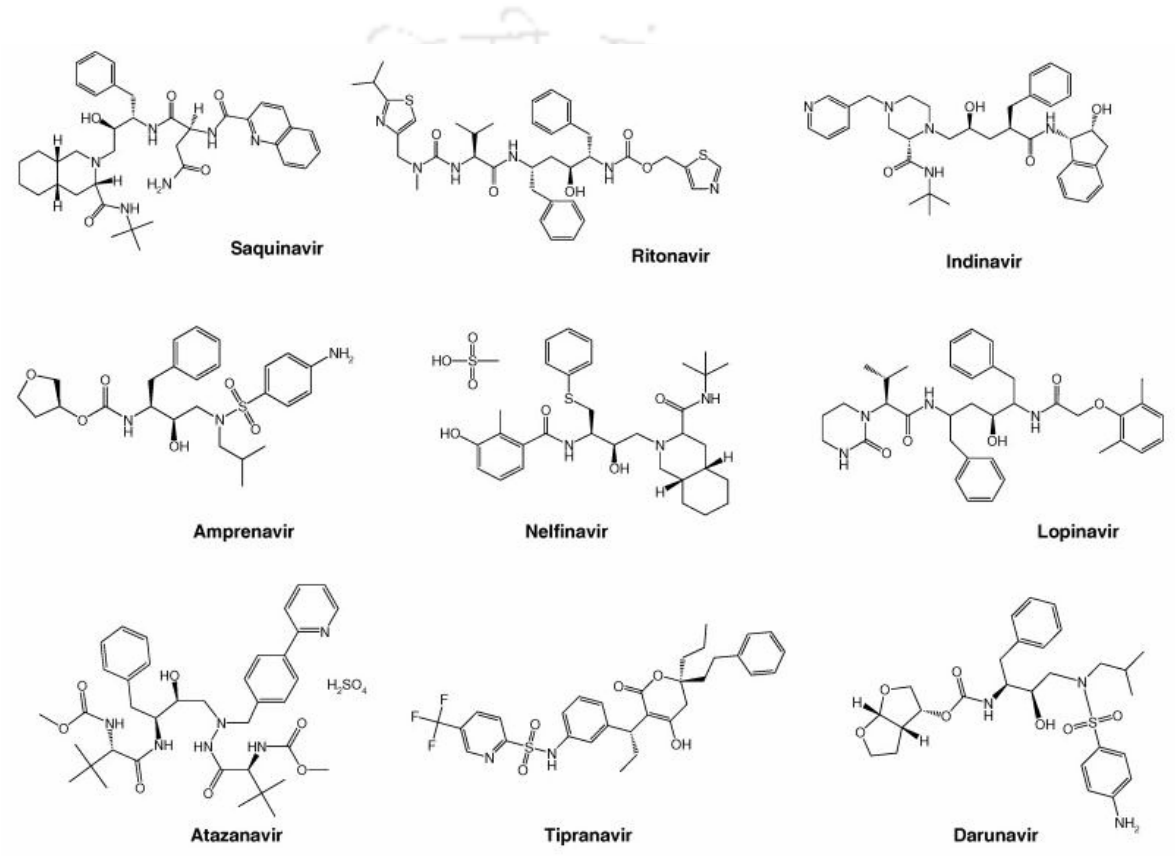


Figure 1.3: *Inhibitors of HIV-1 protease that are approved by FDA to treat AIDS and AIDS related malignancies. (Adapted from a review "HIV-1 protease inhibitors: effects on HIV-2 replication and resistance. L. Menendez-Arias and Jozsef Tozser. Trends in Pharmacological Sciences Volume 29, Issue 1, January 2008, Pages 42-49.)*

1.3. Major Objectives of the work:

The general objective for this work was to understand both the biological and technical aspects (i.e. the dependence of the dynamics on the details of simulation setups) of

TH-772_04610604

conformation and dynamics of HIV-pr in understanding the molecular basis of drug resistance. More specifically the objectives were:

1) To provide significant information for designing effective new inhibitors against I47V mutants by modifying the shape of the experimental inhibitor JE-2147.

2) To determine if any generalization related to drug resistance of HIV-pr may be made by comparing the results of I47V mutation study with previously published works.

3) To understand the conformation and dynamics of HIV-pr under high pressure and its importance in regulation of size of the active site.

4) To comprehend the polarization effect on protein conformation highlighting on HIV-pr's flexibility/rigidity behavior.

5) To understand the effect of different force fields on the flexibility of HIV-pr and its significance in drug design against HIV-pr *in silico*.

1.4. Protein structure and function:

Proteins, the most versatile biological macromolecules are the building blocks of cells and organs of living organisms. It plays an essential role in cellular structure, function and is ultimately the mechanisms through which many diseases have their effect. The morphology, function and activity of a cell are all dependent on the proteins it expresses. In other words, proteins affect almost all property that exemplifies a living organism. Proteins act as messenger between cells and organs of higher organisms in the form of peptide hormones. They also play the role of electron carriers in crucial physiological functions such as photosynthesis and respiration. Multimeric proteins organize themselves so as to facilitate selective passage through the membranes. A type of proteins called as immunoglobulins act as sentinels of our body against pathogens. Proteins also have pivotal role in regulating the transcription of genes and thus, determining the pattern of gene expression. Structural proteins are known to constitute the scaffold of the cells. Moreover, the key to appreciating how different proteins function in these different ways is based on a detailed understanding of protein structures.

Despite these varied biological functions, proteins are virtually a homogeneous class of molecules. They are composed of twenty standard amino acid residues, which differ from

TH-772_04610604

each other by their side chain groups. Main chains of amino acid residues are made up of a nitrogen atom, a carbon atom named as alpha carbon (Cα) to which the side chain atom is covalently bonded, and a carboxyl group. The dihedral angle of the bond between a main chain nitrogen atom and the Cα atom is called the angle phi (φ), and the dihedral angle of the bond between the Cα atom and the carboxyl carbon atom is called the psi (ψ) angle. Statistical studies reveal that the angles φ and ψ are distributed in certain regions in the whole φ and ψ plane, which is called the Ramachandran plot [Ramachandran *et al.* 1968]. Amino acid residues are connected to form a chain through the covalent bonds between the carboxyl carbon atom of a residue and the nitrogen atom of the next residue to form a peptide bond.

The structural stability of the proteins comes from various interactions among molecules in the proteins through covalent and non-covalent interactions. Covalent interactions involve the peptide bond among amino acids and non-covalent interactions involve the ionic, hydrophobic and van der Waals interactions among the atoms of amino acid residues. Besides these interactions, hydrogen bonding interaction plays an important role in protein structure determination.

1.4.1. Hydrogen Bonding:

A hydrogen bond (H-bond) is the attractive force between the hydrogen attached to an electronegative atom of one molecule and an electronegative atom of a different molecule (a polar atom and another nearby polar atom). The polar atom that is covalently bonded to the hydrogen atom is called the H-bond donor and the polar atom that is close but not covalently bonded to the hydrogen atom is called the H-bond acceptor. The major part of the H-bond is an electrostatic interaction between the dipole of the covalent bond to the hydrogen atom, in which the hydrogen atom has a partial positive charge (δ^+), and on the other electronegative atom has a partial negative charge (δ^-) as shown in the Figure1.4 for the H-bonding in water molecules. The H-atom is unique in being able to interact strongly with one electronegative atom while being covalently attached to another.

TH-772_04610604

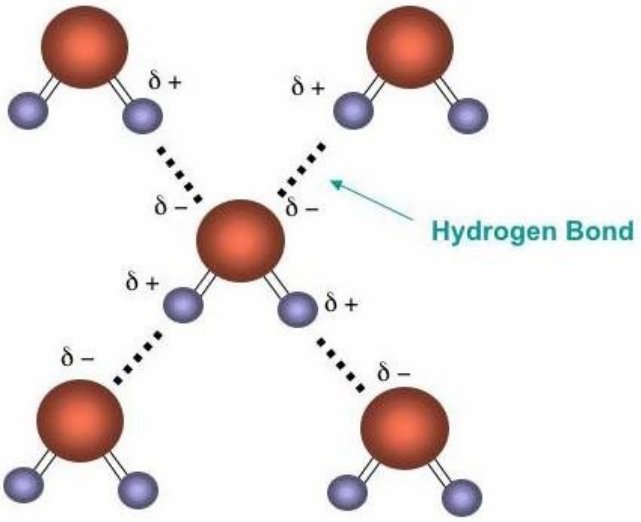


Figure 1.4: *Hydrogen bonding in water molecules*.

As such H-bond is energetically poor, but due to its partial electrostatic and covalent properties, it assumes collinear conformation, thus becomes thermodynamically favorable [Taylor and Kennard 1984]. The length and strength of the H-bond are dependent on the electronegativities of the acceptor and donor atoms. More the electron acquiring tendency of the acceptors and donors, lesser the length and stable is the H-bond. By strengths H-bonds are classified as weaker and longer (as compare to covalent bonding), because they have a tendency to form and broken easily under normal biological conditions. The bond length of a H-bond is 1.97 Å as compare to the normal covalent bond length of 0.96 Å. Although H-bonds are weaker and longer than the covalent bonding, there presence in large amounts gives a solid structure to the biomolecules.

Stereochemistry of the H-bond is helpful in the interpretation of protein structure, function and stability which is governed by (i) the electronic configuration of acceptor atoms, (ii) the steric accessibility of donor atoms and (iii) the conformation of amino acid side-chains [Ippolito *et al.* 1990]. Mutational studies on proteins [Shirley *et al.* 1992, Byrne *et al.* 1995, Yu *et al.* 1995, Myers and Pace 1996, Yamagata *et al.* 1998] have also been used to estimate the hydrogen bonding contribution to protein structure and stability. In these studies, out of the two polar amino acids involved in the hydrogen bond

TH-772_04610604

formation, one is replaced by a non polar amino acid. Then the variation in the stability between the wild and mutant protein is compared to estimate the contribution of hydrogen bonding towards it. Myers and Pace [Myers and Pace 1996] used experimental values of such free energy differences between the wild and mutant protein and concluded that a hydrogen bond contributes −1 to −2 kcal/mol towards protein stability. However, Honig and Yang [Honig *et al.* 1995] contradicted the above finding, claiming that the loss of one partner in the hydrogen bond is not the true indicator of the contribution of hydrogen bond to protein structure, since it leaves an unsatisfied polar group which is buried and not exposed to water, associating itself with desolvation penalty. Assessment of the contribution of hydrogen bonds to protein structure and stability are further hampered by factors as hydrophobicity, packing and conformational entropy [Byrne *et al.* 1995]. However, hydrogen bonds play a key role in determining the structure of proteins is a well documented.

1.4.2. Ionic and van der Waals Interactions

The most fundamental non-covalent interaction is that between electrostatic charges which acts over some distance of separation. The quantitative expression for the effect of the variables like charge, charge and distance on electric force is known as Coulomb's law. Coulomb's law states that the energy of the electrostatic interactions between two charged atoms is directly proportional to the product of the quantity of charge on the two atoms and inversely proportional to the distance between two atoms. In equation form, Coulomb's law can be stated from the following figure as:

Figure 1.5: *Ionic interactions between two atoms Q_1 and Q_2 at a distance R.*

If the charges on the two atoms are opposite in sign, the energy decreases as they attract each other leading to a favorable interaction. On the contrary, if two charges are similar,

TH-772_04610604

they repel each other. At large distances, there is a negligible energy of attraction between the two atoms, but as they are brought closer together, they are attracted to one another. On the other hand, the atoms are actually repelled at small distances due to the positively charged atom's nuclei.

However Coulomb's law is applicable only for two point charges in vacuum. For biological systems like protein-aqueous solution, the electrostatic interaction is changed by other interactions and is decreased by the dielectric constant**.** So the above expression for biological systems can be written like follows:

$$\Delta E = K \frac{Q_1 Q_2}{D R_{12}} \quad \text{.... ..... (1.1)}$$

Where ΔE is the change in potential energy, K is the constant, Q_1 and Q_2 are the two charged atoms at a distance R_{12} and dielectric constant D. At short distances, the charge of an atom is separated between the nucleus and the scatter electron cloud, so it cannot be treated as a point charge. In proteins the charge separation problem is even more. The charges of the ionized groups of the side chains of amino acids like Lysine (Lys), Arginine (Arg), Histidine (His), Aspartate (Asp) and Glutamate (Glu) residues and the α-amino & α-carboxyl groups are distributed over two or more hydrogen or oxygen atoms [Matthew 1985, Warshall *et al.* 1984]. Also in proteins there are salt bridges, generally formed by the interactions between very closely placed oppositely charged groups with the combined effort from electrostatic interactions and H-bonding [Bray *et al.* 1984].

The van der Waals forces are very weak attractions (or repulsions), which occur between atoms at close range. When two uncharged atoms are brought very close together their surrounding electron clouds influence each other. Random variations in the positions of the electron around one nucleus may create a transient electric dipole, which induces an opposite transient electric dipole in the nearby atom. The two dipole weakly attract each other bringing the two nuclei closer. These weak and close-range attractions are called van der Waals interactions, which varies as the sixth power of the distance between the

TH-772_04610604

atoms. As the two nuclei draw close, their electron cloud begins to repel each other because the electrons on the different molecules cannot be in the same part of space at the same time. The repulsive energy increases with the inverse of the 12^{th} power of the distance between the centers of the two atoms [Lennard-Jones 1931]. At the point where the van der Waals attraction exactly balances their repulsive force, the nuclei are said to be in van der Waals contact. Van der Waals interactions are commonly characterized by an energy potential as a function of distance that comprises both the attractive force and the repulsion at close range. The best known of these is the Lennard -Jones potential of the form:

$$V(r) = 4\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right] \qquad \ldots\ldots\ldots\ldots\ldots\ldots (1.2)$$

Where ε is the depth of the potential well and σ is the (finite) distance at which the inter-atomic potential is zero and "r" is the distance between the atoms.

1.4.3. Hydrophobic Interactions

Hydrophobic interactions are very important in biological systems. Water is a very poor solvent for non-polar molecules as compare to other organic solvents [Creighton 1993]. The interactions of water molecules are not favorable with non-polar surfaces. Non-polar molecules try to avoid the most important interactions, i.e. H-bonding in the aqueous solution. This virtual absence of interactions between non-polar molecules and water directs the favorable interactions between the non-polar groups themselves. So, non-polar molecules greatly prefer non-polar environment. This phenomenon of preference of non-polar atoms of nonaqueous environment is called the hydrophobic interaction [Kauzmann 1959, Ben-Naim 1978].

TH-772_04610604

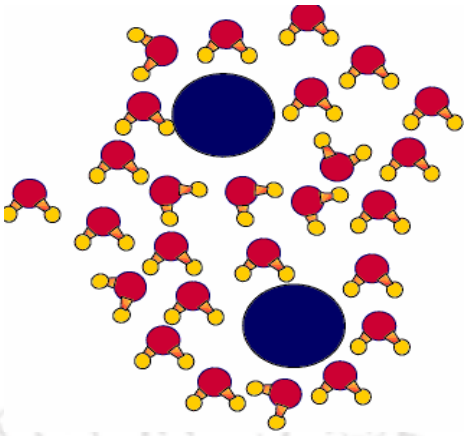


Figure 1.6: *A schematic diagram showing the Hydrophobic Interactions*.

Figure 1.6 shows the schematic diagram of hydrophobic interactions, where the two globular molecules are encapsulated with a hydrophobic zone. For which the hydrophobic regions of a protein will preferentially locate away from the surface of the molecule [Privalov *et al.* 1989]. The hydrophobic interaction has the uncommon property of decreasing in magnitude at lower temperatures to form the H-bonded clathrate-like structures around the nonpolar molecule [Baldwin 1986].

To explain the compactness of protein molecules, hydrophobic interactions have become especially popular among protein chemists. There are indeed many non-polar groups in protein macromolecules and many of them are clustered together as if avoiding contact with water. Therefore it was appealing to regard a globular protein in aqueous solution as an oil drop in water stabilized by hydrophobic interactions. This point of view was supported by the finding that the unfolding of a compact protein structure, i.e. the process which can be regarded as a transfer of internal non-polar groups of a protein to water, proceeds with a rather small enthalpy at room temperature [Hermans *et al.* 1965]**,** suggesting that the entropy factor might play some role in the stabilization of the compact state of a protein.

Role of hydrophobic effect in maintenance of protein structure and stability has also been studied by site-directed mutagenesis in the amino acid sequence of the protein [Yutani *et al.* 1987, Kellis *et al.* 1988, Matsumura *et al.* 1988, Kellis *et al.* 1989, Shortle *et al.* 1990,

TH-772_04610604

Sandberg *et al.* 1991, Eriksson *et al.* 1992, Serrano *et al.* 1992]. In these experiments, a hydrophobic residue is substituted with another hydrophobic residue. The free energy of unfolding for the wild-type and the mutant is then compared to find the effect of the substitution on protein stability. The results obtained from point mutation are higher and differ substantially from solvent transfer experiments. Factors accountable for this variation in values obtained from solvent transfer and mutation studies may be due to cavity formation, modification in other parts of the protein or the resemblance of the protein more to a solid than a liquid [Richards 1977]. These resultants of mutation render the comparison of values from solvent transfer experiments and mutational studies complicated. Thus, using protein engineering for the quantification of the hydrophobic effect is ridden with its own pitfalls. Nonetheless, it is universally acknowledged that, hydrophobic effect plays a pivotal role in protein structure and folding.

1.5. Protein dynamics:

Without motion there cannot be life. This is particularly true for proteins, as exemplified in a work by Petsko and co-workers published in *Nature* [Rasmussen *et al.* 1992]. Almost all biological processes critically depend on protein activities. Though many of these processes dependent on the protein structure and function, protein dynamics also have significant role in these processes. Practically all these processes that involve motion find their origin in protein dynamics. The mechanism of muscle contraction is an ideal instance to justify the above statement. Muscle contraction occurs by the action of actin and myosin protein cross-bridge cycling [Geeves *et al.* 1999, Goldman 1998]. Function of kinesin and F_0F_1-ATPase also originate from protein dynamics [Gelfand *et al.* 1991, Block. 1998, Suzuki *et al.* 2003]. Dynamics also plays very crucial role in a host of proteins. For example, flexibility of structural conformation is the basic requirement for proper functionality of transport proteins, signal transducers, immunoglobulins, and enzymes [Stryer, 1998]. In many enzymes, conformational changes are essential for trapping the substrate in its interior and avoiding release. Also, proper fitting of the substrate to the active site is ensured by dynamic changes of enzyme structure e.g. HIV-1 protease and in lysozyme [Berendsen *et al.* 2000, Hayward *et al.* 1998, Rose *et al.* 1998]. Immunoglobulin flexibility is vital for antigen binding and adaptation to diverse

TH-772_04610604

molecular shapes and sizes [Davies *et al.* 1990, Davies *et al.* 1993]. Further, G-proteins acting as signal transducers in second messenger cascade too modulate its structure on binding to hormones [Hamm 1998, Helmerich *et al.* 1996, Skiba *et al.* 1998]. Allosteric proteins as haemoglobins too undergo conformational shift for its optimum functionality [Changeux. *et al.* 1993, Monod *et al.* 1965]. Substrate binding to one subunit of these multimeric proteins triggers a conformational change that alters the substrate affinity of the other subunits, thereby sharpening the switching response of these proteins. The conformational changes involved range from very subtle, local changes, as in the case of e.g. myoglobin, to global conformational changes, involving motions of significant amplitude for large parts of a protein (e.g. hemoglobin) [Nelson and Cox 2000]. Role of protein dynamics is not restricted to its diverse functionality only, but it also regulates the protein folding process. Also protein dynamics is vital for various biological functions like signal transduction, immune system, cellular transport and several enzymatic activities, which ranges from picosecond (ps) to millisecond (ms) time scale [Cavanagh and Venters 2001]. Conversely, local motions in proteins, such as side chain and backbone fluctuations take place at picosecond and nanosecond (ns) [Kungl *et al.*, 1998] while protein folding, allosteric modulation and enzymatic catalysis takes place at microsecond (μs) to millisecond (ms) time scale [Roder and Shastry 1999, Eisenmesser *et al.* 2002, Daniel *et al.*, 2003]. From NMR relaxation technique, it is shown that during catalysis of human cyclophilin A, internal dynamics of enzyme takes place on hundred of microsecond time scale and the rates of conformational dynamics of the enzyme strongly correlate with the microscopic rates of substrate turnover [Eisenmesser *et al.* 2002]. The overall dynamics of enzymes are highly coordinated during catalysis, which is indicated from the involvement of global movement of enzyme along with segmental movement of active site [Eisenmesser *et al.* 2005]. So it confirms that the intrinsic dynamics of an enzyme triggers catalysis.

Protein assumes its native state through step-wise dynamic transitions. X-ray crystallography studies throw light on the structure of the protein but not on its dynamics. Nuclear Magnetic Resonance (NMR), used for studying internal mobility of proteins, has provided valuable information on the picosecond- microsecond time scales of motion in

TH-772_04610604

HIV-pr [Nicholson *et al.* 1995; Ishima *et al.* 1999] and how ligand binding affects these motions [Yamazaki *et al.* 1994a & 1994b; Yamazaki *et al.* 1996; Ishima *et al.* 1999]. However, NMR alone has not been able to provide a detailed mechanism of how the protein changes during ligand binding and till now no experimental techniques have been come up to monitor protein conformational changes at atomic resolution level in nanosecond time scale. Even though, several proteins have been structurally characterized at different intermediate conformations, the time resolution of structural studies improves steadily [Borgstahl *et al.* 1997]. Detailed data on the pathways followed during the modulation is yet to be explored. At the current scenario, computer simulation techniques are the only tool to derive information about protein dynamics at atomic resolution. However recently there are reports on protein dynamics study with combined MD simulations and NMR spin-relaxation methods in the picoseconds to nanosecond time range [Trbovic *et al.* 2008]. Nevertheless, details on the pathways between different known conformations often remain ambiguous.

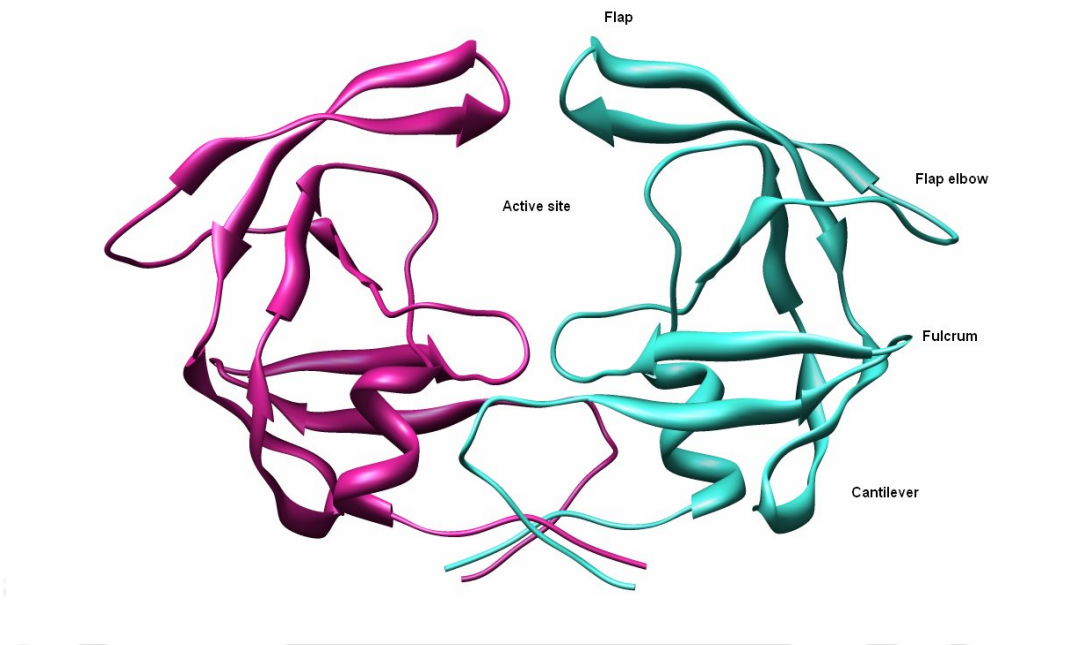
From the dynamics of protein point of view, it is interesting to check the conformation and dynamics of the key protein HIV-pr, which catalytic activities and size of the active site are controlled by its flexible dynamic parts. In order to get the atomic details and the conformational ambiguity related to HIV-pr dynamics, we use computational models/MD simulation tools based on the physical forces that act on each atom in a protein.

1.6. HIV-1 Protease Conformation and dynamics:

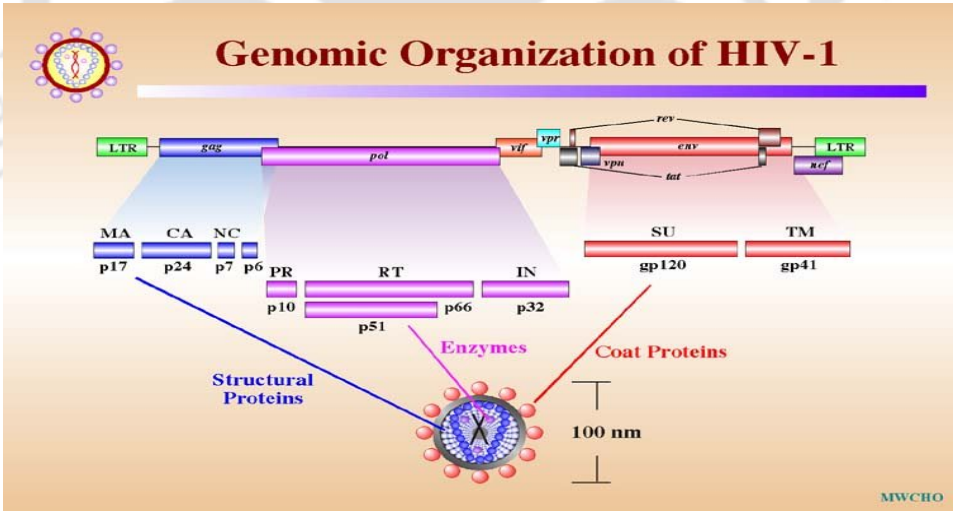
Structurally HIV-pr is a homodimeric aspartyl protease [Fitzgerald and Springer 1991, Hong *et al.* 1998, Baca *et al.* 1993] whose activity is indispensable to the virus [Kohl *et al.* 1988, Seelmeier *et al.* 1988, McQuade *et al.* 1990]. It contains two identical chains of length 99 amino acids each with a molecular weight of ~ 22 Kd. The protease in its monomeric form is inactive and exhibits C2 symmetry in its unliganded state. The residues of HIV-pr are numbered as 1–99 and 100–198 for each monomer (chain A and B, respectively). Flap (residues 43–58 and 142-157), flap elbow (residues 35–42 and 134-141), fulcrum (residues 11–22 and 110-121), cantilever (residues 59–75 and 158-174),

TH-772_04610604

and ligand binding regions, the active site for both the chains-A & B are represented in the Figure 1.7(a) and the HIV-pr cleavage sites are represented in Figure 1.7(b).



(a)



(b)

Figure 1.7: *(a) Structure of HIV-1 protease and (b) the Gag and Pol sequence cleaved at the blue and pink regions shown. (Adapted from MWCHO)*

MA, CA, NC, p6 are the structural proteins that form the viral shell and capsid; RT is the reverse transcriptase and IN is the integrase enzymes. Note that HIV-pr itself is also a

TH-772_04610604

part of the POL polypeptide chain [Debouck *et al.* 1999]. The most important region of HIV-pr is the active site or ligand binding region, which is characterized by the highly conserved *catalytic triad* sequence Asp-Thr- Gly in the dimer interface with contribution of each monomer as Asp25(25')-Thr26(26')-Gly27(27') [Davies 1990]. The Asps are essential to the protease both catalytically and structurally while the Thr & Gly are buried in the active site. The active site is capped by two identical flexible glycine rich ß-hairpins, or flaps that restrict access to the active site. The mobile flap, residues 43-58, in either chains contains three characteristic regions: side chains that extend outward (Pro44, Met46, Phe53, Lys55, Arg57), hydrophobic chains extending inward (Lys43, Lys45, Ile47, Ile54, Val56, Asp58), and a glycine rich region (Gly48, Gly49, Gly51, Gly52) with Ile50 remains at the tip of the turn. The tips of the flaps, notably the glycine rich region, are highly flexible and this is thought to be necessary for substrate binding and product release. The tips curl into the cleft resulting in an open space wide enough for a substrate to enter. Schiffer *et al.* suggested that, upon curling, the electronegative active site is exposed and a neutral or positively charged substrate could potentially be guided into a conformation optimal for binding [Scott *et al.* 2000]. It is already notified that the curling of each monomer does not occur in a symmetric fashion given the symmetric axis. The active site wall region is present just below to the flap region, which contains the P1-loop (residues 78-81) in either chain. However the residues 78-80 do not have direct interactions to substrates or inhibitors. They have an indirect role in ligand binding through their contact with the flap region. X-ray diffraction studies [Spinelli *et al.* 1991, Lapatto *et al.* 1989, Wlodawer *et al.* 1989] showed that in the apo-form of HIV-pr, the flaps are packed onto each other loosely in a *semi-open* conformation, and the overall structure of the protease is less compact compared to the crystal structure of the inhibitor bound protease where the flaps are packed tightly, i.e., closed-structure [Louis *et al.*1998, Jeyabalan *et al.* 2000].

For HIV-pr, there are a large number of published works, both computational and experimental, exploring its structural and mechanistic aspects. Mutations in HIV-pr are of two types, one at or near the active site and the others are far from the active site. Near the active site mutations may reduce the affinity of the enzyme for its substrate and

TH-772_04610604

inhibitors by change in their direct interactions. However, non-active site mutations usually called as compensatory mutations, can act through long-range electrostatic interactions affecting the ligand binding in the active site region [Ohtaka *et al.* 2003]. In accordance to the NMR and computational studies, certain mutations change the conformational dynamics especially the flap dynamics of HIV-pr. This in turn changes the binding affinity of the ligands by alteration of cavity size leading to various drug resistant mutant forms of the protease [Rose *et al.* 1998]. Some mutations may confer both direct and indirect effect [Bandyopadhyay and Meher 2006]. Mutations have been identified in at least 50 positions within the HIV-pr [Gulnik *et al.* 2000, Kozal *et al.* 1996, Shafer *et al.* 1999, Boden *et al.* 1998, Hertogs *et al.* 2000]. Lists of mutations that occur on the HIV-pr backbone are shown in the Figure 1.8. A molecular level understanding of drug resistance requires the knowledge of both direct and indirect effects of mutation.

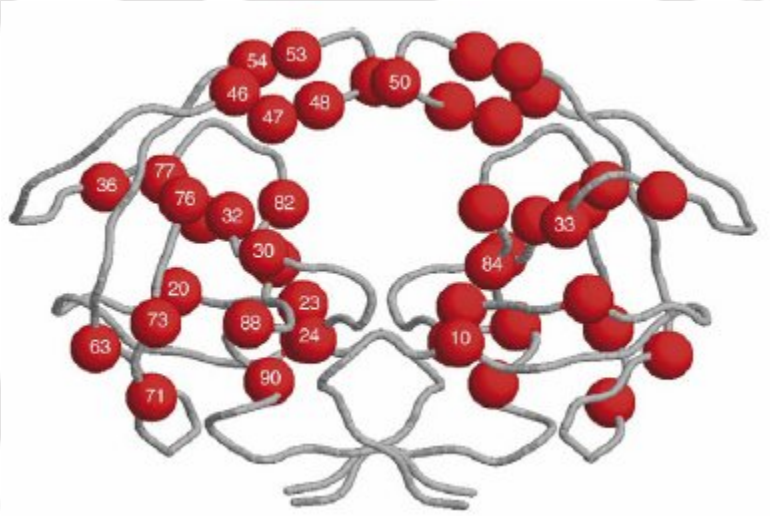


Figure 1.8: *The structural distribution of the most common mutations associated with drug resistance in the HIV-1 protease. Mutations can occur anywhere in the protease structure. Mutations within the binding cavity are very conservative and operate by distorting the shape of the cavity. Conformationally constrained inhibitors have difficulties in adapting to the altered geometry and lose significant binding affinity.*

(Adapted from Overcoming HIV-1 resistance to protease inhibitors. Drug Discovery Today. Disease Mechanisms | Infectious diseases Vol. 3, No. 2 2006)

TH-772_04610604

For every protein/enzyme, its structure and dynamics play crucial role in the functional activities. For HIV-pr, the proper folding of the two monomers to form a catalytically competent dimer and the conformational dynamics especially the flap dynamics has special impact on its catalytic efficiency. From that point of view, it is necessary to gain knowledge about how the protein folds to its optimum function and how the dynamics of the protein help in the catalytic processes. A brief discussion regarding the flap dynamics and the folding dimerization of HIV-pr along with a general review of the works carried out by several groups is presented here.

1.6.1. Flap dynamics:

The flexibility of the flap tips is known as flap dynamics. It opens and closes the flaps determining the cavity size, which changes with several mutations leading to various drug resistant mutant forms of the protease by lowering the affinity for drugs. Understanding the mystery behind flap dynamics is an important step in designing highly potent new drugs against HIV with lesser drug resistance. Thus several earlier computational studies intended at understanding flap dynamics. In recent past, MD has been used to study the dynamics of flaps: how the protease flaps move to allow substrate access to the active site and drug escape from the active site, how is the arrangement of the flaps during liganded and unliganded form and how mutations induce the conformational flexibility of the flap region [Collins *et al.* 1995, Scott *et al.* 2000, Piana *et al.* 2002, Freedberg *et al.* 2002, Perryman *et al.* 2004, Meagher *et al.* 2005, Bandyopadhyay and Meher 2006, Hornak *et al.* 2006, Perryman *et al.* 2006, Toth *et al.* 2006, Lauria *et al.* 2007, Seibold *et al.* 2007, Sadiq *et al.* 2007, Tozzini *et al.* 2007, Singh and Senapati 2008].

There has been a vast major of works performed by many groups in order to understand the drug resistance behavior of HIV-pr in a variety of ways. Many groups has explored on the different mutation positions and their effects on HIV-pr conformation, while others have concentrated on dimerization and folding of HIV-pr. However, there has been inadequate argument regarding the role of different simulation protocols and physiological conditions on the conformational dynamics or flap dynamics of HIV-pr.

TH-772_04610604

1.7. Present Investigation:

The present study focuses on both the biological as well as the technical aspects of HIV-pr conformation. The effects of protein backbone mutations, physiological parameters (like pressure) and simulation protocols (polarization of the system and the force fields used) on the conformational dynamics of HIV-pr have been studied using MD simulations. In order to achieve the above objectives four different works were investigated and evaluated of their potential. Following were the detailed investigations performed in the present thesis:

1.7.1. Mutation effect on HIV-pr conformation:

The major problem associated with the existing drugs against HIV-pr is molecular resistance, which is mainly due to the various mutations of the protein backbone leading to the conformational flexibility of the protein. However, the molecular basis of drug resistance of HIV-pr is not fully understood. JE-2147, an experimental inhibitor of HIV-pr, shows a resistance profile different from that of known drugs. Although it is less susceptible to several common mutations, it is still susceptible to few mutations like I47V which appears to be specific for JE-2147. In the first part of my work, the molecular details of the effect of I47V mutation are investigated using molecular dynamics simulation. Four simulations of apo and complexed proteins in their wild type (WT) and mutant forms have been performed. It is found that the mobility of the side chain of mutant Val47 in chain B of HIV-pr about the inhibitor increases significantly relative to WT Ile47 in chain B. This is due to loss of optimized packing of the inhibitor to the residue 47 in chain B of the mutant when compared with WT enzyme. There also are subtle differences in motion involving residues in the flap region, which are more prominent in the apo form.

1.7.2. Effect of Pressure:

The internal protein dynamics for HIV-pr proved to be vital for ligand binding and determination of the cavity size, which varies with common physiological parameters like temperature, pressure, pH conditions and the protein backbone mutations that cannot be

TH-772_04610604

ruled out. In the final part of my work, the effect of 1 bar (0.987 atm) and 3 Kbar pressure conditions was studied on the conformation and dynamics of HIV-pr *in silico*. Literature searches testify that at high pressure, hydrophobic residues of the proteins are uncovered to the solvent exposed area, which eventually leads to change in dynamics and hence conformation of the protein. Our dynamics studies showed that, although the collective dynamics was constrained under pressure there are exceptions too. A significant perturbation of the protein secondary structures was observed from the secondary structure analysis. It can be seen that turns and bends are favored under high pressure at the expense of α-helices and β-sheets resulting in the reduction of structural variability. Solvent accessible surface (SAS) area of both the low and high pressure simulations showed remarkable differences. It was also observed that with the increase in pressure, the hydrophobic effect is decreased. All these changes in the internal protein dynamics at high pressure conditions are expected to have significant impact on the protein conformation and drug binding affinity to the active site region, which may have a direct/indirect effect on the drug resistance behavior of HIV-pr.

1.7.3. Effect of polarization:

In the second part of my work, the effect of polarization in biomolecular force field is studied by performing MD simulation of HIV-pr using two AMBER force fields, namely ff99 (non-polarizable) and ff02 (polarizable). The results of simulation show that the overall structural fluctuation of HIV-protease is reduced in the polarizable simulation as found from the RMSD values. Comparison with the NMR order parameters with the calculated values show that although some residues are less flexible in the ff02 simulation, the dynamics of two β-hairpins (flaps), the most flexible part of the protein, is relatively insensitive to the effect of polarization. The flap-active site distance, a measure of flap opening, is distinctly more in the non-polarizable simulation. The water count and radial distribution functions are investigated near a representative residue of three types - charged, polar and hydrophobic. Both water count and radial distribution function differ significantly near the charged residue (catalytic Asp25) between the force fields. However, the water movement is similar near the polar (Ser37) and hydrophobic (Ile85)

TH-772_04610604

residues. The results of this investigation show that polarization is likely to influence both global and specific local motions of protein and solvent.

1.7.4. Effect of Simulation protocol and force fields:

In the third part of my work, the effect of force field on the dynamics of HIV-protease is investigated by performing molecular dynamics simulation using three AMBER force fields, namely ff99, ff99SB, and ff03. HIV-pr has complex dynamics as revealed by various computer simulation studies and NMR experiments. The structure of HIV-pr is characterized by two identical flexible beta-hairpin loops (flaps) covering the active site and by the presence of several other flexible loops and beta-strands. The dynamics of the flexible part of the protein controls the entry of inhibitor and substrate to the active site. The models for the conformational dynamics came from the previous simulations are not all consistent with each other and not all are supported by the NMR results. The generalized order parameters for amide backbone are calculated from the three force fields and compared with the NMR S^2 values. The flap movement is also investigated by examining various structural parameters. The results show that the ff99SB and ff03 force field calculated order parameters agree reasonably well with the NMR S^2 values and ff99 calculated values deviate most from the NMR order parameters. However, between ff99SB and ff03 there are several differences, most notably in the loop regions. It is found that these loops are, in general, more flexible in the ff03 force field. This results into a larger active site cavity in the simulation with the ff03 force field. Detail analysis of interactions suggests that it is likely that a combination of different factors (such as H-bonding, different torsion parameter) is contributing to this difference in fluctuation. The effect of this difference in computer aided drug design against flexible receptors is discussed.

1.8. Conclusion:

The investigations of drug resistance of HIV-pr due to mutations and conformational dynamics of the enzyme are reported here in this thesis. Taken together these studies indicate that conformational dynamics of HIV-pr is sensitive enough to the simulation setups and force fields difference. Also mutation in some specific region has both direct

TH-772_04610604

and indirect effects on the conformation and dynamics of HIV-pr. All these studies were carried out through MD simulations and few are compared with the experimental observations published by others. All these works have noteworthy applications in getting in to drug resistance and designing of new potential inhibitors against HIV-pr.

TH-772_04610604

CHAPTER 2

Theory and Methods

TH-772_04610604

2. Theory and Methods:

Theoretical basis of the methodology applied throughout the thesis is described in detail in the current chapter. Some insight into the molecular modeling methods, in particular with MD simulations and empirical force fields are presented.

2.1. Molecular Dynamics Simulations:

Molecular Dynamics (MD) simulations are useful tools for the theoretical study of biological molecules. It facilitates the understanding of fluctuations and conformational changes of proteins and nucleic acids at an atomic level and is also useful in molecular structure refinement in X-ray crystallography and NMR experiments. In MD, the time evolution of a molecular system through phase space is calculated by integrating Newton's laws of motion. The result of an MD calculation is a trajectory of the time evolution of the positions and velocities of the individual atoms constituting the system. The first MD simulation carried out was a study of the interactions of 32 hard spheres performed by Alder and Wainwright in the late 1950´s [Alder and Wainwright 1957, 1959]. Their studies threw light on the behaviour of simple liquids. The next major advance was in 1964, when Rahman carried out the first simulation using a realistic potential for liquid argon [Rahman 1964]. Subsequently, Rahman and Stillinger performed the first MD simulation of a realistic system in their simulation of liquid water in 1974 [Stillinger and Rahman 1974]. This work along with other theoretical works by Stillinger has led to a deeper understanding of the physics of glass forming liquids using the theory of inherent structures [Weber and Stillinger 1984, Stillinger and Weber 1988]. The first protein simulation was reported in 1977 with the vacuum simulation of the bovine pancreatic enzyme inhibitor (BPTI) [McCammon *et al.* 1977]. This study replaced the notion that proteins are essentially rigid structures. Since then, due to the development of computationally inexpensive and precise models for water molecules [Jorgensen *et al.* 1983], solvent effects have been studied to observe the behaviour of biomolecules in physiological environment. More realistic treatments of the system

TH-772_04610604

boundaries and accurate treatment of long-range electrostatic forces are the other developments in this area.

In the current scenario, MD simulations of solvated proteins, protein-DNA complexes and membrane lipid systems are quite common. MD simulations have been used to understand the basic thermodynamic properties of simple model systems as well as biological processes like conformational transitions, ligand binding, small protein folding, enzymatic mechanisms and ion transport in biological systems. Although a protein is not rigid, it can however undergo a variety of faster and slower structural rearrangements termed as conformational changes or transitions. The conformational flexibility and structural fluctuations of biomolecules are crucial to their biological functions like enzyme activity and substrate binding. But these motions are usually spontaneous and unsynchronized and thus difficult to probe in ensemble-averaged experiments. Hence, the study of biomolecular conformational changes is of huge importance in theoretical biophysics or biomolecular simulations [Levitt 1982, Karplus and McCammon 2002]. Also the protein-folding problem *i.e.*, the problem of how the amino acid sequence of a polypeptide determines its three dimensional native structure, is of vast significance in theoretical biophysics. The importance of understanding the protein folding process is evidenced by the Levinthal paradox [Levinthal 1968] which states that if the folding mechanism were to occur by random attempts of the system to explore all it's accessible configurations (which for an average protein are of the order of 10^{30}), assuming a characteristic time of 10^{-12} sec for the probing of each configuration, the average folding time would be of the order of 10^{10} years. The fact that folding processes in nature occur in immensely shorter times (in femtoseconds), proves that folding cannot be the result of a mere random search of the correct functional form through configurational space. In this effect, several attempts have been made to simulate the folding of proteins. For example, a protein-folding pathway has been observed during a 1-microsecond dynamics simulation in explicit water [Duan and Kollman 1998]. However, the main barrier in resolving the detailed mechanisms of protein folding is the lack of effective exploring of the conformational space due to the large number of degrees of freedom of a polypeptide chain. To solve the protein-folding problem,

TH-772_04610604

theoreticians are focusing on the simulation of conformational dynamics of small peptides, which serves as a model for protein-folding.

Karplus [Schaefer, *et al.* 1998] and van Gunsteren have attempted several times to simulate small peptides both with explicit and implicit solvent models [Daura, *et al.* 1998]. Yet the number of internal degrees of freedom for small peptide sequences is too large for complete sampling of configurational space with modern computers. For example, the backbone of a 30 amino acid residue peptide contains 60 dihedral angles. Theoretically, each dihedral angle can assume one of only two possible states, these alone would span a conformational space of approximately 2^{30} ($\sim 10^9$) distinct conformers. Although this extremely large number of degrees of freedom would seem to rule out any possibility of obtaining significant results from MD simulations, the subset of biologically interesting degrees of freedom turns out to be significantly smaller. Special sampling techniques such as umbrella sampling [Torrie and Valleau 1977] have been developed to explore conformational space of proteins and peptides efficiently [Guo *et al.* 1997, Mu and Stock 2002]. The sampling problem is particularly important for the calculation of thermodynamic properties from time trajectories. Before performing an MD simulation, it is essential to select an initial configuration of the system and to ensure that configurations do not have high potential energy interactions. To rule out instabilities during the simulation, energy minimization is required prior to running the simulation. An initial distribution of velocities must also be assigned to the system. The equilibrium distribution of velocities for a thermodynamic system is a Maxwell-Boltzman distribution; usually, it is more convenient to assign a Gaussian distribution of velocities initially and to run a short equilibration phase of dynamics to allow the kinetic and potential energy of the system to reach their equilibrium distributions spontaneously.

Further, a function describing the potential energy of the system must also be selected. Given an initial distribution of velocities and a starting set of coordinate values for the atoms of the system, a trajectory of the time evolution of the molecular system can then be obtained by integrating the potential energy function of the system numerically, using sufficiently small time steps, usually of the order of one femtosecond. Finally, the

TH-772_04610604

obtained trajectories must be stored and subsequently analyzed. In order to extract desired information and parameters, care must be taken to save coordinates (and velocities if needed) often enough so as not to lose the information. The potential energy function or force field describing the forces acting between the atoms as a function of their positions is the most important aspect of the simulation. Possibilities of obtaining physically reasonable results from an MD simulation are dependent on the accuracy of the applied potential energy function.

2.1.1. Theory of Molecular Dynamics Simulations:

The statement of the problem addressed by MD simulations appears very simple at a first look: an MD algorithm simply calculates the classical time evolution of the system. This is done by integrating Newton's law of motion:

$$F_i = m_i a_i \qquad \{i = 1, \ldots, N\} \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (2.1)$$

$$\frac{F_i}{m_i} = \frac{d^2 x_i}{dt^2} \qquad \{i = 1, \ldots, N\} \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (2.2)$$

Where, F_i is the force exerted on particle i, m_i is the mass of the particle, and a_i is its acceleration. For a mechanical system the force can also be expressed as the gradient of the potential energy:

$$F_i = -\nabla i V \qquad (i = 1, \ldots, N\} \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (2.3)$$

The difficulty lies in the fact that for "realistic" systems with a large number of atoms, analytical solutions cannot be derived and the equations have to be solved numerically. The discontinuous nature of computer based numerical calculations, in contrast to the continuous nature of realistic potentials, requires the equations of motion to be integrated by breaking the calculation into a series of very short time intervals (typically 1 or 2 femtoseconds) at atomic detail. At each step, the forces on the atoms according to the molecular mechanics force field are computed and combined with the current positions and velocities to generate new positions and velocities at the following time step. The

TH-772_04610604

total force on each atom at a time *t* is calculated as a vector sum of its interaction with other atoms in a pair-wise additive model. During the integration time step the forces are assumed to be constant. The atoms are then moved to the new positions; an updated set of forces is computed, and so on and so forth.

In this way an MD simulation generates a trajectory that describes how the system evolves through phase space as a function of time. Currently, several computationally efficient algorithms exist for integrating the equations of motion. All of them use Taylor series expansions of the positions and dynamic properties:

$$r(t+\delta t) = r(t) + \delta t v(t) + \frac{1}{2}\delta t^2 a(t) + \frac{1}{6}\delta t^3 b(t) + ... \quad \text{..... (2.4)}$$

Where *v* is the velocity (the first derivative of the positions with respect to the time), *a* is the acceleration (the second derivative), *b* is the third derivative, and so on. One of the most commonly used algorithms is the Verlet algorithm [Verlet 1967]. It uses the positions and accelerations at time *t*, and the positions from the previous step to calculate the new positions at time *t*+*dt*.

$$r(t+\delta t) = r(t) + \delta t v(t) \frac{1}{2}\delta t^2 a(t) + ... \quad \text{..... (2.5)}$$

$$r(t-\delta t) = r(t) - \delta t v(t) \frac{1}{2}\delta t^2 a(t) - ..... \quad \text{..... (2.6)}$$

Adding these two equations gives:

$$r(t+\delta t) = 2r(t) - r(t-\delta t) + \delta t^2 a(t) \quad \text{..... (2.7)}$$

The velocities do not explicitly appear in the Verlet integration but they can be calculated in a variety of ways. A simple approach is to divide the differences in positions at times

TH-772_04610604

$t+\delta t$ and $t-\delta t$ by $2\delta t$; or to estimate the velocities at the half-step. A variation on the Verlet algorithm is the leap-frog algorithm [Hockney 1970]. It uses the following relationships:

$$r(t+\delta t) = r(t) + \delta t v(t+\frac{1}{2}\delta t) \quad \text{..... (2.8)}$$

$$v(t+\frac{1}{2}\delta t) = v(t-\frac{1}{2}\delta t) + \delta t a(t) \quad \text{..... (2.9)}$$

In the leap-frog algorithm, the velocities $v(t+1/2\delta t)$ are first calculated, from the velocities at time $t-1/2\delta t$ and the accelerations at time t. The positions $r(t+\delta t)$ are then deduced from the velocities just calculated together with the positions at time t. The velocities at time t can be calculated as the average:

$$v(t) = \frac{1}{2}\left[v(t+\frac{1}{2}\delta t) + v(t-\frac{1}{2}\delta t) \right] \quad \text{..... (2.10)}$$

The advantage over the standard Verlet algorithm is that the velocities are explicitly included and that it does not require the calculation of differences of large numbers. After running the simulation, a trajectory is obtained containing the changes of the atomic coordinates of the system with time. From this trajectory dynamic parameters can be extracted and evaluated. Furthermore, stable structures and transition states can be analyzed and visualized.

2.2. Force Fields – the Empirical Potential Energy Function:

The fact that molecular mechanics works is due to the validity of several assumptions. The first to be named is the Born-Oppenheimer approximation. Without this assumption it would be impossible to contemplate writing the energy as a function of the nuclear coordinates at all. Other, more drastic approximations concern the single energy terms in the force field; these are represented by empirical potential energy functions. The functional form adopted for the potential energy is designed to allow efficient

TH-772_04610604

computation of the energy of a system as a function of the coordinates. The potential energy $V(\mathbf{r})$ can be described as the sum over the bonded and non-bonded energy terms of all the atoms in the system.

$$V(r) = V_{bonded} + V_{non-bonded} \qquad \text{……..……} \quad (2.11)$$

2.2.1. Bonded Interactions:

In most commonly used force fields bonded interactions are described as a sum over four simple harmonic terms that describe bond stretching and angle bending. Rotation about single bonds (torsions) is governed by sinusoidal energies. The planarity of groups (e.g., the amide planes of proteins) can also be enforced by harmonic potentials known as improper torsions.

$$V_{bonded} = V_{bonds} + V_{angles} + V_{torsions} + V_{impropers} \qquad \text{…….} \quad (2.12)$$

(a) Bond Stretching:

An empirical potential energy function for a typical bond is the Morse potential which has the form:

$$V_{Morse}(l) = D_e \left\{1 - \exp\left[-a\left(l - l_0\right)\right]\right\}^2 \qquad \text{……..} \quad (2.13)$$

$$a = \omega\sqrt{\left(\frac{\mu}{2D_e}\right)} \qquad \text{……..} \quad (2.14)$$

Where, D_e is the depth of the potential minimum, l is the bond length *i.e.*, the distance between the two bonded atoms, l_0 is the equilibrium value of the bond length, μ is the reduced mass, and ω is the frequency of the bond vibration in the approximation of small displacements from the equilibrium position *i.e*, if $l \approx l_0$. In this case the potential

TH-772_04610604

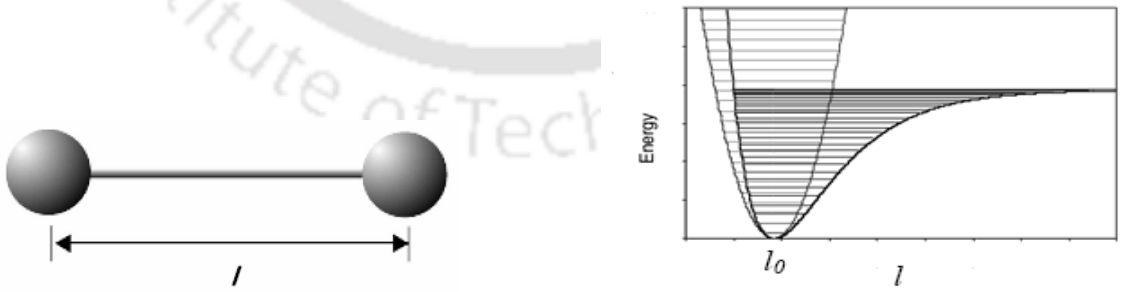
function can be approximated by a harmonic well and ω is related to the stretching constant of the bond, k, by:

$$\omega = \sqrt{\frac{\kappa}{\mu}} \qquad \text{..... (2.15)}$$

Although the Morse potential gives a precise description of the potential energy as a function of the bond length, it is generally not used in molecular mechanics force fields. This is mainly due to the computational expense of evaluating the exponential part of Eqn. 2.13 and also to the fact that three parameters are required per bond. The Morse curve describes a wide range of behaviour from equilibrium conditions to complete dissociation of the bond. At normal temperatures for biological systems in equilibrium it is rare for bonds to deviate significantly from their equilibrium values, under these assumptions a simple harmonic potential (Hooke´s law) is an adequate approximation for the description of bond stretching energies.

$$V_{bonds}(l) = \frac{k}{2}(l - l_0)^2 \qquad \text{..... (2.16)}$$

This has the advantage of being computationally inexpensive and requiring only two parameters per bond type. The harmonic approximation to the Morse curve is illustrated in Figure 2.1.



(a) (b)

Figure:2.1 *(a) Bonds between atoms separated by a distance "l". (b) Near to the equilibrium value "l_0 " the harmonic potential is a good estimate for the more accurate Morse curve.*

TH-772_04610604

Both, equilibrium bond length l_0 and force constant k are specific for each pair of bonded atoms, *i.e.*, they depend on the chemical type of the constituent atoms. Values for force constants and bond lengths can be evaluated from experimental data such as infrared stretching frequencies high-resolution crystal structures, microwave spectroscopy data or theoretically from quantum mechanical calculations.

(b) Bond Angle Bending:

The deviation of angles from their equilibrium values is also described using a harmonic potential:

$$V_{angle}(\theta) = \frac{k}{2}(\theta - \theta_0)^2 \qquad \text{..... (2.16)}$$

The contribution of each angle is characterized by a force constant k and an equilibrium value θ_0. Vibrational motions involving angle bending normally occur at lower frequencies than those of typical bond vibrations, less energy is required to distort an angle from its equilibrium value than to stretch a bond. This fact is reflected in the smaller force constants used for angle terms compared to those of bond terms in most force field implementations. The definition of angle bending terms is illustrated in Figure 2.2.

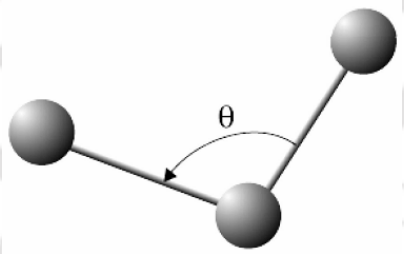


Figure: 2.2. *Angle bending terms are three atom terms characterized by a harmonic potential dependent on the angle between the three atoms.*

(c) Torsional Terms:

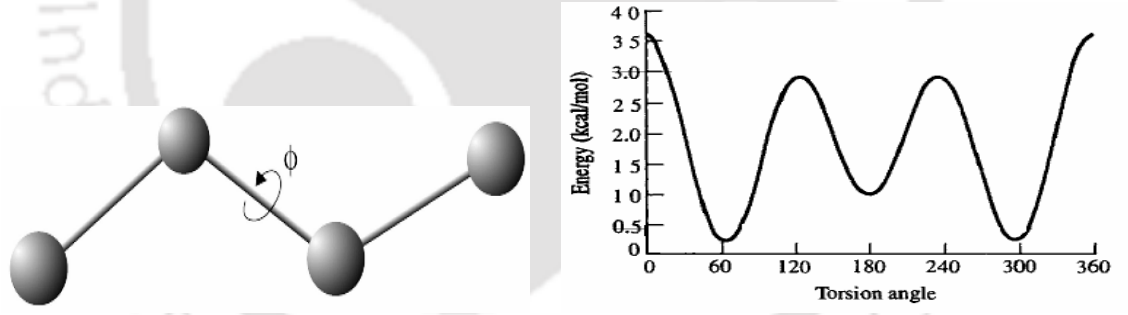
The torsional angle terms in the potential energy function model the presence of effective barriers for the rotation around chemical bonds. These are due to steric interactions between atoms or groups of atoms separated by three covalent bonds (1, 4-interactions).

TH-772_04610604

The proper modelling of barriers of rotation around chemical bonds is fundamental for the reproduction of structural properties of molecules and of conformational transitions. The potential is periodic and often expressed by a cosine function:

$$V_{torsion}(\phi) = K_{\phi}\left(1-\cos\left(n\phi-\phi_0\right)\right) \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (2.17)$$

Where $K\phi$ represents the barrier height, Φ is the torsion angle between the 1,4-pair, Φ_0 is an offset which defines the angular position of the first minimum in the potential, and n the multiplicity which gives the number of minima in the function as the bond is rotated through 360°. In most force fields two or more torsion terms may be assigned to the same 1,4-pair with different values of K, n, and Φ, summing up together in the energy evaluation. This allows the reproduction of very complex shapes for the rotational energy barrier. An example of this is illustrated in Figure 2.3 along with the definition of the torsion angle Φ.



(a) (b)

Figure: 2.3 *(a) Definition of the torsion angle Φ̃ (b) Energy curve for the torsional terms.*

(d) Improper Torsions:

The improper torsion is used to maintain chirality on a tetrahedral extended heavy atom or to maintain planarity of certain atoms like s*p2*-hybridized carbon atoms. These are 4-atom terms as for normal torsions, the potential energy is a harmonic function of the angle Φ between two planes defined as illustrated in Figure 2.4. The improper torsion functional form is given by:

TH-772_04610604

$$V_{improper}(\phi) = K_{\phi}(\phi - \phi_0)^2 \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (2.18)$$

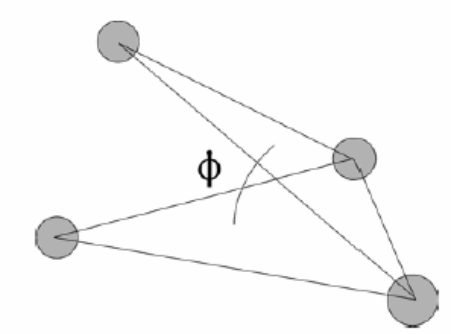


Figure: 2.4. *Improper torsional angle potential defined as a harmonic function of the angle Φ between two planes.*

2.2.2. Non-Bonded Interactions:

The energy term representing the contribution of non-bonded interactions has basically two components: the van der Waals interaction energy and the electrostatic interaction energy:

$$V_{non-bonded} = V_{VDW} + V_{electrostatic} \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (2.19)$$

The calculation of these contributions in MD is the most time consuming part because they contain long-range interactions of the atoms in the system. The interaction energy of one atom and the rest of the system are calculated as a sum of pair-wise (atom to atom) interactions. However, the simultaneous interaction of three or more atoms is not calculated. Also certain polarization effects are not explicitly included in the force field.

(a) The van der Waals Interactions:

The van der Waals interaction between two atoms arises from a balance between repulsive and attractive forces. The attractive forces are long-range forces whereas the repulsive forces act at short distances. The attractive contribution is due to dispersive

TH-772_04610604

forces arising from local electron fluctuations, which generate instantaneous dipoles in an atom or a molecule. An instantaneous dipole in a molecule can in turn induce a dipole in neighboring atoms, giving rise to an attractive inductive effect. The short-range repulsive contributions are due to the electron-electron interactions. These are often referred to as exchange forces. The attractive interactions are long ranged than the repulsive ones but as the distance becomes shorter the repulsive interaction becomes dominant. The van der Waals interaction is most often modeled using the Lennard-Jones 6-12 potential:

$$V(r) = 4\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right] \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (2.20)$$

The Lennard-Jones potential contains only two adjustable parameters: The collision diameter σ (the separation for which the energy is zero) and the well depth "ε". It is characterized by an attractive part that varies as r^{-6} and a repulsive part that varies as r^{-12}. Although this functional form is used in most force fields, there is plenty of evidence that r^{-12} is a poor representation of the repulsive potential, an exponential function; $e^{-r/r^{\circ}}$ would be a much better approximation as it is more faithful to the exponential decay of atomic wave functions and of their overlap at large distances, which is responsible for repulsion. To use the Lennard-Jones potential in MD simulations one would have to calculate all the pair-wise interactions of all the atoms in the system because the potential decays to zero only at infinite distance. At a distance $r = 2.5\ \sigma$ the potential has only 1% of its value at $r = \sigma$. This reflects the r^{-6} distance dependence of the dispersion interaction. In order to reduce the number of interaction terms, and thus the calculation time, the Lennard-Jones potential is often truncated. This can be done by defining an appropriate cut-off distance and calculating the pair-wise interactions only for the atoms lying within this distance and all van der Waals interactions of atoms beyond the cutoff distance are set to zero. The cut-off criterion for the potential energy function can be realized in two different ways. One solution is to simply truncate the potential at the cut-off distance: at $r \geq r_c$ the potential is set to zero and interactions beyond that distance are neglected. A problem of this method is the discontinuity introduced in the force at the cut-off distance. At the cut-off distance, the force will have a finite value, which drops

TH-772_04610604

suddenly to zero just beyond that cut-off. An alternative method is the use of a switching function. A switching function decreases the interaction potential over a predefined range of distances. The potential takes its usual value up to the first cut-off and is then switched to zero between the first and the second cut-off. This model suffers from strong forces in the switching region due to the rapid change in the derivative of the potential energy function. The Lennard-Jones potential and the two cut-off methods are illustrated in Figure 2.5.

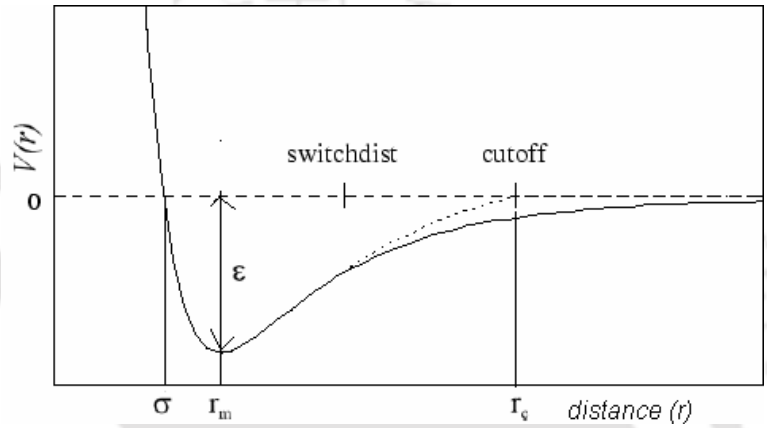


Figure: 2.5. *The Lennard-Jones potential (straight line). A shifted (dashed line) function can be used to realize a cut-off criterion: beyond the cut-off distance "r_c"the potential is zero.*

(b) Electrostatic Interactions:

Accurate reproduction of the electrostatic properties of a molecule is of fundamental importance in force field development. Electrostatic interactions act at longer range than van der Waals interactions. Truncation schemes must be applied carefully, and in some cases should be entirely avoided. A common approach to modelling these interactions is to distribute a number of fractional point charges throughout the molecule. These sets of charges are designed to reproduce the electrostatic properties of the molecule. In most cases the charges are restricted to the nuclear centres; these are referred to as partial atomic charges. The electrostatic interaction between two molecules (or between different parts of the same molecule) is then calculated as a sum of interactions between pairs of point charges, using Coulomb's law:

TH-772_04610604

$$V_{electrostatic} = \sum_i \sum_j \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \qquad \text{..... (2.21)}$$

Where q_i, and q_j are the partial atomic charges, ε_0 is the effective dielectric constant and r_{ij} is the relative distance between the two particles. The strength of Coulomb interactions decreases with r^{-1}. Because of this the use of cut-off methods in most cases is not appropriate and can cause significant errors.

In summary, the potential energy in a molecular mechanics force field is calculated as the sum of all the bonded and the non-bonded interactions over all the atoms of the system:

$$V(r) = \sum V_{bonds} + \sum V_{angles} + \sum V_{torsions} + \sum V_{VDWs} + \sum V_{electrostatics} \text{ . (2.22)}$$

Molecular mechanics force fields provide a reasonably good compromise between accuracy and computational efficiency. In some cases molecular mechanics can provide answers that are as accurate as high-level quantum mechanical calculations, in only a fraction of the computer time.

The standard MD simulation package AMBER 7 and 8 [Case *et al.* 2002 & 2004] were used in the present work. *Sander*, one of the AMBER modules carries out the energy minimization, molecular dynamics and NMR refinements. It provides direct support for several force fields (non-polarizable and polarizable) for proteins and nucleic acids, and for several water models and other organic solvents. The basic force field implemented here has the following form, which is about the simplest functional form that preserves the essential nature of molecules in condensed phases and that describes the non-polarizable (or additive) force fields:

TH-772_04610604

$$V(r) = \sum_{bonds} K_l (l - l_{eq})^2 + \sum_{angles} K_\theta (\theta - \theta_{eq})^2 + \sum_{dihedrals} \frac{V_n}{2}(1 + \cos[n\phi - \gamma]) + \sum_{i<j}^{atoms} \frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} + \sum_{i<j}^{atoms} \frac{q_i q_j}{\varepsilon R_{ij}} \quad \text{..... (2.23)}$$

2.2.3. Non-polarizable and Polarizable force fields:

The general force fields mostly used in biomolecular simulation describe the electrostatic interaction in terms of fixed and atom-centered charges, which is a characteristic of the additive or non-polarizable force fields [Mackerell *et al.* 1998; Cornell *et al.* 1995] of which the potential is described in equation 2.23. However in a real physical system like high dielectric water medium in a condensed-phase simulation, polarization has a special impact on the molecular charge distribution. Also, for a neutral body in the gas phase has the effect of polarization, when a strongly charged system approaches towards it. The main problem with the non-polarizable force field is that the charges do not respond to the change in the local environment, i.e. remain fixed throughout the simulation. However, it is likely that fixed charge model may describe the local electrostatics poorly for cases such as interaction of a charged residue with water and also with salt bridge formation. To challenge this issue, polarizable force fields have been developed by several groups [Kaminski *et al.* 2002, Cieplak *et al.* 2001 and Ren *et al.* 2003], where the electrostatic property of atom changes with the change in environment. This is according to the principle of electronegativity equalization, which states that the charge flows between atoms until the instantaneous electronegativities of the atoms are equal. For instance, in the ff02 force field of AMBER program package, the polarization is described by a combination of fixed charge and induced dipole. The induced dipole

TH-772_04610604

moment of an atom depends on the electric field coming from rest of the system and on the atom centered polarizability. Addition of a polarization term to the additive force field makes it "Non-additive" and is based on atom-centered dipole. The polarization term can be added by the following equation.

$$E_{pol} = -\frac{1}{2}\sum_{i}^{atom} \mu_i \bullet E_i^{(0)} \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (2.24)$$

Where μ_i is an atomic polarizability. Also, the off center charges except the atomic centered charges can be included in polarizable force fields. In milieu of MD simulation, the equations for polarizable force field are competently solved using the extended Lagrangian method [Rick *et al.* 1994] at a computational cost slight more than that required for a fixed-charge, pairwise-additive force field.

2.2.4. Water Models: TIP3P

A water model is described by its geometry, together with other parameters such as the atomic charges and Lennard-Jones parameters. The first water model came in 1983 from Jorgensen at al. describes about three interaction sites, i.e., TIP3P (Transferable Intermolecular Potential 3-Point) [Jorgensen *et al.* 1983]. However after that many water models came up with 3 to 6 point interaction sites. The simplest water models assume the water molecule as rigid and depend only on non-bonded interactions. The electrostatic interaction is modeled using Coulomb's law where as the Lennard-Jones potential account for the attractive and repulsive forces. The potential for TIP3P model is represented by

$$E_{ab} = \sum_{i}^{\text{on } a}\sum_{j}^{\text{on } b} \frac{k_C q_i q_j}{r_{ij}} + \frac{A}{r_{\text{OO}}^{12}} - \frac{B}{r_{\text{OO}}^{6}} \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (2.25)$$

Where k_C is the electrostatic constant, q_i and q_j are the partial charges relative to the charge of the electron; r_{ij} is the distance between two atoms or charged sites; and A and B are the Lennard-Jones parameters. The charged sites are on the three atoms. In most

TH-772_04610604

water models, the Lennard-Jones term applies only to the interaction between the oxygen atoms.

As mentioned above, the TIP3P model has three interaction sites, which correspond to the three atoms of the water molecule. A point charge is assigned to each atom; also the oxygen atom gets the Lennard-Jones parameters. This model is widely used for molecular dynamics simulations because of its simplicity, reasonable structural and thermodynamic descriptions and computational efficiency. The model is robust enough to explain the properties of liquid water at other thermodynamic points and biological water. The model is advanced over the two-site models in both static and dynamic properties. The H-O-H angle for the water molecule is of 104.5° and the O-H distance is 0.957 Å. The simple models for TIP3P water is represented as below:

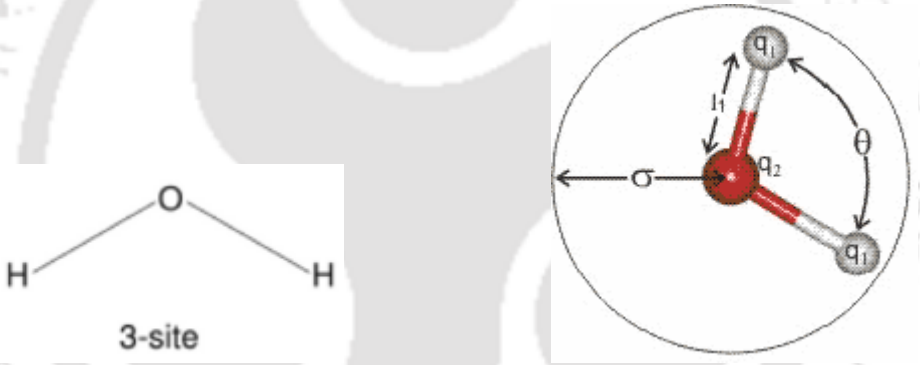


Figure 2.6: *Schematic representation of the TIP3P water models.*

2.3. Energy Minimization procedures:

For most of the systems the potential energy is a complicated, multidimensional function of the 3N cartesian coordinates. The minima on this multidimensional energy surface are of special interest for both quantum mechanics and molecular mechanics calculations. Indeed, some molecular mechanics force field parameters can be obtained by fitting molecular properties calculated by quantum chemical methods in the 'ground state' (*i.e.*, in the global potential energy minimum). Minimization algorithms are used to identify geometries of the system corresponding to minima of the potential energy surface for

TH-772_04610604

which the numbers can be very large. This is true especially for biomolecular systems consisting of thousands of atoms and a huge number of degrees of freedom.

$$\frac{\partial f}{\partial x_i} = 0; \frac{\partial^2 f}{\partial x_i^2} > 0 \qquad (i = 1, \ldots, N) \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots (2.26)$$

Minimization algorithms are also very important for MD simulations, where it is important to start from well-minimized structures to avoid unwanted high energy interactions. Given a function f which depends on the variables $\{x_1, x_2, \ldots, x_n\}$, a minimum of f is defined as a point where the first derivative of the function with respect to each of the variables is zero and the second derivatives are all positive:

For analytical functions, the minima can be found using standard calculus methods. However, this is not possible for systems such as biomolecules where the shape of the potential energy landscape cannot be expressed by a simple analytical formula. In these cases, minima are located using numerical methods that gradually change the coordinates to produce configurations with lower and lower energies until the minimum is reached. The most common minimization algorithms use derivatives of the energy in respect to the coordinates to predict the location of the closest minimum. The two major issues to be considered in the development of a minimization algorithm are the speed of convergence of the method and the memory requirements. No single algorithm has yet been proved to be the best for all problems; rather a combination of different algorithms is commonly used. Most minimization algorithms can only go downhill on the energy surface. Thus, they can only locate the minimum that is nearest (in a downhill sense) to the starting point. To locate more than one minimum or the global energy minimum with these algorithms many minimizations from different starting points are required. However, for a true exploration of the potential energy landscape, energy minimization alone is not sufficient.

The energy minimization procedures used in this work are a combination of two derivative minimization methods: the steepest descent and the Newton-Raphson method.

TH-772_04610604

When discussing derivative methods it is useful to write the potential energy function V as a Taylor series expansion about the point $\mathbf{r}_0$:

$$V(r) = V(\mathrm{r}_0) + (\mathrm{r}\text{-}\mathrm{r}_0)V'(\mathrm{r}_0) + (\mathrm{r} - \mathrm{r}_0)^{\mathrm{T}} \frac{1}{2} V''(\mathrm{r}_0)(\mathrm{r}\text{-}\mathrm{r}_0) + \ldots \quad \ldots. (2.27)$$

Where $\mathbf{r}_0$ and $\mathbf{r}$ are $3N$ dimensional vectors. V' ($\mathbf{r}_0$) is the vector of the partial derivatives of V with respect to each coordinate and V'' ($\mathbf{r}_0$) is the Hessian matrix; it contains the partial second derivatives and is of $3N$ x $3N$ dimensions. In the vicinity of a minimum $V(\mathbf{r})$ can be approximated by truncating the series at the second order. This approximation is used also for normal mode calculations.

The direction of the first derivative of the energy indicates where the minimum lies and the magnitude of the gradient indicates the steepness of the local slope. By moving each atom in response to the force (negative gradient) acting on it the energy of the system is lowered. Second derivatives indicate the curvature of the function, information that can be used to predict where the function will change direction.

The energy minimization methods can be divided into two classes: the first-derivative techniques like steepest descent and conjugate gradient; and the second-derivative methods like the Newton-Raphson and related algorithms.

2.3.1. Steepest Descent Method:

The steepest descent method is a first-order minimization algorithm. It gradually changes the coordinates of the atoms as it moves the system closer and closer to the minimum point. The direction of movement on the potential energy surface is parallel to the net force, which in geographical analogy corresponds to walking straight downhill along the steepest descent. The minimum is located iteratively using a line search algorithm. The method of steepest descent is very robust, meaning that the minimum is found even when the starting structure is far away from the minimum; so it is a good method to start with. On the other hand, the close-range approach to the minimum may require quite a large

TH-772_04610604

number of iterations and, due the inflexible determination of the search direction unwanted oscillations around the minimum may occur.

2.3.2. Conjugate Gradient Method:

The conjugate gradient method is also a first-order minimization algorithm, which accumulates the information about the function from one iteration to the next. With this proceeding the reverse of the progress made in an earlier iteration can be avoided. For each minimization step the gradient is calculated and used as additional information for computing the new direction vector of the minimization procedure. Thus, each successive step continually refines the direction towards the minimum. The computational effort and the storage requirements are greater than for steepest descent but conjugate gradients is the method of choice for larger systems. The greater total computational expense and the longer time per iteration are more than compensated by the more efficient convergence to the minimum achieved by conjugate gradients.

2.3.3. Newton-Raphson Method:

The Newton-Raphson method is a second-order derivative method, which uses the inverse Hessian matrix for energy minimization. For a purely harmonic function the method finds the minimum in only one step. As the real potential energy function contains anharmonic terms, again a tedious repetition is required. But the number of iterative steps required to reach a minimum is generally less in comparison to first-order derivative methods. The efficiency of the method increases as convergence is approached. The computational effort and the storage requirements for calculating larger systems are disadvantages of this method.

For most applications a combination of steepest descent and Newton-Raphson minimization is favored. With the steepest descent method the structure can be efficiently brought close to the minimum and with subsequent Newton- Raphson minimization the minimum is located within a few steps.

TH-772_04610604

2.4. Use of charges and solvents:

The molecular mechanics calculations are generally carried out under vacuum conditions (ε = 1). For non-polar hydrocarbons, the effect of the explicit inclusion of solvent is negligible as compared to gas phase calculations. The investigation of molecules containing charges and dipoles however requires the consideration of solvent effects [Burkert *et al.* 1982] to prevent the excessive influence of strong electrostatic interactions. The force field will try to maximize the attractive electrostatic interaction, resulting in energetically strongly preferred but unrealistic low-energy conformations of the molecule. This can be prevented by employing the corresponding solvent dielectric constant [Eliel *et al.* 1965]. In contrast to macromolecules, the electrostatic field of small molecules is considered to be homogeneous. Experimentally determined dielectric constants for a large number of solvents are found in the literature, which can be applied for a correct treatment of the Coulombic term of solvated molecules. An accurate way to care for the problem of charges and solvation in the course of a molecular mechanics optimization is to perform the calculation without taking charges into consideration. This usually yields satisfying results and is especially recommended if the results of a conformational analysis are to be minimized, because usage of charges may markedly alter the conformation by electrostatic interactions. The strength of the electrostatic interaction decreases with distance**.** Therefore, in some force fields the dielectric constant can be chosen to be distance-dependent in order to simulate the effect of displacement of solvent molecules in course of the approach of a ligand molecule to a macromolecular surface.

Other than resorting to dielectric constant to simulate solvent effects, a solvent box containing discrete solvent molecules can be created around the molecule. The additional computational effort and the limitations in regard to the limited number of solvents that can be used in most of the available force fields are severe weakness of this method.

2.4.1. Ewald Summation Techniques:

A major problem in correctly evaluating the electrostatic contribution to the potential energy is that, if one uses periodic boundary conditions, where a charged particle

TH-772_04610604

interacts with all the other charges in the simulation box and with all of their images in an infinite array of periodic cells. The system thus contains an infinite number of charges. The total electrostatic potential energy, Φ ($\mathbf{r}_i$) at the position $\mathbf{r}_i$ of any charge i of the primary cell is given by the finite difference of two infinite, diverging series:

$$\phi(\mathrm{r}_i) = q_i \sum_{j+}^{\infty} \frac{q_{j+}}{\left|\mathrm{r}_i - \mathrm{r}_{j+}\right|} + q_i \sum_{j-}^{\infty} \frac{q_{j-}}{\left|\mathrm{r}_i - \mathrm{r}_{j-}\right|} \qquad \text{..... (2.28)}$$

The Ewald summation approach, first proposed by Ewald [Ewald 1921], solves this problem by splitting the potential into two well behaved and rapidly converging parts, the first being represented in real space and the second in reciprocal space. A simple illustration of the method can be made for a one-dimensional ion lattice with a charge distribution as shown in Figure 2.7; the extension to three dimensions is insignificant. The original lattice is supplemented by a set of Gaussian charge distributions of opposite sign to form an auxiliary lattice (lattice 1, in Figure 2.7):

$$\rho(\mathrm{r}) = -q_j \left(\frac{\eta^2}{\pi}\right)^{\frac{3}{2}} e^{-\eta^2 (\mathrm{r}-\mathrm{r}_j)^2} \qquad \text{..... (2.29)}$$

A further lattice (lattice 2, in Figure 2.7) is then introduced to compensate the additional Gaussian distributions, such that "lattice 1 + lattice 2 = original lattice". The contributions of the two lattices to the potential energy are computed separately. The effect of lattice 1 at large distance, due to the compensating effect of the Gaussian distributions, tends rapidly to zero. The narrower the Gaussian distributions, the more they resemble the delta functions of the original charge distribution and the more the compensating effect of the Gaussian distributions is efficient *i.e.*, the larger the value of η the faster the series will converge. The effect of lattice 1 is therefore best computed in real space, where the series will rapidly converge. For lattice 2 the potential sum is evaluated in k-space. When the Gaussians are broad *i.e.*, when η is small, a smaller number of Fourier components is required. By suitably adjusting the parameter η, optimal convergence of both series can be achieved. Ewald summations are computationally quite

TH-772_04610604

expensive to implement. Nonetheless, the Ewald method is the most correct way to accurately include all the effects of long-range forces in computer simulations.

In the Ewald procedure, computation of the reciprocal energy from lattice 2 is the most expensive part in the evaluation of long-range electrostatic contributions.

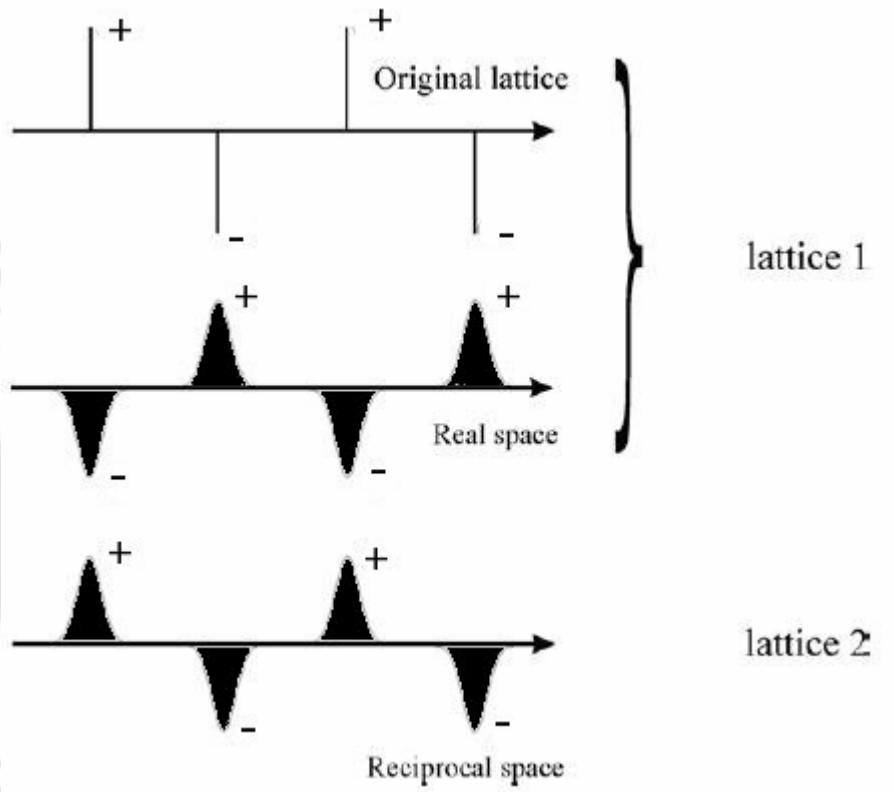


Figure: 2.7. *The Ewald summation method: the original set of charges is summed to a Gaussian distribution (calculated in real space) and to a canceling charge distribution (calculated in the reciprocal space).*

(a) Particle Mesh Ewald Method:

To handle long-range electrostatics and make their computation more efficient, a variety of methods have been developed [Leach 1996]. The method used in this work to calculate the electrostatics is the Particle-Mesh Ewald method (PME) [Essmann *et al.* 1995]. The PME method considerably speeds up the process. The direct sum in real space is evaluated explicitly using cut-offs while the reciprocal sum is approximated using fast Fourier Transforms (FFT) with complications on a grid in which charges are interpolated

TH-772_04610604

to the grid points. PME evaluates the forces by analytically differentiating the energies, thus significantly reducing memory requirements.

2.4.2. Periodic Boundary Conditions:

A periodic boundary is an important technique in a molecular dynamics simulation. It is a strategy to make a simulation consisting of only a few hundred atoms to experience forces as if they are in a bulk solution of infinite size. The reason of its requirement is to eliminate the dominating effects of the surface. The figure below illustrates the concept of periodic boundary conditions.

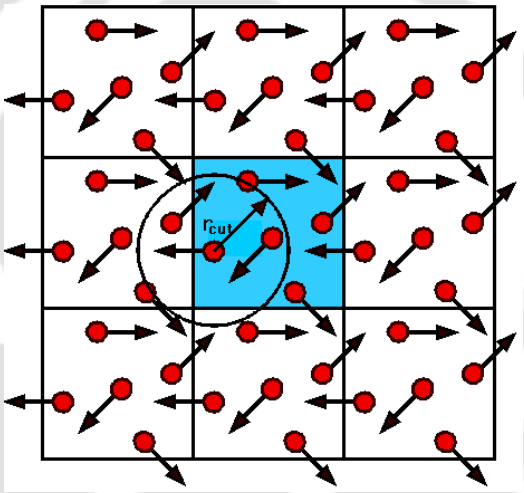


Figure: 2.8. *Periodic boundary condition (PBC)*

The colored box represents the system we are simulating, while the surrounding boxes are identical copies in every detail. In other words, every particle in the simulation box has an exact duplicate (having same velocities too) in each of the surrounding boxes. This arrangement is imagined to fill the whole of space. Whenever an atom leaves the simulation cell, it is replaced by another from the opposite cell face. So, the number of atoms in the cell is conserved. Furthermore, no atom feels any surface forces, as these are now completely removed. In the figure r_{cut} is the cutoff radius that is normally applied when calculating the force between two atoms. As the figure indicates, an atom may interact with one in the neighboring cell (which is an image of one of the atoms in the simulation cell) because it is within the cutoff radius. It ignores the equivalent atom in the

TH-772_04610604

simulation cell because it is too far away. In other cases the interaction comes from an atom in the simulation cell itself. Thus the interaction that is calculated is always with the closest image. This is known as the minimum image convention.

The cutoff is chosen such that an atom in the primary box can interact with only one image of any given atom. This means that r_{cut} cannot be greater than half the width of the cell.

2.5. SHAKE Algorithm:

In the MD simulations the size of time step should be smaller than the motions characterized by the highest frequencies. In classical MD simulations these are typically the bond stretching motions involving hydrogen atoms. In order to use a larger time step, all the bonds involving hydrogen atoms are constrained to their equilibrium position with the SHAKE algorithm [Armstrong 1998; Ryckaert *et al.* 1977]. In this algorithm at each step a correction $\vec{g}_a^{(r)}$ directed along the bond, is introduced in the forces such as to guarantee that the constraint is satisfied.

$$\vec{r}(t+\delta t) = \vec{r}(t+\delta t) + \frac{\delta t^2}{m_a}\vec{g}_a^{(r)} \qquad \text{..... (2.30)}$$

where $r\ (t+\delta t)$ is the position that the system would have reached in the absence of the constraint. The procedure is then applied on the next constraints. In a chain of atoms the correction applied on constraint $i+1$ would partially disrupt constraint i. For this reason the correction is applied cyclically on each constraint in the system, until the desired convergence is reached for all the constraints.

2.6. The Berendsen thermostat.

Constant temperature MD simulations could be obtained by coupling to a Berendsen thermal bath [Berendsen *et al.* 1984]. At each step the velocities are scaled by a factor

TH-772_04610604

$$\chi = \left(1 + \frac{\delta t}{\tau_T} \left(\frac{T}{T_0} - 1 \right) \right) \qquad \text{(2.31)}$$

Where T_0 is the reference temperature and τ_T is a time constant that determines the strength of the coupling between the system and the thermal bath. A similar algorithm can be used to obtain constant pressure by periodic scaling of the simulation cell size and atomic positions.

2.7. Setting up Molecular Dynamics Simulations.

2.7.1. Initialization:

An MD simulation for biomolecules can be started taking an initial configuration as a starting point from an X-ray crystal structure or an NMR structure obtained from the Brookhaven Protein Data Bank. Homology modeling based theoretical structure can also be used as the starting structure for an MD simulation. Careful selection of the initial structure influences the quality of the simulation and hence it is often good to prefer a structure close to the state that one wish to simulate. Generally, the initial structures may have a possibility of having strong van der Waals interactions, which may lead to local structural distortion, resulting in an unstable simulation. So prior to start an MD simulation, it is wise to do an energy minimization of the structure thereby eliminating any strong van der Waals interactions that may exist.

The initial X-ray crystal structure may contain some water molecules, but the amount is usually inadequate for solvation. So, in order to solvate the protein, it is required to add explicit water molecules. So, water molecules are added explicitly to the system. To get rid of the surplus water molecules overlapping the protein, a suitable cutoff value is adjusted. Re-minimization of the protein fixed in its energy minimized position is done to readjust the water molecules to the protein. Once the waters are adjusted, the restraints on the protein are removed to enable the whole system (protein + water) to progress in time.

TH-772_04610604

2.7.2. Heating Equilibration:

During the heating stage, initial velocities are assigned to each atom of the system at a low temperature by randomly selecting from a Maxwell-Boltzmann distribution and Newton's equations of motion are integrated to propagate the system in time. At times, new velocities are assigned at a slightly higher temperature and the simulation of protein/water system is allowed to continue. This process is repeated until the desired temperature is obtained. Once the optimum temperature is reached, properties like the structure, pressure, temperature and energy are examined. The objective of the equilibration phase is to run the simulation until these properties become stable with respect to time and to bring the system to equilibrium from the starting configuration.

2.7.3. Production phase:

The production phase commences after the system get equilibrated and is the final step of the simulation that can be spanned up to the desired time length from several hundred picoseconds (ps) to nanoseconds (ns) or more. It is during the production phase that thermodynamic properties and other data can be calculated. During the production phase no velocity scaling is performed and hence the temperature becomes a calculated property of the system. Various properties (mostly the atomic coordinates and velocities) are routinely calculated and stored during the production stage for subsequent analysis and processing.

2.7.4. Analysis of Trajectory:

In an MD simulation, coordinates and velocities of the system are saved for further analysis. Most of the time dependent properties can be displayed graphically, where one of the axes corresponds to time and the other to the quantity of interest, such as energy, RMSD, RMSF, B-factors and time dependent properties such as correlation functions etc. Average structures can be calculated and compared to experimental structures. The quantities that are usually calculated from an MD simulation include:

TH-772_04610604

(a) **Mean Energy:** $\langle E \rangle = \frac{1}{N}\sum_{i=1}^{N} E_i$ ..... (2.32)

(b) **RMSD:** $\left\langle \left(r_i^{\alpha} - r_i^{\beta} \right) \right\rangle^{1/2} = \sqrt{\frac{1}{N}\sum_i \left(r_i^{\alpha} - r_i^{\beta} \right)^2}$ ..... (2.33)

(c) **RMSF:** $\sqrt{\frac{1}{N_f}\sum_f \left(r_i^{f} - r_i^{ave} \right)^2}$ ..... (2.34)

(d) **B-factors** (Temperature factors): $B_i = \frac{8}{3}\pi^2 \sqrt{\frac{1}{N_f}\sum_f \left(r_i^{f} - r_i^{ave} \right)^2}$ .... (2.35)

(e) **Radius of Gyration** (Rg): $\sqrt{\frac{1}{N_i}\sum_i \left(r_i - r_{cm} \right)^2}$ ..... (2.36)

where r_i - r_{cm} is the distance between atom *i* and the center of mass of the molecule.

(f) **Time-correlation function**: $C(t) = P2(t) / R(t)$ ..... (2.37)

where *P2(t)* is *<P2 (v(0).v(t))>* with *P2* is the legendre polynomial of order 2 and "*v*" is the vector chosen and *R(t)* is *<b^3(0) b^3(t)>* with b as the length of the vector chosen.

(g) **S^2 order parameter**: It is the measure of the fluctuations in the orientation of an N-H bond internuclear vector. It is a time dependent properties such as correlation functions and is described by

$$S^2 = \lim_{t \to \infty} C(t) \quad \text{..... (2.38)}$$

TH-772_04610604

2.8: Molecular Modeling and Visualization.

Molecular dynamics simulations can help visualize and understand conformational changes at an atomic level when combined with molecular graphics programs. Programs like Xgrace [Turner 1995, Portland] can display the structural parameters of interest in a time dependent way and the programs like CHIMERA [Pettersen *et al.* 2004], RasMol [Sayle *et al.* 1995], Sybyl [Tripos International], VMD [Humphrey *et al.* 1996], PyMol [DeLano Scientific LLC 2002], SPDBV [Guex *et al.* 1997] are useful tools for protein visualization and modeling.

TH-772_04610604

CHAPTER 3

Drug resistance of HIV-1 Protease against JE-2147: I47V mutation investigated by Molecular Dynamics simulation

TH-772_04610604

Drug resistance of HIV-1 Protease against JE-2147: I47V mutation investigated by Molecular Dynamics simulation.

3.1. Introduction:

HIV is a retrovirus and its genome is encoded by RNA, which is reverse-transcribed to viral DNA by the viral reverse transcriptase (RT) upon entering a new host cell. This is followed by integration, transcription, translation and finally assembly and budding of HIV-proteins. Drugs against HIV have been developed by targeting key players at different points of the viral life cycle. To date, Food and drug administration (FDA) has approved anti-HIV drugs targeting three different viral proteins, HIV-RT, HIV-protease (HIV-pr) and HIV-gp41. Nine FDA approved drugs target HIV-pr, a member of the aspartyl protease enzyme family, which functions at the late stage of infection by cleaving viral Gag and Gag-Pol polyproteins thereby generating mature infectious virions. The knowledge that incomplete processing of these virions produces immature, non-infectious viral particles lead to intense research to find HIV-pr inhibitors [Swain *et al.* 1990, Thaisrivongs *et al.* 1996, Wang *et al.* 2000, Perryman *et al.* 2004]. This is also one of the major success stories of developing drugs using structure-based techniques using a combination of x-ray crystallography and computational work [Wlodawer *et al.*1998].

The major problem associated with the existing drugs against HIV-pr is drug resistance of different kinds. The drugs against HIV-pr are losing their effectiveness due to rapid point mutations in the genome of HIV. Two kinds of mutations have been identified for HIV-pr. One kind occurs at or near the active site by affecting the binding of drugs directly by reducing the van der Waals (VDW) contacts, increasing steric hindrance and increasing the number of unfavorable electrostatic interactions between HIV-pr and inhibitors [Wittayanarakul *et al.* 2005]. The other kind occurs far from the active site with enhancement of the enzymatic function [Piana *et al.* 2002b] by conformational changes that increase the affinity of the protease for substrates over inhibitors, the so-called compensatory mutation. Computational studies have been extremely valuable in

TH-772_04610604

investigating the molecular basis of drug resistance since from static X-ray structures alone it is difficult to understand the protein conformational change and dynamics.

In the current work we focus on HIV-pr by considering a very high resolution (1.09 Å) crystal structure of HIV-pr with ligand JE-2147 [Reiling *et al.* 2002]. This high resolution structure allows closer look into the drug resistance mechanism of HIV-protease. JE-2147, shown in Figure-3.1, is a peptidomimetic protease inhibitor developed by the company Agouron (Pfizer). *In vitro* data suggest that JE-2147 can be more efficient than the currently marketed HIV-pr inhibitors/drugs [Yoshimura *et. al.* 1999, Ohtaka *et al.* 2003]. JE-2147 is also interesting because of its unique resistance profile. There are two major mutations, which reduce the efficacy of JE-2147, I84V and I47V. While I84V is common for other similar ligands, I47V seems to be specific for JE-2147.

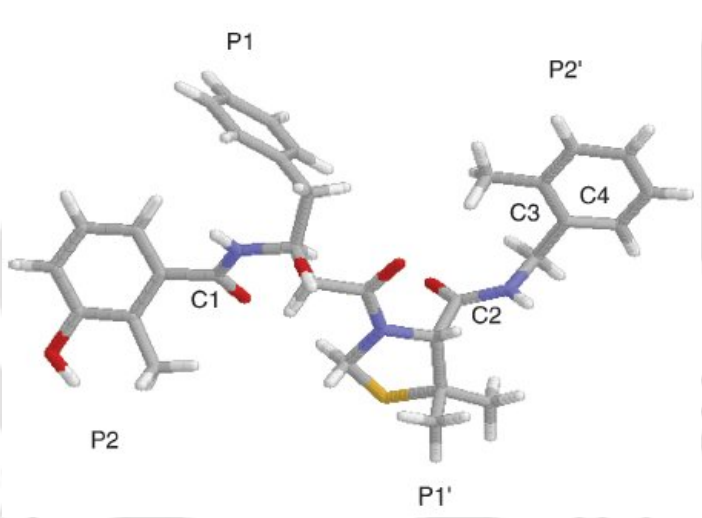


Figure 3.1: *Structure of JE-2147. Atoms H, C, O, N and S are shown in color white, gray, red, blue, and yellow respectively. P1, P2, P1' and P2' represent different substituents. Atoms C_1, C_2 , C_3 and C_4 were used to define structural parameters (defined in the text) for analysis.*

We have investigated the effect of I47V mutation using molecular dynamics simulation for both apo and complexed protein (for both WT and mutant) to get insight to this complex process. The outcome of the simulations shows that the mobility of the ligand

TH-772_04610604

differs on mutation and also the mobility of residues in the flap region differs especially for the apo form.

3.2. Computational Details:

All four molecular dynamics (MD) simulations were started using the high resolution crystal structure of HIV-pr with JE-2147 (pdb ID 1KZK). Leap module of AMBER 7 program package [Case *et al.* 2002] was used to prepare the system for the simulation. The AMBER 99 force field was used for the simulation [Cornell *et al.* 1995]. Charges of JE-2147 were calculated using the RESP [Bayly *et al.* 1993] procedure at the Hartree-Fock level with 6-31G* basis set after minimizing the molecule at the AM1 semi-empirical level [Dewar *et al.* 1985]. All four systems were immersed in a water box of size containing more than 8000 water molecules. TIP3P [Jorgensen *et al.* 1983] model was used to represent the water molecules. The positive charges of protein-water system were neutralized by chloride ions.

For the simulations of the complexed protein, charged state of the ASP25' (one of the two catalytic aspartates) was taken as protonated. This is in accord with the finding in reference [Wang *et al.* 1996b] that ASP-25' was protonated. Detailed calculations [Wittayanarakul *et al.* 2005] for HIV-pr and Saquinavir, a ligand similar to JE-2147 also found that one of the ASPs is protonated. The electrostatic interactions were calculated with the particle mesh ewald method [Essman *et al.* 1995]. Constant temperature and pressure conditions in the simulation were achieved by coupling the system to a Berendsen's thermostat and barostat [Berendsen *et al.* 1984]. Bonds involving the hydrogen atoms were constrained to their equilibrium position with the SHAKE algorithm. The whole system was minimized for 200 steps. Then the system was heated to 300 K over 20 ps with 1 fs time step. Subsequently, 180 ps MD run was performed for equilibration. The time step for MD simulation for the production run was 2 fs. The system was then run for 3 ns and the stability of the trajectories was carefully monitored and first 300 ps were removed from the analysis. Since there is no structure available for the I47V mutant, 1KZK structure was taken as a template and the Ile47 (and 47') was replaced by Valine.

TH-772_04610604

3.3. Results and Discussions:

There are a number of attempts to understand drug resistance of HIV-pr using computational techniques. They differ in the particular mutation and particular inhibitor studied. Also, the lengths of molecular dynamics simulation covered a wide range (600 ps in reference [Wittayanarakul *et al.* 2005] to 22 ns in reference [Perryman *et al.* 2004]). Analysis of the molecular dynamics trajectories was also performed in a variety of ways. The main purpose of the present work is to understand the direct and indirect effects of I47V mutation. The following analysis is done keeping that in mind. In particular, the opening of flaps and protein-ligand movements are considered in the analysis.

3.3.1. Stability of the trajectories:

The stability of the trajectories of all four simulations was monitored by plotting the rmsd values of the C_α atoms as shown in Figure 3.2. It can be seen that the rmsd of all the trajectories are similar from the starting structures during the course of the simulations with values around 1.2-1.6 Å ensuring stable trajectories.

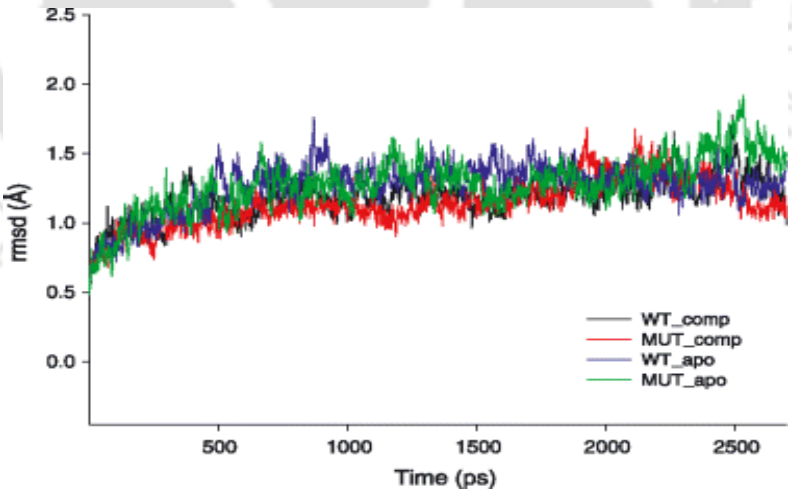


Figure 3.2: *RMSD values for the C_α atoms for the WT and mutant simulation for both apo and complexed forms.*

TH-772_04610604

3.3.2. Comparing the apo proteins: WT vs. Mutant

Figure 3.3 shows the difference of isotropic temperature (B) factor between mutant and WT for each residue. The difference of B factor can give idea about the structural fluctuation of different regions of WT and mutant. The maximum changes in B-factor between WT and mutant are for the residues in the flap elbows of the two chains (34-35, 37, 35'-41'), flaps of the two chains (48-51, 46'-54'), part of the cantilever region (65-70, 65'-68') though there are several other regions where B-factor differs. Thus from the B-factors, fluctuation of several residues especially the ones at the flap tips, flap elbows and cantilever region seems different between WT and mutant. In the next section we investigate these in more detail.

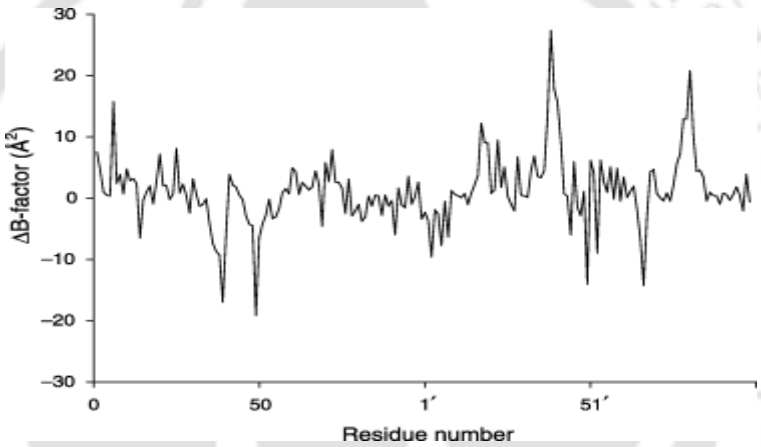


Figure 3.3: *Difference of B-factor values from molecular dynamics (MD) simulation for WT and mutant HIV-pr simulation of the apo protein (mutant B-factor–WT B-factor).*

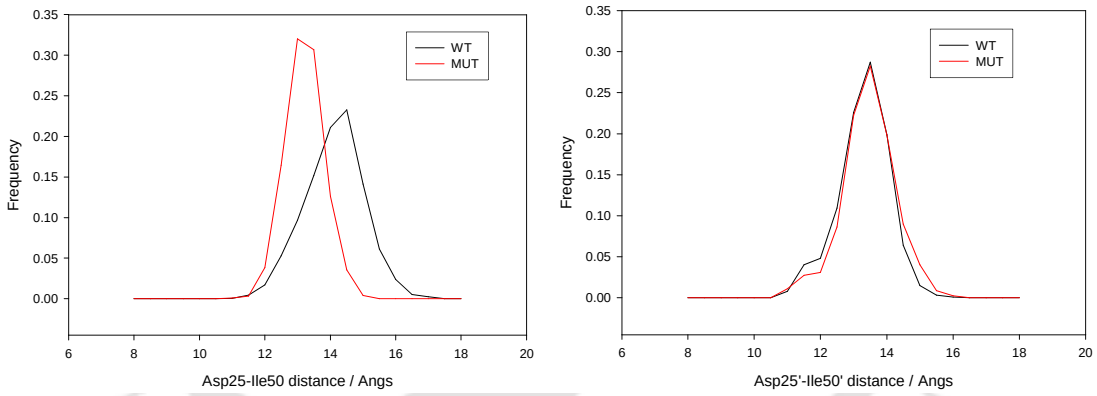
3.3.2.1. Local fluctuations:

In this section, local structural differences between WT and mutant are investigated. Especially the flap movement is investigated in detail. It is known that flap dynamics affects both drug binding and enzyme catalysis of HIV-pr. Several mutations affect dynamics of flap. For instance, L90M, G48V, V82F/I84V mutations open the flap more in the mutant than the wild type. On the other hand, M46I mutation makes the flap more closed. Recently a very detailed work by McCammon and co-workers on V82F/I84V mutant used a variety of parameters to find the flap dynamics. We have used some of those to enquire the extent of flap motion.

TH-772_04610604

3.3.2.2. Flap tip to active site distance:

The distance between the flap tip (C_α of Ile50(50')) and catalytic aspartates (C_α of Asp25(25')) was measured from the simulation and the distributions of the same are shown in Figure 3.4(a) and 3.4(b) for chain A and B respectively. As seen from the figure, the distance between flap tip and catalytic site in chain A is clearly different for WT and mutant. The mean of WT distribution is 14.6 Å and standard deviation is 1.0 Å. For the mutant, the mean and standard deviation are 13.5 Å and 0.6 Å respectively. So, the mean of these two distributions differ by more then 1.0 Å and WT distribution covers wider values. However, for chain B, the distributions are almost overlapping. The simulation results suggest that for chain A, the average flap tip to active site distance is less in the case of mutant. The difference in motion of chain A and chain B of HIV-pr has been observed in other simulation also [Wittayanarakul *et al.* 2005].



(a) (b)

Figure 3.4: *(a) Distributions of Asp25–Ile50 distance for both wild type (WT) and mutant HIV-pr simulation of the apo protein; and (b) distributions of Asp25'–Ile50' distance for WT and mutant HIV-pr simulation of the apo protein.*

3.3.2.3. Ile50-Ile149 distance:

The distance between C_α of Ile50- C_α of Ile149 is shown in Figure 3.5. This will give the distance between the flap tips in the two chains. This quantity will can shed light on the difference of flap motion between WT and mutant. It is clearly seen from the figure that for the mutant the distribution has one peak around 8 Å, while for the WT the main peak is distributed around 7 Å and two other peaks around 9 and 6 Å. The mean and standard deviation of WT distribution are 7.7 and 0.9 Å and that of mutant are 8.2 Å and 0.7 Å

TH-772_04610604

respectively. While there is considerable overlap between the two distributions the distance between the flap tips are fluctuating more in the case of WT than that of the mutant.

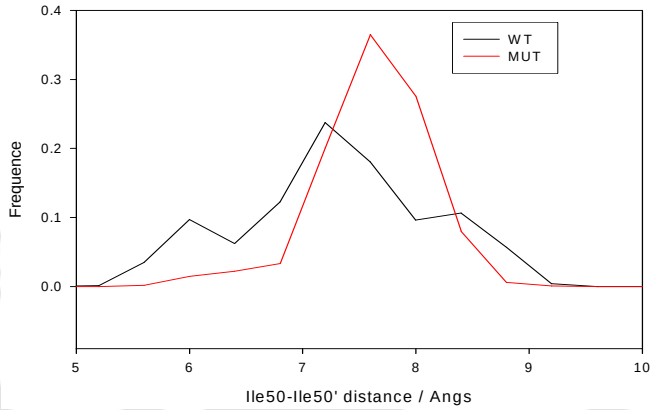


Figure 3.5: *Distributions of Ile50–Ile149 distance for WT and mutant HIV-pr simulation of the apo protein.*

Phe53'(C_α)-Cys67'(C_α) distance was also monitored (not shown in the figure) since there are large differences in the B-factors in those residues. The distributions are found to be different with some overlap. Though the B-factor is quite different for 53' and 67', the distance between them is not drastically different for WT and mutant. This is because the WT B-factor is more for 53' but less for 67', so the balance between these makes 53'-67' average distance not drastically different for WT and mutant. Thus both from differences in B-factor and analysis of several parameters, we can conclude that the results of simulations indicate the dynamic motion of WT and mutant are different in certain residues. In particular, the distances between residues in the flap elbow and flap tips are different in the case of mutant and WT.

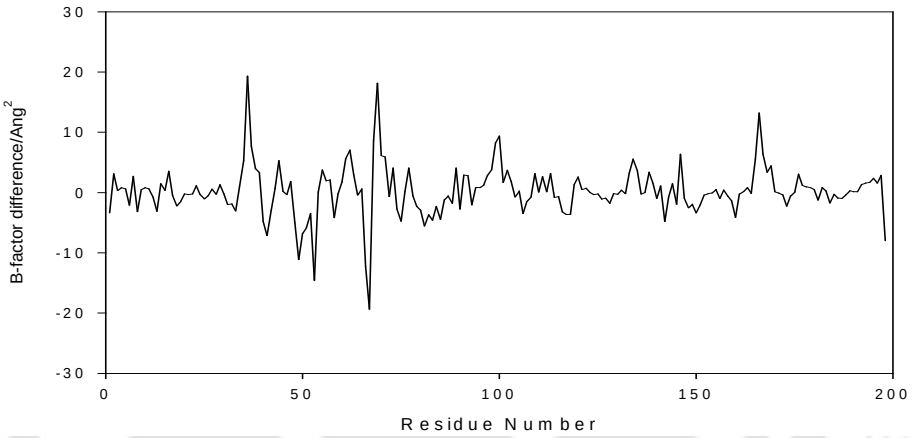
3.3.3. Comparing the complexed form of the protein: WT vs. Mutant.

3.3.3.1. B-factor analysis:

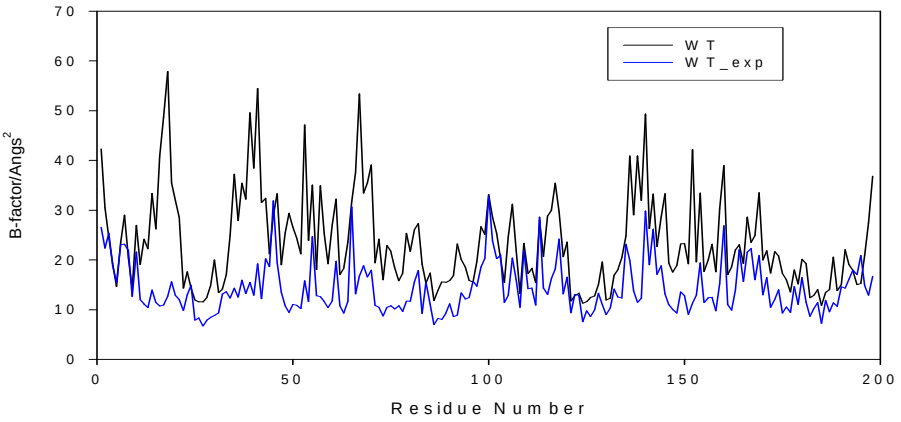
The difference of temperature factors (B-factor) of the protein in its complexed form is shown in Figure 3.6(a). It can be seen that compared to the apo protein, the difference between WT and mutant has reduced for most of the residues. However, there are significant difference for the residues 36-37 (flap elbow of chain A), 48-52 (flap tips of

TH-772_04610604

chain A), 66(66')-69(69') (part of the cantilever regions in both chains). Figure 3.6(b) shows a comparison between X-ray structure B-factor and calculated B-factor for the WT. It can be seen that the calculated B-factors are much higher than that of the X-ray structure although the general pattern is mostly similar with few exceptions. It is to be noted that for high-resolution crystal structures (such as the one used here) the calculated B-factors can be much higher than experimental one. This is because the high-resolution crystals are usually very well packed, which is difficult to achieve in simulation done in pure solvent.



(a)



(b)

Figure 3.6: *(a) Difference of B-factors for wild type (WT) and mutant complexed HIV-pr (mutant B-factor–WT B-factor); (b) B-factor for the complexed HIV-pr (calculated and experimental results for the WT).*

TH-772_04610604

Analysis of Local fluctuations of inhibitor complexed WT and mutant:

Next several key local motions were investigated. Three quantities out of those are discussed in the following (a) Asp25(25')-Ile50(50') and Ile50-Ile50' distances which indicate the movement of the flap and flap-active site movement, (b) Asp25(25')-inhibitor distance, which will be an indicator of the protein-ligand motion, (c) dihedral angles of Ile-47' and the P2' moiety of the JE-2147, which will shed light on the orientation of the protein and inhibitor.

3.3.3.2. Asp25(25')-Ile50(50') distance and Ile50-Ile50' distance:

The Asp25(25')-Ile50(50') distances were monitored during the course of the simulation to compare with observed differences found in the case of apo proteins. It was found that for chain A the distribution is much narrower compared to the apo protein. Moreover, the WT and mutant distributions have significant overlap. This indicates that in the ligand bound state, the distance between flap tips and the active site does not differ significantly on mutation for chain A. For chain B, there is more difference between the distributions compared to chain A. To know the relative motion of the flap tips Ile50-Ile50' distance was also monitored. Again it is much narrower than the apo form, and the difference between WT and mutant is much smaller.

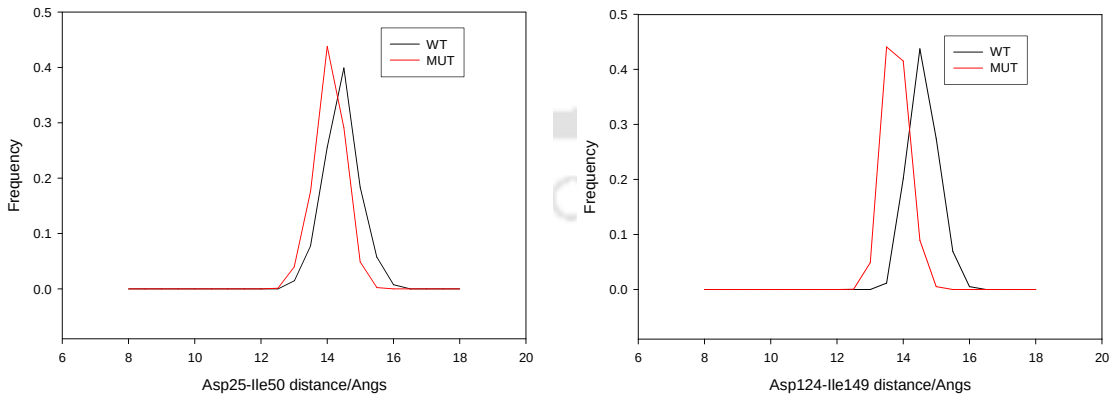


Figure 3.7: *Distributions of Asp25-Ile50 and Asp124-Ile149 distances for both WT and mutant simulation for the complexed protein.*

TH-772_04610604

The results of this section indicate that though Ile-Ile distance is similar in the complex there is still difference in the Asp-Ile distance in chain B. These three distributions are shown in the following Figures 3.7 and 3.8.

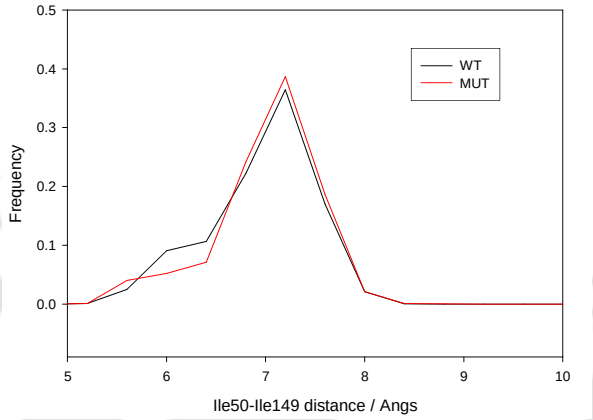


Figure 3.8: *Distributions of Ile50-Ile149 (flap-flap) distance for the WT and mutant for the complexed protein.*

3.3.3.3. Protein-ligand distance:

In previous work [Piana et. al. 2002b], it has been shown that the displacement of the substrate to the active site of the protein is coupled with the complex motion of the entire protein. We have used the definition given by Rothlisberger group [Piana et. al. 2002b] to get the protein-ligand distance in a simple manner; this is the average distance between the C_α of two catalytic Asp (Asp25 and Asp25') and the two carbons of the ligand (C_1 and C_2 in figure 2). The distributions of the average of these four distances are shown in Figure 3.9.

This shows that the distributions are essentially same for WT and mutant, which indicates that the ligand is bound strongly to the catalytic aspartates and mutation does not have any significant effect.

TH-772_04610604

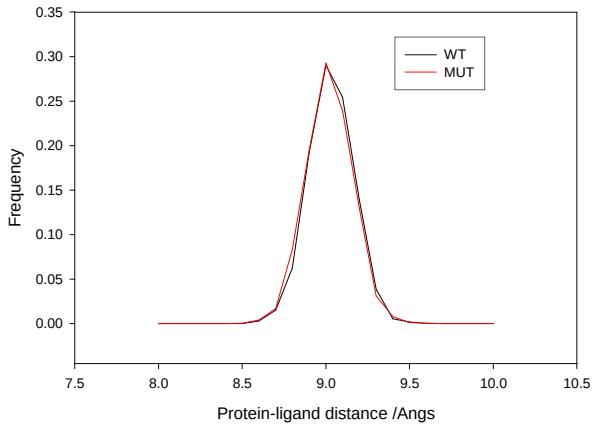
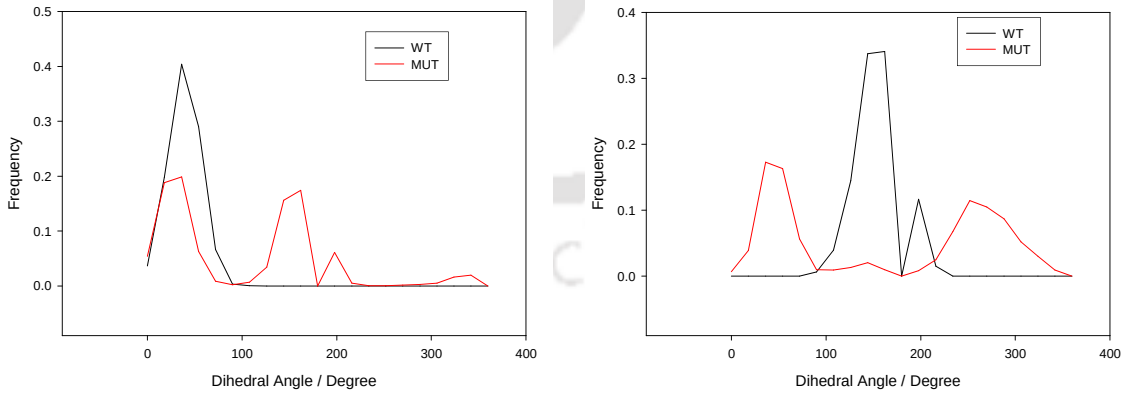


Figure 3.9: *Distributions of protein–inhibitor distances (see text) for both wild type (WT) and mutant HIV-pr simulation.*

3.3.3.4. Dihedral angle reflecting the orientation of protein and ligand:

Residue Ile47' in the WT HIV-pr interacts with the P2' site of the inhibitor in the crystal structure. The loss of such molecular interactions may be a key direct effect of mutation. To explore this, two dihedral angles involving C_β–$C_{\gamma 1}$ ($C_{\gamma 2}$) of Ile-47' (and Val47' for the mutant) and C_3–C_4 of JE-2147 were monitored by simulations to give an indication of relative orientation of the P2' group of the inhibitor with respect to the side chain of the 47' residue of HIV-pr.



(a) **(b)**

Figure 3.10: *Distributions of dihedral angles used to obtain orientation of P2' moiety of the inhibitor relative to the 47' side chain for both WT and mutant. (a) CG1 and (b) CG2 dihedrals.*

TH-772_04610604

Distributions for the two angles (henceforth denoted by CG1 andCG2) indicated significant differences between the WT and mutant HIV-pr (Figure 3.10).

The dihedral angle measurements for mutant HIV-pr showed a greater fluctuation when compared with that of WT, including values both in the regions near 0° and 360°. The mean and Standard Deviation of the WT dihedral angle were 36.5° and 13.9°, respectively, for CG1 and 146.4° and 17.5° for CG2, respectively. The mean and Standard Deviation of the mutant dihedral angles were 134.9° and 138.1°, respectively, for CG1 and 237.6° and 72.2°, respectively, for CG2. Thus, the relative orientation of residue 47' changes more rapidly and with a wider range of angles relative to the P2' group in the case of mutant HIV-pr. This results from of the loss of predominantly hydrophobic interaction between the P2' group of JE-2147 and mutant HIV-pr which effects increased mobility for the inhibitor. Both WT and mutant HIV-pr were analyzed for movement of water molecules, and simulations revealed that only one water molecule came within the 3 Å from the center of the four atoms chosen to define the dihedral angle. Therefore, the effect of water in the direct interaction between 47' residue of HIV-pr and P2' of JE-2147 is predicted to be minimal.

3.3.4. Overall molecular analysis of I47V HIV-pr resistance to JE-2147

In this study, drug resistance of 147V HIV-pr against an experimental inhibitor, JE-2147, was investigated to understand the effect of this mutation at the molecular level. For this purpose, four different simulations involving apo and complexed proteins (with and without the 147V mutation) were performed representing all atoms of the protein and inhibitor with thousands of explicit water molecules. The quality of the simulations were ensured by the rmsd values and by comparing with the B-factor obtained from X-ray for the complexed form of the WT. The results show that for the apo protein, there is difference in dynamic motion of WT and mutant involving the residues in the flap elbow and flap tip regions. In the case of the complexed HIV-pr, the difference in dynamic motion between WT and mutant was less than that of the apo protein relative to the motion of residues in the flaps and flap elbows. The average protein–ligand distance is

TH-772_04610604

not affected by mutation. The most distinct motion for the HIV-pr complex was the movement of the side chain of Val47' about the inhibitor and this was greater than for Ile47' in the WT. Hence, such results correlate to losing one $–CH_2$ group in the side chain between mutant and WT proteins, and consequent decreased hydrophobic interactions of Val47' mutant HIV-pr for inhibitor binding in the case of JE-2147. Such findings suggest that a larger group at the P2' position JE-2147 might provide recovery of such binding properties. Furthermore, the Val47' mutation causes a change in the dynamics of the flaps and such disruptions are likely to play a role in the binding and hence to the resistance.

3.3.5. Comparison of I47V with other mutants:

A comparison of the 147V mutation with other studies focused on different mutations is important to further interpret the findings described in this study. In this regard, we have considered the mutations G48V (flap region) and I84V (active site) to compare with I47V. The G48V mutation has been found for the clinical inhibitor Saquinavir (SQV), whereas I84V mutation exists for all clinical inhibitors. It is known that the major direct effect of G48V mutation is increased steric effect between protein and inhibitor, whereas the I84V mutation affects the dynamical motion of the protein.

The highly conserved flap tips of HIV-pr I47–G48–G49–I50–G51–G52–F53 are extremely flexible. Nuclear magnetic resonance (NMR) studies by Torchia and colleagues [Ishima *et al.* 1999] have shown that residues 49–52 have rapid motion in <10 ns time scale and fluctuation of F53 is likely coupled to the motion of the entire flap. Such G48V simulation studies [Wittayanarakul *et al.* 2005] have determined that there is difference in rmsd for the residues in this region upon mutation. It is noted that these reported [Wittayanarakul *et al.* 2005] simulations were performed with complexed HIV-pr, whereas in the present study simulation were conducted with both apo and complexed HIV-pr. It is likely that simulation studies of apo forms of G48V mutation would show more differences in flap motion, and such a finding has been verified in a recent report [Hamelberg *et al.* 2005] wherein accelerated MD calculations showed decreased flap motion in the G48V (apo form) to be substantially different from the WT. In the present study, we conclude both direct effects (VDW and steric for residues 47 and 48,

TH-772_04610604

respectively) and indirect effects (change of flap dynamics) are functionally involved. For the I84V mutation, the key effect is the dynamical motion of the protein. An examination of the V82F/I84V double mutant shows that it becomes more flexible (the change is much more than the I47V mutant). This would likely compromise inhibitor binding, since a large enthalpic cost would be necessary to close the mutant flap.

3.4. Conclusion:

1) The loss of one $-CH_2$ group in case of mutant Ile→Val side chain decreases in hydrophobic interactions of Val47' mutant HIV-pr for inhibitor binding.

2) Positioning a larger group at the P2' position of JE-2147 might provide recovery of the loss of –CH2 group while going mutant from Ile to Val and may increase the binding properties between the ligand-receptor to diminish the resistance. (As shown in Figure 3.11).

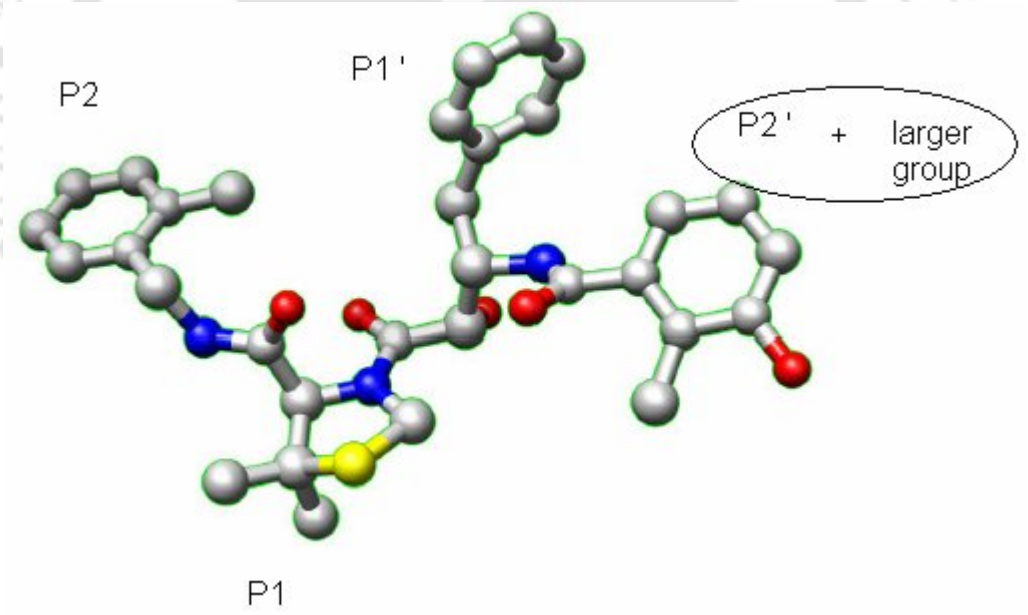


Figure 3.11: *Region showing the P2' Position in JE-2147, where the larger group can be added.*

3) A generalization has been made by comparing the I47V mutants with others stating I47V have both direct and indirect effects on drug resistance.

TH-772_04610604

Concluding Remarks:

Unquestionably, the understanding of the molecular basis of drug resistance is a complex task. The specific details obviously depend on position, mutation and the particular inhibitor involved. Nevertheless, the outcome of this study as well as previous studies on HIV-pr provides evidence for both direct effects (i.e. protein residues in the active site) and indirect effects (e.g. protein residues in the mobile flap region).

TH-772_04610604

CHAPTER 4

Pressure induced conformational dynamics of HIV-1 protease: A Molecular Dynamics Simulation study

TH-772_04610604

Pressure induced conformational dynamics of HIV-1 protease: A Molecular Dynamics Simulation study.

4.1. Introduction:

HIV-pr, an essential enzyme for HIV replication, is an important target for drug design strategies to combat AIDS. It acts at the late stage of infection by cleaving the Gag and Gag–Pol polyproteins to yield mature infectious virions. If the HIV-pr cannot ensue with its activity, the viral replication will be stopped and that will lead to the formation of immature virions. Keeping that in mind several drugs has been developed and also approved by the FDA for the treatment of AIDS. Unfortunately, resistance to these drugs has been built up quickly. Thus, understanding the drug resistance mechanism and knowledge of designing potent HIV-pr inhibitors with novel mechanisms of action is most demanded.

Structurally, HIV-pr is a homodimer. The active site region of the protein is covered by two glycine rich, antiparallel β-hairpins refereed as flaps. The volume of the active site and the accession of ligand to the active site are controlled by the dynamics of the two flaps [Piana *et al.* 2002a, Piana *et al.* 2002b]. Conformations of the protein determine its functions. The conformational dynamics of HIV-pr is known to be essential for the ligand binding and determination of cavity size, which changes with several common physiological parameters like temperature, pressure, and pH conditions [Kovalesky *et al.* 2005]. Also the effect of the protein backbone mutations can't be ruled out [Bandyopadhyay and Meher 2006]. The change in the conformation of the protein leads to the decreased binding affinity of the drugs and eventually escort to the drug resistance phenomenon.

Pressure plays an important role in the conformation of the proteins as revealed from a number of experimental studies [Hilson *et al.* 1999, Kharakoz *et al.* 2000, Paci *et al.* 2002]. These studies show that proteins behave somewhat differently under high pressure. Molecular dynamics simulation studies of protein under high pressure have

TH-772_04610604

shown the positional shift of absolute minimum of free energy resulting in a reconfiguration of the free energy landscape. At modest pressure, proteins remain in their native state and show elastic behavior indicating reversibility. At higher pressures exceeding 200–500 MPa the absolute minimum is shifted noticeably [Paci *et al.* 2002] with a set of completely different conformations, i.e, the high pressure denatured state. The pressure effect on protein denaturation is quite different from temperature dependent protein denaturation. Pressure-induced structural changes have been reported for deoxymyoglobin [Yamato *et al.* 1993], lysozyme [McCarthy *et al.* 2006a, Paci *et al.* 1996, Akasaka *et al.* 1997, Refaee *et al.* 2003], BPTI [Wroblowski *et al.* 1996], myoglobin [Urayama *et al.* 2002], apomyoglobin [McCarthy *et al.* 2006b], ubiquitine [Kitahara *et al.* 2005], Arc repressor [Trzesniak *et al.* 2006], HPr [Canalia *et al.* 2004], and various other proteins [Heremans *et al.* 1998, Gruner *et al.* 2004, Hummer *et al.* 1998] . HIV-pr is being studied under normal temperature and pressure conditions *in vitro* and *in silico* as well. In comparison, no report has been filed investigating the dependence of protein dynamics and conformation on high pressure for HIV-pr. It has been shown that, under high pressure the hydrophobic core of the protein bulges out to the solvent exposed area, which eventually changes the dynamics and hence conformation of the protein [Hummer *et al.* 1998]. HIV-pr has several hydrophobic residues in its most important regions; i.e. flap and active site region. So to illustrate the pressure induced conformational changes of HIV-pr and hence the possibilities of emergence of drug resistance behavior, we performed two simulations in the normal 1bar pressure and in a high 3Kbar pressure conditions. We observed that the volume of the protein shrunk with respect to the increase in pressure and the secondary structures have substantial alteration at high pressure with respect to the simulation time. As expected in accordance to the previous studies, we didn't find considerable changes in the hydrophobic core region, rather the hydrophobicity first increased and eventually decreased at high pressure. Interestingly the rate of change of hydrophobic SAS and total SAS seems to be more in the case of high pressure 3 Kbar than in 1 bar system.

TH-772_04610604

4.2. Computational Details:

The crystal structure of the wild type HIV-pr (PDB code 1HHP; 2.7 Å resolutions) [Spinelli *et al.* 1991] with semi-open conformation was used as the starting structure for the MD simulations. All missing hydrogen atoms were added using the LEaP module of the AMBER 8 program package [Case *et al.* 2004]. The *ff99SB* [Hornak *et al.* 2006] force field with TIP3P [Jorgensen *et al.* 1983] water models was used for both the 1bar (1 atm = 1.01325 bar = 101.325 kPa) and 3Kbar simulations. The system was solvated with the TIP3P waters in the periodic box of size 83.5 x 62.5 x 68.9 $Å^3$ containing more than 8000 water molecules. A cutoff of 10 Å was used along the three axes to discard any water molecule if it is farther than the cutoff from any solute molecule. An appropriate number of Cl^- counterions were added to neutralize the net positive charge on the system. The protonation states of catalytic aspartates (Asp-25 and Asp-124) vary depending on the binding of inhibitors or substrates. As the starting model in this simulation is a ligand free structure, the protonation states of both the catalytic aspartates were kept deprotonated. A cutoff of 8.5 Å was used for Lennard-Jones interactions, and the long-range electrostatic interactions were calculated with the particle mesh ewald (PME) method [Essman *et al.* 1995]. Constant temperature and pressure conditions in the simulation were achieved by coupling the system to a Berendsen's thermostat and barostat [Berendsen *et al.* 1984]. The SHAKE procedure was used to constrain all bonds involving hydrogen.

The system was minimized in two phases. In the first phase, the system was minimized giving restraints (30kcal/mol/$Å^2$) to protein and crystallographic waters for 500 steps with subsequent second phase minimization of the whole system. Then the system was heated to 300K over 25 ps with a 1 fs time step. The protein atoms were restrained with force constant of 30 kcal/mol/$Å^2$ at the NVT ensemble. After that the force constant was reduced by 10 kcal/mol/ $Å^2$ in each step to reach the unrestrained structure in three steps of 10 ps each. The system was then switched over to the NPT ensemble and equilibrated without any restraints for 180 ps. The system was equilibrated in total of 235 ps in 1 bar. The convergence of energies, temperature, pressure and global RMSD was used to verify the stability of the systems.

TH-772_04610604

Then the system was divided in to two sets. The first set was slowly submitted to increasing the pressure, whilst the second was kept at atmospheric pressure. The pressure in the first system was elevated in 22 concatenated runs during a total 2.2 ns as follows: 1–10 bar for 100 ps; 10–50 bar for 100 ps; 50-100 bar for 100 ps; 100–1000 bar (100 bar increase during 100 ps for each run); 1000–3000 bar (200 bar increase during 100 ps for each run). The second system was allowed to run for an equivalent 2.2 ns MD simulation period at 1 bar. Being achieved this point both the systems were allowed to evolve during one ns each, in order to produce comparably equilibrated starting points for the low and high pressure systems. After this both the 1 bar and 3Kbar trajectories were run for 30 ns. The time step for MD production run was 2 fs. The long 30 ns trajectories were used to calculate the average structure for both the low and high pressure systems. Analysis of both the MD trajectories i.e., RMSD, RMSF etc was carried out using the Ptraj module of AMBER 8. The secondary structure changes during the MD simulations was analyzed using the "Dictionary of Secondary Structure for Proteins (DSSP)" definitions [Kabsch *et al.* 1983] and was calculated using the do_dssp program from the GROMACS package. Also the solvent accessible surface (SAS) area of each residue was calculated and analyzed by the program g_sas from the GROMACS package [van der Spoel *et al.* 2004]. Ramachandran plots for both the low and high pressure system were calculated using the PROCHECK program [Lakowski *et al.* 1993]. Graphic visualization and presentation of protein structures were done using Chimera [Pettersen *et al.* 2004].

4.3. Results and Discussions:

4.3.1. Mobility Analysis: RMSF

In order to get the information about the relative mobility of the atoms of protein at normal and high pressure conditions, we have analyzed the root mean square fluctuations (RMSF) of α-carbon atoms of all the residues (Figure 4.1).

TH-772_04610604

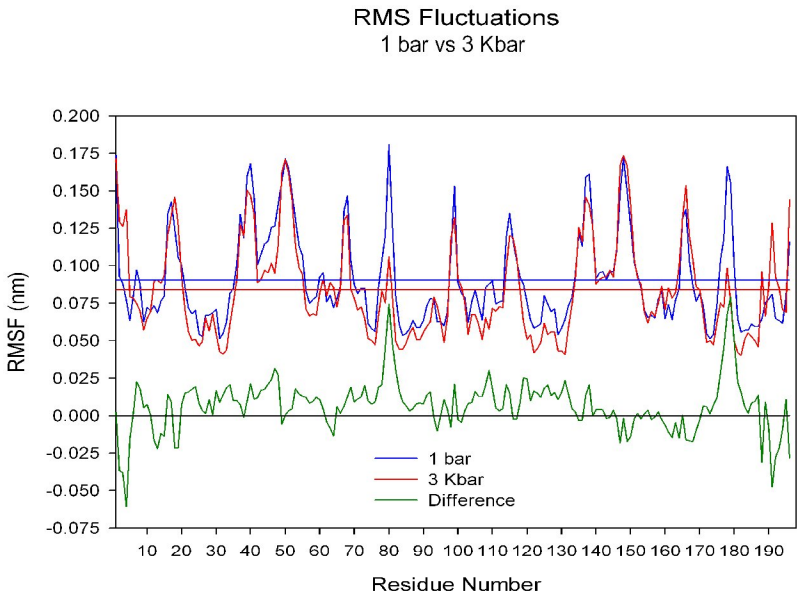


Figure 4.1. *RMSF of α-carbon atoms at 1 bar (Blue) and 3 Kbar (Red) and there difference (Green).*

It was observed that with increase in pressure the average residue mobility is significantly lowered. The blue and red straight lines represent the average RMSF for the 1 bar and 3 Kbar pressure conditions respectively. The average Cα atom mobility at 1 bar is about 1.078 times greater than that at 3 Kbar. This is mostly true for all α-carbons with the noteworthy exceptions of residues like 9, 25, 26, 29, 31, 37, 39, 61, 67, 111-113, 140-142, 152,154, 158 and 171-173. On the other hand, the most notorious reduction in mobility is observed for Thr-80 & 179 in the active site wall region.

4.3.2. Root Mean Square Deviation (RMSD) analysis:

To examine the possible structural similarities and variations during simulations, we have computed the RMSD. The equilibrated structure at 3.435 ns was taken as reference structure to calculate the α-carbon RMSD for both low and high pressure simulations. Figure 4.2 shows the RMSD for both the 1 bar and 3 Kbar systems. It can be seen that the 1 bar system has reached a point of reasonable stability around the value of 1.35 sigma units. Whereas the high pressure 3 Kbar system shows an unstable behavior and appears to be more variable when compared with 1 bar system.

TH-772_04610604

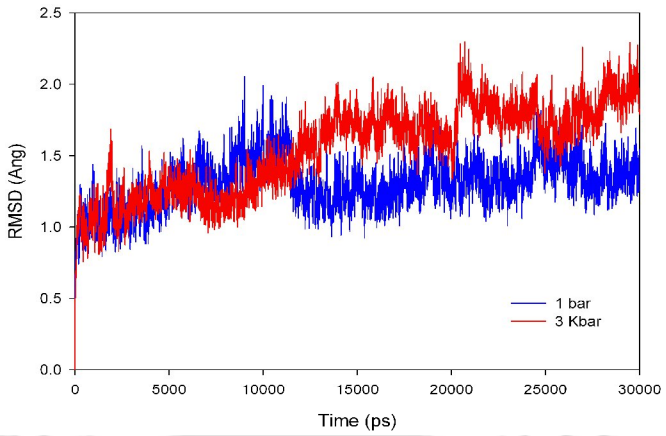
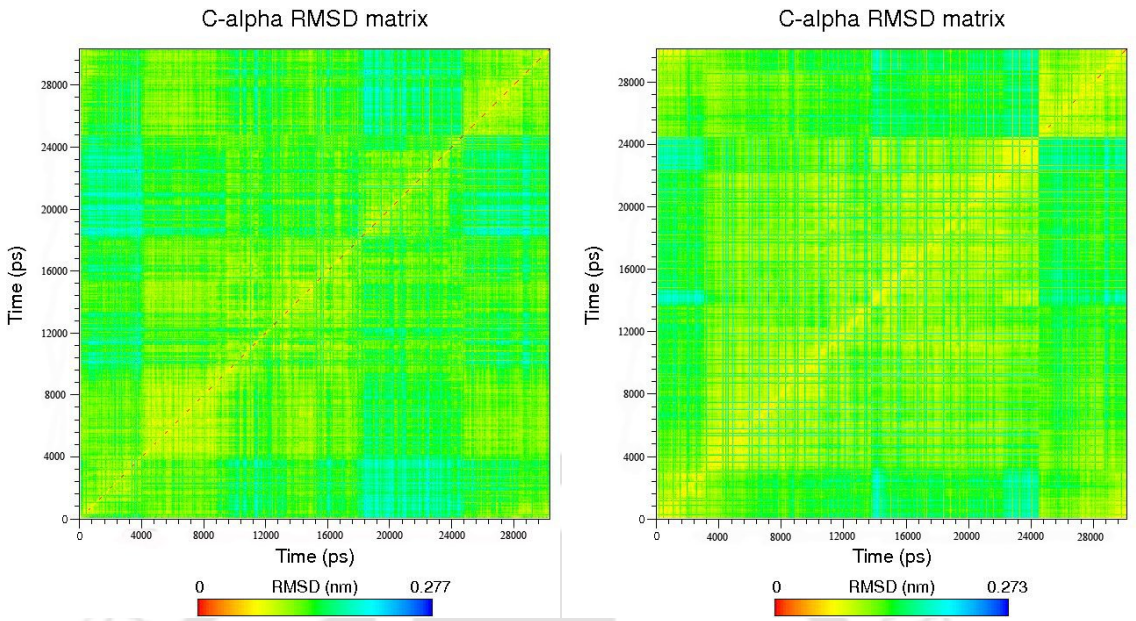


Figure 4.2. *RMSD of α-carbon atoms at 1 bar (Blue) and 3 Kbar (Red) for the entire simulation period.*

As a whole we can conclude that 1bar simulation has reached a reasonably stable conformation but structures at 3 Kbar finding difficult to stabilize and thus do not tend to repeat themselves. Figure 4.3 (a) and 4.3 (b) shows RMSD matrix plots for the entire 1 bar and 3 Kbar systems with 30 ns simulation time which gives more detailed description of these issues. Here the RMSD value was calculated pair wise, covering every recorded frame combination. The results are presented in a color code format. The bar shows the color/nanometer uniformity. For 1 bar system, the average fluctuation in RMSD is well below of 2.1 as we can see in orange-yellow regions (close to slope = 1). However light blue regions are spread to specific regions (yellow=1.15 sigma units; green=2.3 sigma units; light blue=3.45 sigma units) signifying the flexible regions. For system with 3 Kbar pressure condition, as shown in fig 4.3 (b) the effect of pressure dominates as the yellow regions are more prominent in the central region and the light blue regions are lesser than the 1 bar. It signifies the pressure effect in regions which shows local irreversibility. In general the blue regions are significantly less in 3Kbar simulation and they appear only near the edges, indicating that the structures at the initial period and the final period are significantly different from each other while between 4ns to 24ns the structures were almost conserved.

TH-772_04610604



(a) (b)

Figure 4.3: *RMSD matrix plot for the entire (a) 1 bar and (b) for 3 kbar system with 30 ns simulation time. Here the RMSD value was calculated pair wise, covering every recorded frame combination and are presented in a color code format. The bar shows the color/nanometer uniformity. Color can be interpreted in sigma units as: Red=0 sigma units; yellow=1.15 sigma units; green=2.3 sigma units; light blue=3.45 sigma units.*

On the other hand, the structures for 1bar system have more frequent transitions as reflected by many small blue patches. This indicates that structures at high pressure are more stable than in 1 bar pressure condition. Literature searches for other proteins RMSD and comparing the present results testify that HIV-pr is stable enough to high pressure. As a whole with increase in pressure, the global protein mobility is reduced and the local flexibilities are increased which is evident from lesser blue regions and higher yellow regions in the 3 Kbar system.

4.3.3. Secondary Structure analysis:

The secondary structural changes during the whole 30 ns MD simulation period was analyzed using the DSSP definitions and was calculated by the do_dssp program. A detailed comparative analysis of the secondary structural changes (Figure 4.4 upper and 4.4 lower) during the simulations is outlined below:

TH-772_04610604

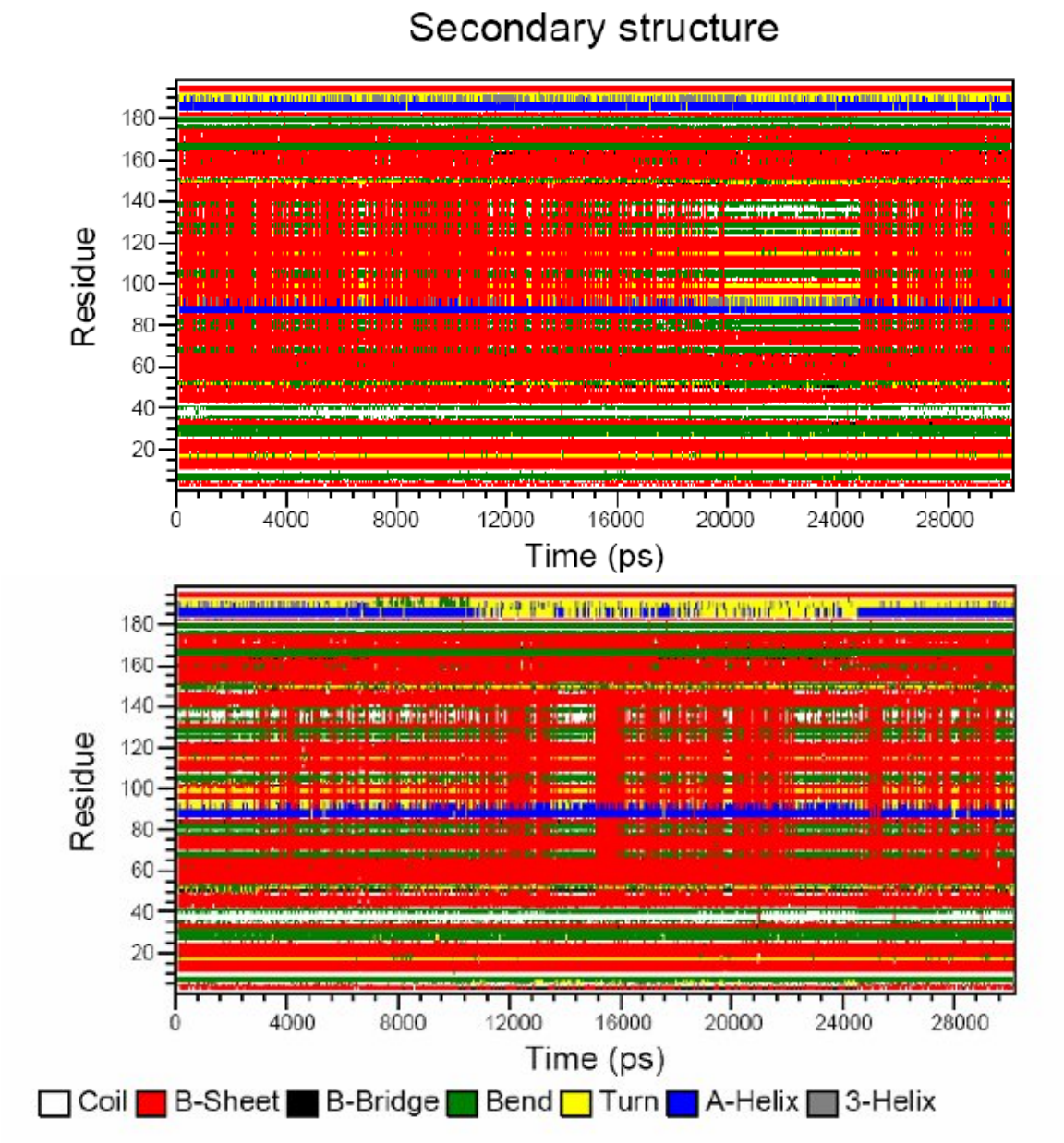


Figure 4.4. *Secondary structure matrixes for the entire 1 bar (Upper) and 3 kbar (Lower) system simulation times of 30 ns. Color code may be interpreted from the legend.*

Residues 1-4 (No secondary structure assigned by DSSP): Unchanged from 1 bar to 3 Kbar in the initial but appearance of β-bridge in later.

Residues 5-7 (Bend): Stable bend at 1 bar but fluctuate between bend, turn and coil for 3 Kbar.

TH-772_04610604

Residues 8-9 (No secondary structure assigned by DSSP): Unchanged from 1 bar to 3 Kbar.

Residues 10-14 (Extended β-strand): Stable β-strand unchanged from 1 bar to 3 Kbar.

Residues 15-18 (Turn): Turn with intermittent bend properties from 1 bar changes to stable turn at 3 Kbar.

Residues 19-24 (β-sheet): Stable β-sheet unchanged from 1 bar to 3 Kbar.

Residue 25 (No secondary structure assigned by DSSP): Oscillates between coil, turn and sheet.

Residues 26-31 (Bend): Stable bend at 1 bar. Unchanged from 1 bar to 3 Kbar.

Residues 32-34: Oscillates between β-sheets and coil. Mostly unchanged from 1bar to 3 Kbar.

Residues 35-42 (Bend/loop): Oscillates between bend and coil. Coil is more prominent than bends. Mostly unchanged from 1bar to 3 Kbar.

Residues 43-48 (β-sheet): Stable β-sheet unchanged from 1 bar to 3 Kbar.

Residue 49 (β-sheet): Mostly replaced by coil from 1 bar to 3 Kbar.

Residue 50-51 (Bend): Oscillates between bends and turns and appearance of coil in between moving from 1 bar to 3 Kbar.

Residues 52-66 (β-sheet): Stable β-sheet unchanged from 1 bar to 3 Kbar.

Residues 67-69 (Bend): Oscillates between bends and sheet at the initial time, but formation of stable bends around 20-25 ns at 1 bar. But for 3 Kbar stable bends are seen in the initial up to 12 ns and later occupied by intermittent β-sheet.

Residues 70-78 (β-sheet): Stable β-sheet unchanged from 1 bar to 3 Kbar.

Residues 79-83: Same as that of 67-69 residues.

Residues 84-86 (β-sheet): Stable β-sheet unchanged from 1 bar to 3 Kbar.

Residues 87-90 (α-helix): Stable α-helix unchanged from 1 bar to 3 Kbar.

Residues 91-93: Oscillates between helix, turn and sheets at 1 bar. Unchanged at 3 Kbar but α-helix occurrence is prominent as compared to turns and sheets.

Residues 94-95: Oscillates between β-sheet and turn with prominence around 20-25 ns for 1 bar. Unchanged from 1 bar to 3 Kbar, but turn is seen throughout the simulation time with little intermittent β-sheet for 3 Kbar.

Residues 96-97: Stable β-sheet unchanged from 1 bar to 3 Kbar.

TH-772_04610604

Residues 98-99: Same as that of 94-95 residues.

Residues 100-103: Same as that of 94-95 residues.

Residues 104-106: Same as that of 67-69 residues.

Residues 107-114 (β-sheet): Stable β-sheet unchanged from 1 bar to 3 Kbar.

Residues 115-116: Same as that of 94-95 residues.

Residues 117-123 (β-sheet): Stable β-sheet unchanged from 1 bar to 3 Kbar.

Residue 124: Oscillates between sheet and coil at 1 bar. Unchanged at 3 Kbar.

Residues 125-130: Same as that of 67-69 residues.

Residues 131-133: Same as that of 32-34 residues.

Residues 134-141: Oscillates between bends, sheets and coils. Bends and coils are prominent in 20-25 ns at 1 bar. At 3 Kbar bend, coil and sheet oscillates but coil and bend are prominent and seen throughout the simulation with intermittent β-sheets.

Residues 142-148 (β-sheet): Stable β-sheet unchanged from 1 bar to 3 Kbar.

Residues 149-151: Oscillates between bends, turns and coil. Turns is prominent at 1 bar and remain unchanged at 3 Kbar except the coil that seems prominent in the first 4 ns and 22-25 ns.

Residues 152-165: (β-sheet): Stable β-sheet unchanged from 1 bar to 3 Kbar except the residue 160 which oscillates between sheet and bend.

Residues 166-168: (Bend): Stable bend unchanged from 1 bar to 3 Kbar.

Residues 169-175 (β-sheet): Stable β-sheet unchanged from 1 bar to 3 Kbar.

Residues 176-181: Stable bends and coils unchanged from 1 bar to 3 Kbar.

Residues 182-183 (β-sheet): Stable β-sheet unchanged from 1 bar to 3 Kbar.

Residue 184: Coil at 1bar unchanged from 1 bar to 3 Kbar.

Residues 185-188 (α-helix): Stable α-helix at 1 bar. This part of the protein shows distinct differences between 1 bar and 3 Kbar by the occupancy of turns over helix from around 10-25 ns. The helix in the chain-B is lost during this period.

Residues 189-191: Oscillate between helix and turns throughout the simulation period at 1 bar. At 3 Kbar turns are mostly occupied with intermittent helices. Interestingly bends are seen around 8-10 ns.

Residue 192: Coil at 1bar unchanged from 1 bar to 3 Kbar.

Residues 193-196 (β-sheet): Stable β-sheet unchanged from 1 bar to 3 Kbar.

TH-772_04610604

Residues 197-198: No secondary structure assigned by DSSP.

Shortening secondary structure observations:

Comparative analysis of the secondary structures from 30 ns for both the normal and high pressure conditions, it is observed that the global changes in the protein conformation remains unchanged. Marked specific local changes seen with the expense of β-sheets and α-helices to turns and bends.

Residues 49-99 (in chain A) and residues 100-140 (in chain B) regions: This part of the protein shows distinct differences between 1 bar and 3 Kbar. Bends, turns and coils are formed in the expense of β-sheets and are more prominent around 20-25 ns for 1 bar. But in 3 Kbar the changes in structure seems throughout the simulation period indicating the structural instability under high pressure correlating the RMSD shown in Figure 6.2.

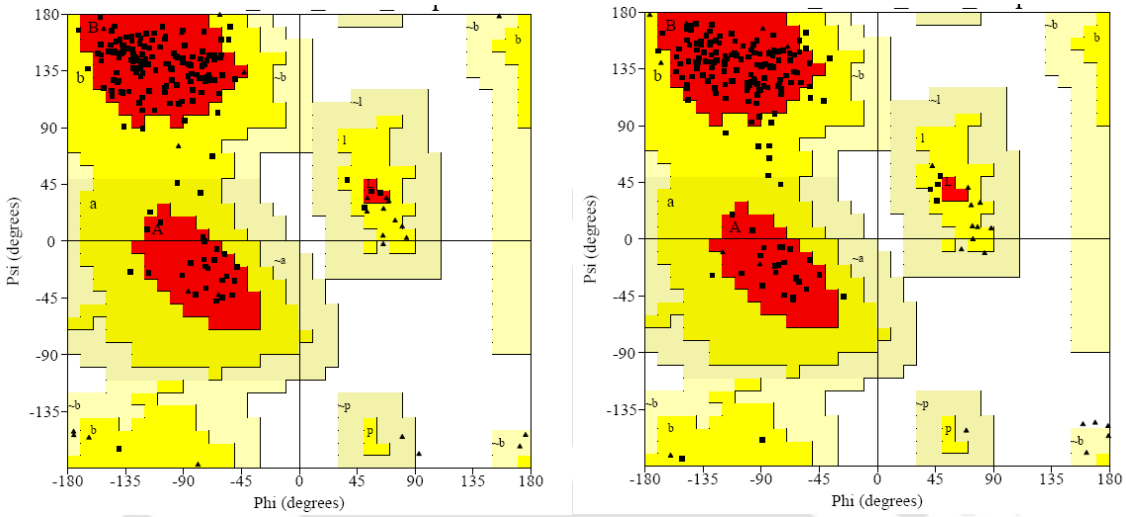
Chain-B helix (Residues 185-191): The most notable difference seen here by the occupancy of turns over helices from around 10-25 ns where the α-helix is being disrupted under high pressure.

4.3.4. Analysis of Ramachandran Plots:

To diagnose the pressure induced changes in stereochemical geometry of each residue, we analyzed the Ramachandran Plot for the average structure obtained from low pressure and high pressure trajectories. The data point (dihedral angles) on the Ramachandran plot for the average structures from both the low and high pressure trajectories doesn't show ample difference. But in the case of 3 Kbar high pressure system, few residues in Ramachandran plot shifts to additional allowed regions from most favored regions. This can be seen from Figure 4.5 (a) & (b). Also we can see the change in secondary structure (sheet to helix and vice versa) of protein from the Ramachandran Plot shown for the 1 bar and 3 Kbar average structures. Mostly the residues in between the sheet and alpha helix regions of Ramachandran Plot undergo changes in secondary structure with change in pressure. Apart from this, with increase in pressure, the residues within the most favored

TH-772_04610604

regions migrate towards the additional allowed regions and those residues in the additional allowed regions also shifts to the boundary that separates the additional allowed regions and disallowed regions. As a whole, the pressure change affects the conformation of each residue and also the stability of the protein structure.



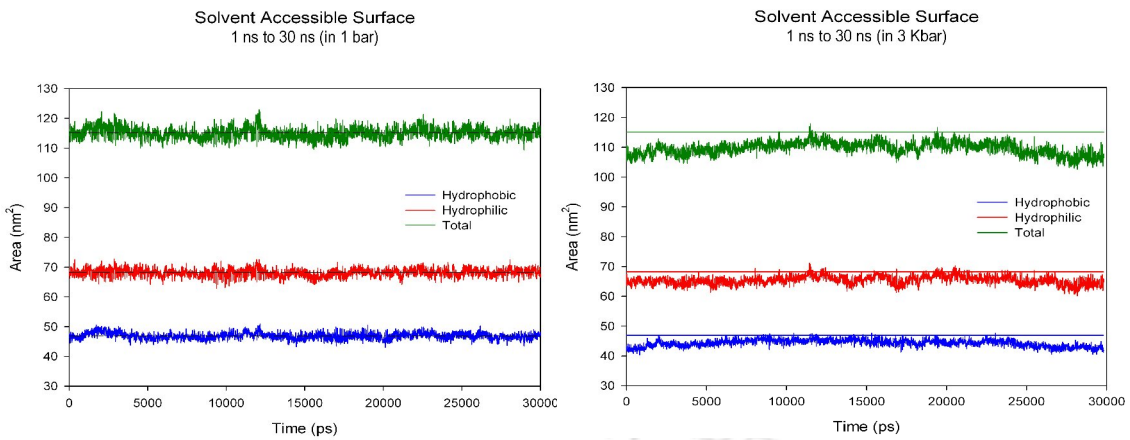
(a) (b)

Figure 4.5: *Ramachandran plot for the (a) 1 bar and (b) 3 kbar averaged structure to check the stereochemical quality of the protein.*

4.3.5. Solvent accessed surface (SAS) analysis:

In order to add-on our observations and conclusions extracted under RMSD and secondary structure analysis and discussion, we now proceed to analyze the protein solvent accessed surface area. Figure 4.6 (a) and (b) shows the solvent accessed surface area for hydrophilic, hydrophobic and total residues of the protein at 1 bar and 3 Kbar respectively. It was found that 1 bar system fluctuates reasonably around apparently stable SAS values. Whereas 3 Kbar system shows a somewhat stable behavior for the hydrophilic exposed surface, whilst the hydrophobic and total SAS shows increasing behavior first and then substantially decreasing behavior. But in no case, the 3 Kbar system never reaches above the 1 bar average. This may be accounted on the basis of restricted flap dynamics with high pressure. In addition to this, the rate of change of hydrophobic SAS and total SAS seems to be more in the case of high pressure 3 Kbar systems than in 1 bar system.

TH-772_04610604



(a) (b)

Figure 4.6: *SAS for the whole 30 ns simulation time at (a) 1 bar and (b) 3 Kbar. The dashed black line shows the average SAS for each type in (a) and the solid line in (b) shows the average SAS for each type from 1 bar, not from 3 Kbar.*

6.3.6. Radius of gyration (Rg) analysis:

Further to check the volume changes in protein structure with the change in pressure, the radius of gyration (Rg) was calculated by the g_gyrate program from GROMACS. It can be seen from Figure 4.7, that there is considerable reduction in volume of protein with pressure. This is due to the fact that the structure becomes compact with increase in pressure. Thus the overall spread of the protein molecule decreases with high pressure.

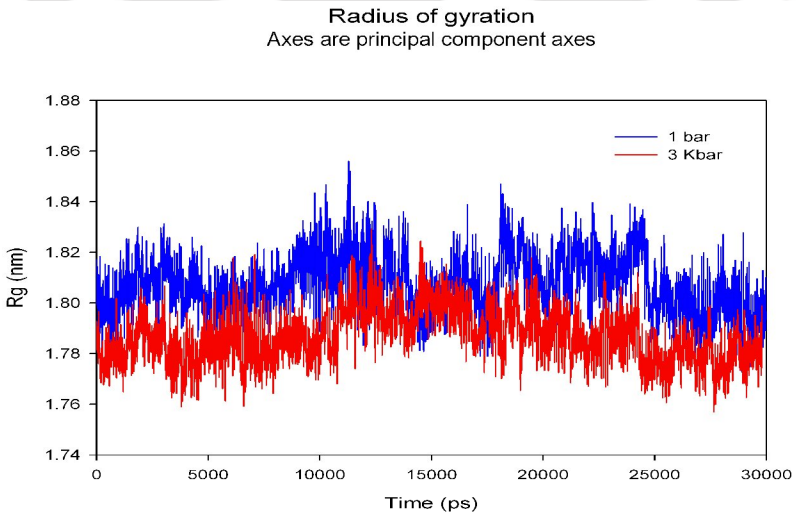


Figure 4.7. *Radius of gyration of α-carbon atoms at 1 bar (Blue) and 3 Kbar (Red).*

TH-772_04610604

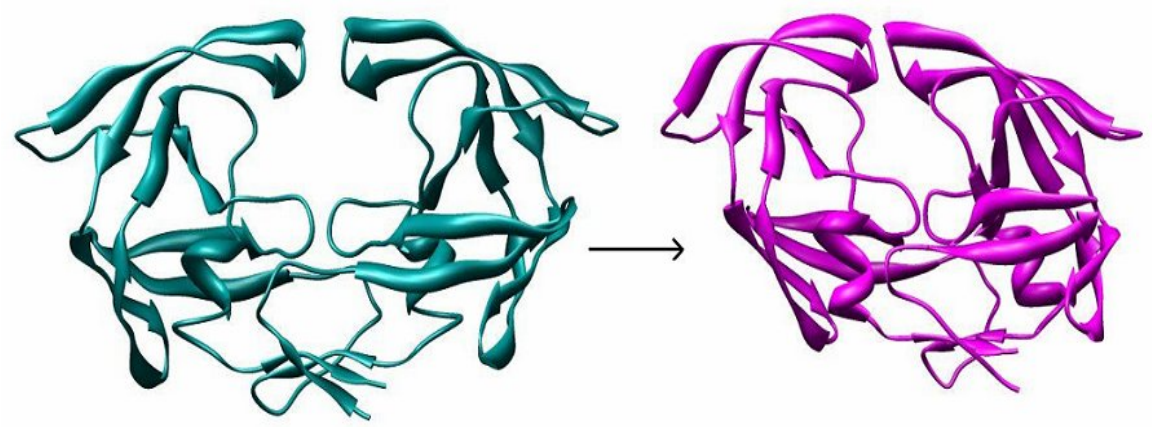


Figure 4.8. *Comparison of the superimposed average structures of 1 bar (Cyan) and 3 Kbar (Pink) from 30 ns simulation time.*

Figure 4.8 shows the superimposed average structures of protein at 1 bar and 3 Kbar conditions presenting clearly the compact deformed conformation of the protein with increase in pressure from 1 bar to 3 Kbar.

4.4. Conclusion:

Comparative study of the HIV-pr at 1 bar and 3 Kbar pressure, gives a clear picture of the conformational changes induced through high pressure. A short simulation time is not good enough to have a concrete idea upon the effect of pressure on the dynamics of protein. Nevertheless some interesting facts can be extracted from the partially evolved protein structure.

1) The present study reproduces not only the global behavior of HIV-pr, but also is able to account for anomalous behavior of regions in and around flaps and active site.

2) The results also demonstrate the increase in compactness of structure under high pressure that affects the secondary structure of the protein with reasonable changes in α-helix and β-sheets to turns and bends.

TH-772_04610604

3) Regarding the protein mobility study the results confirms the general decrease in the structural degrees of freedom of this protein under high pressure.

4) The difference in SAS with pressure increase shows that pressure acts as a unique conformer selector consequently reducing the structural variability.

5) We can also see that the active site region of the protein was shrunk to a greater extent with pressure. This may affect considerably the binding affinity of drugs to the active site region. Further investigation is required to throw light on this.

TH-772_04610604

CHAPTER 5

Molecular dynamics simulation of HIV-1 protease with Polarizable and Non-polarizable force fields

TH-772_04610604

Molecular dynamics simulation of HIV-1 protease with Polarizable and Non-polarizable force fields.

5.1. Introduction

Computer simulation of molecular and biomolecular systems has reached a stage, where problems of immense practical importance can be studied in chemical, biological and material science [Karplus and McCammon 2002, Behler *et al.* 2008]. In biological science, various phenomena at different length and time scales have been studied using different simulation techniques. Some of the interesting examples are simulation of ribosome, ion channel simulation, understanding the mechanism of bacterial flagella [Sanbomatsu *et al.* 2005, Chanda *et al.* 2005, Arkhipov *et al.* 2005]. In the all-atom simulation of biomolecules, the most prominent tool is molecular dynamics (MD) simulation [Frenkel and Smit 2001], where Newton's equation of motion is solved to get thermodynamic and dynamic properties of the system. Some of the challenges in the MD simulation of biomolecules are development of faster simulation algorithm for longer simulation, treatment of interaction among the atoms, convergence of simulation trajectories etc. Along with MD, Monte Carlo (MC) and Brownian dynamics (BD) simulations are also used to study macromolecules at different level of resolutions [Hansman *et al.* 1999, Rienzo *et al.* 2001].

For describing the interaction between the atoms of a macromolecule, simple potential functions (also known as force fields) consisting of non-bonded (electrostatics and van der waals interactions) and bonded (bond, angle, torsion) terms are used. Some force fields use hydrogen bond term explicitly, while others do not. One major issue in modeling macromolecules is the treatment of electrostatics. Traditionally the force fields came from different groups use fixed point charge on atomic centers (non-polarizable force fields) [MacKerell *et al.* 1998, Cornell *et al.* 1995]. These charges do not respond to the change in the local environment i.e. remain fixed throughout the simulation. However, it is likely that fixed charge model may describe the local electrostatics poorly for cases such as interaction of a charged residue with water, salt bridge formation. To address this issue, polarizable force fields have been developed by several groups

TH-772_04610604

[Kaminski *et al.*2002, Cieplak *et al.* 2001, Ren *et al.* 2003], where the electrostatic property of atom changes with the change in environment. For instance, in the AMBER ff02 force field, the polarization is described by a combination of fixed charge and induced dipole. The induced dipole moment of an atom depends on the electric field coming from rest of the system and on the atom centered polarizability. Although the polarizable force fields are available in various MD packages, the numbers of applications with polarizable force fields are still limited. Even fewer are the systematic comparison between non-polarizable and polarizable force field simulation. One example is the investigation of water dynamics near the protein BPTI with polarizable and non-polarizable force fields by Berne group [Kim *et al.* 2005]. In the present work, we have compared two AMBER force fields, one non-polarizable (ff99) [Cornell *et al.* 1995] and one polarizable (ff02) [Cieplak *et al.* 2001] by running MD simulation on HIV-protease (HIV-pr), a drug target for HIV infection. HIV-pr, a member of the aspartyl protease family of enzyme, is involved in the late stage of HIV-infection. There are several drugs against HIV, targeting different HIV proteins at different stages of the viral life cycle. Currently, three HIV proteins, namely HIV-RT, HIV-pr and HIV-gp41 are targeted. Of these three targets, HIV-pr acts at the late stage of infection by cleaving the Gag and Gag–Pol polyproteins to yield mature infectious virions for continuation of the viral lifecycle. Although there are several protease inhibitors approved by the FDA, the current major problem is drug resistance caused by the drug induced mutations on HIV-pr. Except for the newest drugs, for which the mutations are not yet known, all other existing drugs are becoming resistant against HIV-pr. A major challenge is to design drugs less susceptible to mutations, requiring the knowledge of detailed molecular mechanism of drug resistance. The structure of HIV-pr is shown in figure 1. The important regions of the structure, such as, active site, flap, fulcrum, and cantilever are shown for the two chains (A and B) of the protein in the figure. The movement of flaps is of particular interest, since the flap movement may restrict the entry of inhibitor to the active site, which may reduce the binding affinity of inhibitors. This may be one of the reasons for drug resistance, since computer simulations found differences in flap movement for wild type and mutant forms of HIV-pr. Flaps have several glycine residues making it highly flexible. Because of the importance of flap dynamics, there have been a large number of

TH-772_04610604

simulations to understand the conformational dynamics of HIV-protease, mostly by MD simulation. Some works attempted to understand the flap opening mechanism [Scott *et al.* 2000, Hornak *et al.* 2006a & 2006b]. Collins *et al.* used activated MD to see the conformation change from close to open conformation [Collins *et al.* 1995] Several works focused on the difference of conformational dynamics between wild type and mutant forms of HIV-pr [Perryman *et al.* 2004, Ode *et al.* 2006, Bandyopadhyay and Meher 2006]. In the present work we have compared the differences in dynamics of HIV-pr between non-polarizable and polarizable simulations. We have performed 10 ns MD simulation of HIV-pr with ff99 and ff02 force fields. The solvent model was taken as TIP3P (a non-polarizable model) and POL3 (a polarizable model) for ff99 and ff02 respectively [Jorgensen *et al.* 1983, Caldwell *et al.* 1995]. We have compared the calculated order parameters for the N-H bond vector with the NMR values. The flap residues are found to have similar order parameter values for the two force fields. The global structural fluctuation shows that the polarizable simulation makes the protein slightly more rigid than the non-polarizable simulation. We have also examined the movement of water molecules around one representative residue in each of polar, charged and hydrophobic group of residues. It has been found that for the charged residue the difference in water movement is most significant. Both the water count and radial distribution function (RDF) between water-residue differs significantly for the Asp25 residue (the catalytic residue of HIV-pr). For the polar and hydrophobic residues the differences are less between the two force fields.

5.2. Computational details:

The MD simulations are done with AMBER programme package [Case *et al.* 2004]. The starting structure for both the simulations is a semi-open structure of 2.7 Å resolution (pdb ID 1HHP) [Spinelli *et al.* 1991]. TIP3P and POL3 are used as the water models for the ff99 and ff02 simulations respectively. The protein is solvated in a water box containing more than 8000 water molecules. The total charge of the whole system was made neutral by addition of chloride ions.

TH-772_04610604

The catalytic Asp-25 and Asp-124 were kept in the deprotonated form in the simulations. The particle mesh ewald (PME) method was used to treat the long range electrostatics as implemented in AMBER [Essman *et al.* 1995]. Berendsen's thermostat and barostat [Berendsen *et al.* 1984] was used to have constant temperature and pressure conditions. The SHAKE algorithm was used to constraint the bonds involving the hydrogen atoms. Order parameters (S^2) were calculated from a plateau region of the N-H internuclear vector autocorrelation function. The system was minimized in two phases, first giving restraints to the protein and crystallographic waters and then removing the restraints. Heating the system up to 300K was also done in two phases, first giving restraints to the protein and crystallographic waters and then removing the restraints. The production run was for 10 ns. The time step kept for ff99 was 2 fs and for ff02 it was 1 fs, as a larger time step made the ff02 trajectory unstable.

5.3. Results and discussion

5.3.1 Root Mean Square Deviation (RMSD)

To check the stability of the trajectories, RMSD of the Cα atoms for the ff99 and ff02 trajectories are calculated and shown in Figure 5.1. It can be seen that the RMSD in the ff99 trajectory is distinctly more than that in the ff02 trajectory. The RMSD for the ff02 trajectory remains below 2 Å while for ff99 it fluctuates mostly between 2 and 2.5 Å. This indicates that the ff99 force field is making the protein more flexible globally in our simulation.

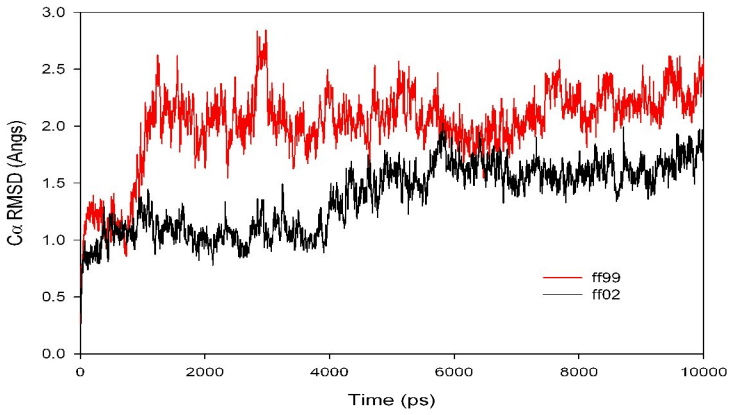
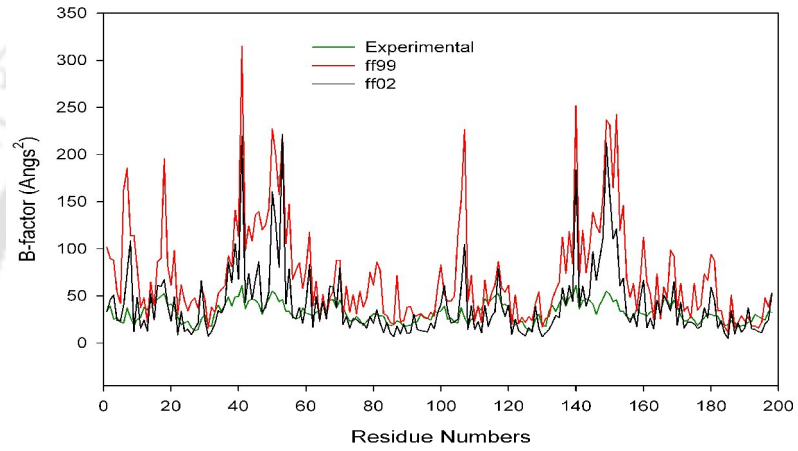


Figure 5.1: *RMSD of the ff99 and f02 trajectories.*

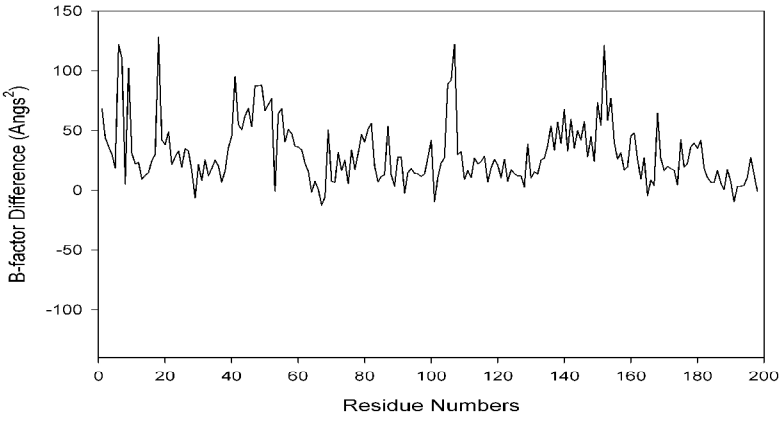
TH-772_04610604

5.3.2. B-factors Analysis:

To check the flexibility in our protein, we have calculated and compared the B-factors. Figure 5.2(a) shows a comparison between X-ray structure B-factor and calculated B-factors for the ff99 and ff02 trajectories. It can be seen that the calculated B-factors are much higher than that X-ray structure B-factors and ff99 has higher values as compared to ff02, which confirms the higher flexibility of protein in non-polarizable simulation and rigidity in the polarizable one. The difference of temperature factors (B-factor) of the protein for polarizable and non polarizable simulation is shown in Figure 5.2(b).



(a)



(b)

Figure 5.2: *(a) B-factors for the X-ray structure, non-polarizable (ff99) and polarizable (ff02) HIV-pr; (b) Difference of B-factors for the HIV-pr (polarizable B-factor – non-polarizable B-factor).*

TH-772_04610604

There exists a broad difference between the two force fields, but specifically certain residues have higher differences like: residues 6-9, 17-19, 40-42, 47-52, 54-56, 80-82 for chain-A and 105-107, 136-140, 144-146, 150-154 for chain-B.

5.3.3. S^2 Order parameters:

Figure 5.3 shows the comparison between N-H order parameter from NMR S^2 (S^2 NMR) and S^2 calculated from ff99 and ff02 force fields (S^2 MD) averaged over the two chains of HIV-pr. It can be seen that the ff99 values deviate more from S^2 NMR. The differences are prominent for the residues Gln2, Leu10, Thr12, Thr26, Leu38, and Thr80. For the flap region (residue 43-58) the calculated order parameters are similar. This is presumably because of the hydrophobic nature of the flap residues. The results suggest that torsion parameters of the residues, especially of glycines are more important than polarization as far as the flexibility of the backbone is concerned.

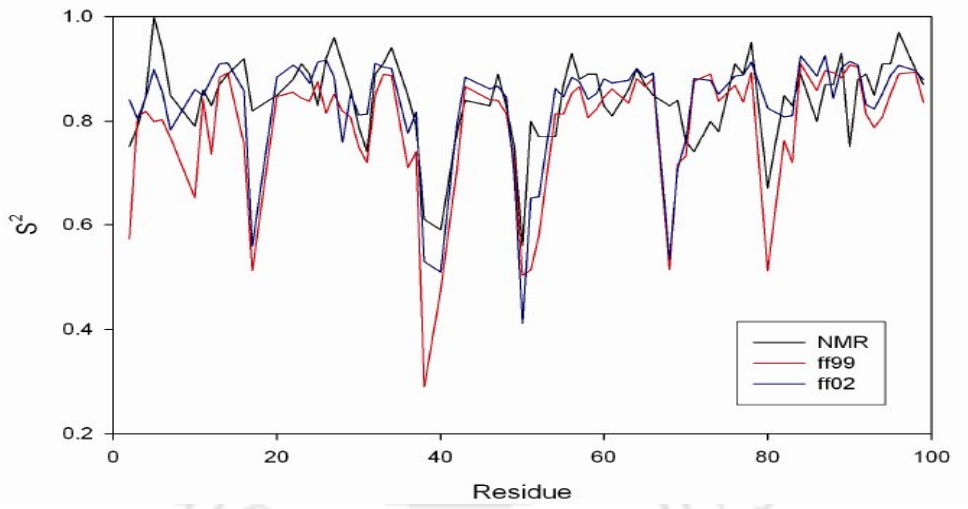


Figure 5.3: *Comparison of NMR order parameters for the N-H bond vector with that calculated from ff99 and ff02 trajectories.*

5.3.4 Flap dynamics:

To understand the degree of flap opening, we have examined the flap-flap and flap-active site distances. Both these distances have been used as simple measures for flap opening in literature. It has been found from our simulation that the flap-flap distance is overlapping between the two simulations (Figure 5.4). However, the flap-active site

TH-772_04610604

distance differs between the two force fields. Figure 5.5 (a) & (b) shows the flap-active site distance (the distance between Cα atom of Asp25 and Cα atom of Ile50 for chain A and the distance between Cα atom of Asp124 and Cα atom of Ile149 for chain B) of HIV-pr for the ff99 and ff02 trajectories. The average values of the distance for chain-A are 18.6 Å (with standard deviation of 2.0 Å) and 13.2 Å (with standard deviation of 0.9 Å) for the ff99 and ff02 trajectories respectively. The same trend was found for chain B also. For chain-B the distances are 18.3 Å (with standard deviation of 2.4 Å) and 15.6 Å (with standard deviation of 2.3 Å) for the ff99 and ff02 trajectories respectively.

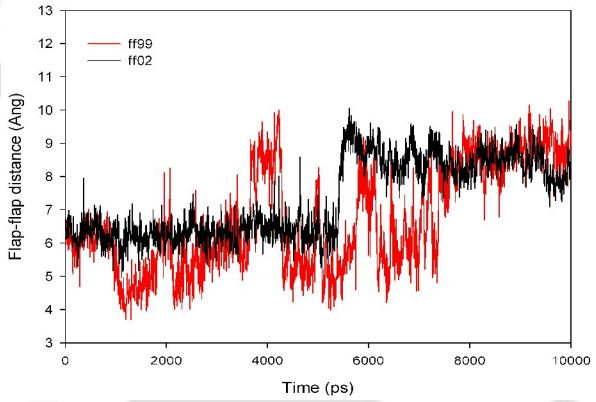


Figure 5.4: *Flap-flap (Ile50Ca-Ile149Ca) distance from the ff99 and ff02 trajectories.*

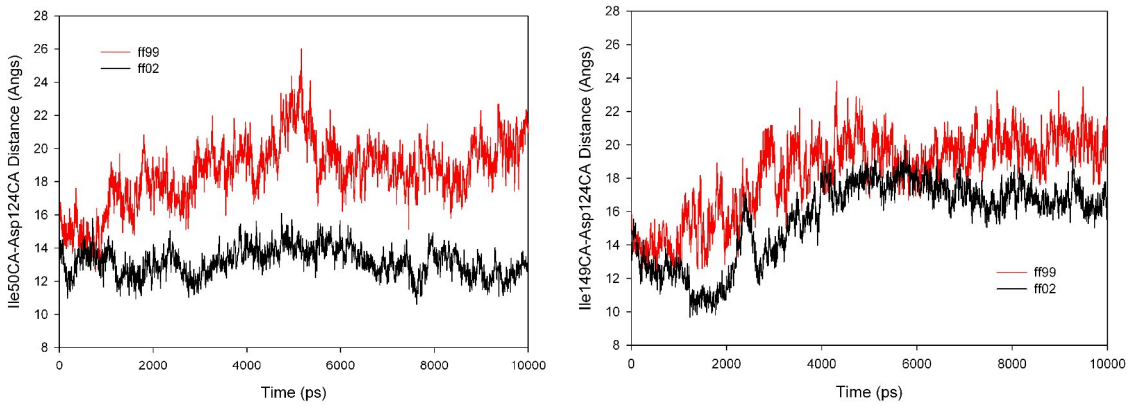


Figure 5.5: *Flap-active site (a) for chain A (Asp25Ca-Ile50Ca) and (b) for chain B (Asp124Ca-Ile149Ca) distance from the ff99 and ff02 trajectories.*

TH-772_04610604

5.3.5. Water movement:

One of the most important effects of polarization is the interaction of water molecules with the protein. To find the behavior of water molecules around the different regions of the protein, we have considered three types of residues: charged, polar and hydrophobic. We have taken one residue in each class. Asp25 as charged, Ser37 as polar and Ile85 as hydrophobic. Figure 5.6 shows that number of water molecules within 8 Å around the concerned residue. It is seen that for the charged residue Asp25 the difference is most significant. In the polarizable simulation, there are mostly 2 to 5 waters around Asp25, while in the non-polarizable simulation there are mostly 8-14 waters around Asp25 with the maximum number of waters is 16. For the Ser37 the water count is similar for the two simulations. For the hydrophobic residue Ile85, except for the last 3 ns of the simulation, the water counts are similar.

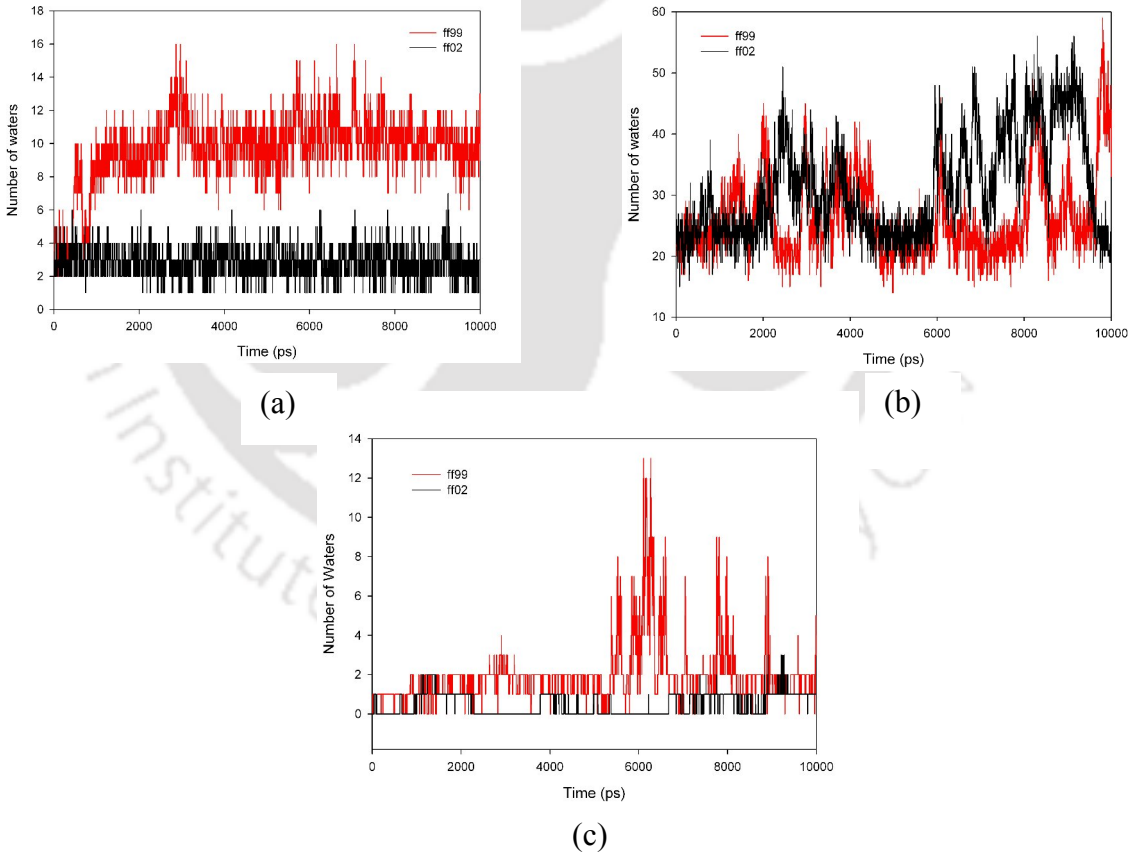


Figure 5.6: *Number of waters within 8 Å from (a) Asp25 (b) Ser37 and (c) Ile85 for the ff99 and ff02 trajectories.*

TH-772_04610604

To understand the microscopic feature of water movement around the residues considered, radial distribution function (RDF) of water residue distance is computed and shown in Figure 5.7. For the case of Asp25, it can be seen that in the first solvation shell the RDFs are similar. However, for the second solvation shell and beyond there is more number of waters (density) in the ff99 force field. Thus the difference in the water count is not coming due to difference in waters in immediate vicinity of Asp25 but from waters away from it. For Ser37, the feature of RDF is similar, although the water density is more for ff02. For the Ile85, water density is more in ff99. It is to be noted that the results discussed for water movement should not be taken as features of residue property only. The water count and movement also depend on the local structure around the concerned residues and convergence of the simulation. However, it is interesting to see that there is significant difference in water movement around the charged catalytic residue Asp25. The difference in water movement is less prominent for the other two types of residues.

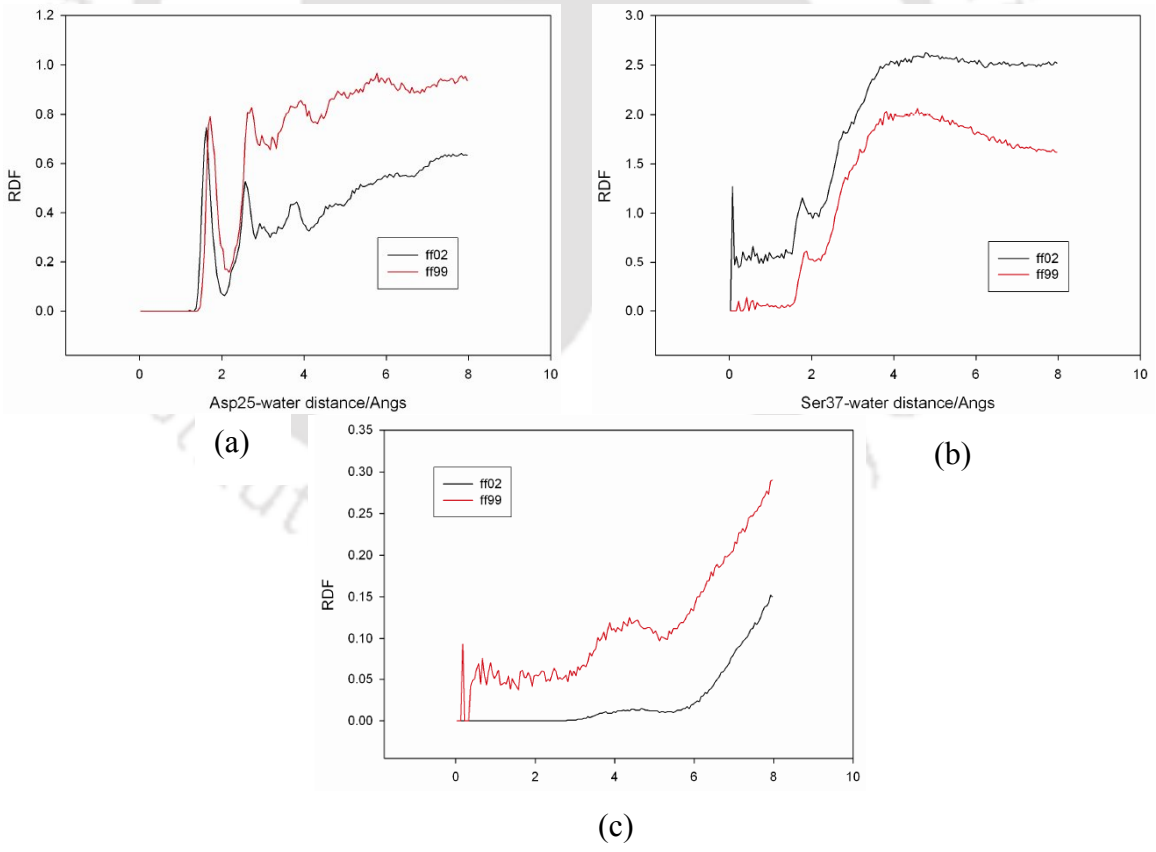


Figure 5.7: *Radial distribution functions for (a) water-Asp25 (b) water-Ser37 and (c) water-Ile85 interaction for the ff99 and ff02 trajectories.*

TH-772_04610604

If we assemble the results from the analysis described above for the two trajectories, some general features emerge. The polarizable simulation makes the protein less flexible as seen from the RMSD values. The order parameters for the N-H bond vector are similar for most of the residues except six residues, where ff99 simulation makes the residues more fluctuating. The flap region's order parameters are similar for the two simulations. Flap-flap distance does not show any major differences between the two force fields. However, the flap-active site distance is several angstroms more in the ff99 simulation. The water-protein interaction is examined for one residue in each class of charged, polar and hydrophobic residues. The maximum effect of polarizable force field is seen for water movement around the catalytic Asp25. In the first solvation shell the water count and RDFs are similar for the two fore fields. However, beyond the first solvation shell, ff99 simulation has about 10 water molecules more within 8 Å. For the polar (Ser37) and hydrophobic (Ile85) the difference between ff99 and ff02 is reduced. It is to be noted that a full investigation of the polarization requires much longer simulation and application on different protein systems. However, the present study shows that there are visible differences in both global and local features of the protein conformational dynamics and water-protein interaction due to inclusion of polarization in the force field.

5.4 Conclusion:

1) The results of the present work show that the polarizable simulation makes the protein more rigid in a global sense.

2) The calculated order parameters of the N-H bond vector are mostly similar except for few residues of the protein.

3) The flap-active site distance, a measure for active site cavity opening, is more in the ff99 force field.

4) The most interesting result of the present work is the differential movement of water molecules around a charged residue (the catalytic residue) Asp25. It has been found that there is more number of waters in the ff99 force field around that residue. Water movement around polar and hydrophobic residues is similar for the two force fields, although some differences exist.

TH-772_04610604

CHAPTER 6

Conformational dynamics of HIV-1 protease: A comparative study with multiple AMBER force fields

TH-772_04610604

Conformational dynamics of HIV-1 protease: A comparative study with multiple AMBER force fields.

6.1. Introduction

Developing better drugs against various HIV proteins is an important goal to combat the HIV virus [Ghosh *et al.* 2008]. HIV-pr, one of the most indispensable enzymes for HIV replication, is an important target for drug design. It acts at the late stage of infection by cleaving the Gag and Gag–Pol polyproteins to yield mature infectious virions for continuation of the viral life-cycle. Several drugs targeting HIV-pr have been developed and approved by the FDA for the treatment of HIV infection [www.UNAIDS.org]. Unfortunately, resistance to these drugs has built up quickly in the form of mutations in the protease. The current goal is to develop drugs less susceptible to mutations. A detailed knowledge of HIV-pr structure, dynamics and its interaction with ligands and substrates is required to design better drugs.

Structurally, HIV-pr is a homodimeric aspartyl protease of which active site is capped by two identical flexible, glycine rich ß-hairpins, or flaps that restrict access to the active site. With the functional Asp residues located at the dimer interface, the protease contains 198 residues with each subunit containing 99 residues. Along with the flaps, movement of fulcrum, flap elbow and cantilever may also control the size of the active site cavity. X-ray diffraction studies [Spinelli *et al.* 1991, Lapatto *et al.*1989, Wlodawer *et al.*1989] showed that in the ligand bound form, the flaps are on top of each other resulting in a closed conformation. On the other hand, in the apo-form of HIV-pr, the flaps move away from each other to form a semi-open conformation [Louis *et al.*1998, Jeyabalan *et al.* 2000]. However, it has been suggested that crystal packing effects may be responsible for the specific semi-open conformations seen in crystal structures. In the absence of crystal packing effects, semi-open conformations may have a wide variety of structures.

The population of mutant strains of HIV-pr increases with the intake of drugs. These mutations in HIV-pr can be classified as two types. One is near the active site and the other is far from the active site. The near active site mutations may change the direct

TH-772_04610604

interaction between the ligand and the protein, which can reduce the binding affinity of the ligand to the mutant. The non-active site mutations are more intriguing. The reduction of binding affinity of the ligands due to the non-active site mutations is likely related to the change in the conformational dynamics of the protein (the indirect effect). Some mutations may have both direct and indirect effects [Bandyopadhyay and Meher 2006]. The flaps of HIV-pr have β-hairpin structure with several glycine residues. Presence of glycine makes the flaps highly flexible. Previous computer simulation studies suggest that the flap dynamics controls the size of the active site cavity and hence affects ligand binding. NMR experiments propose a model for HIV-pr dynamics, where the flaps are described as an ensemble of semi-open structures with rapid movement of the flap tips. It has been suggested that the flap dynamics has two different time scales. The flap tip motion is rapid (in the nano-second range), while the full flap movement is much slower (in the micro-milli second range) [Ishima *et al.* 1999]. Computer simulations of HIV-pr have been highly valuable because of their ability to look the protein at the atomic resolution, which is difficult for experiments. There have been a large number of computer simulation studies to understand the HIV-pr dynamics primarily using molecular dynamics (MD) simulation. In an early work, Collins *et al.* used activated MD simulation to see the flap opening during the time scale of simulation [Collins *et al.* 1995]. Schiffer and coworkers discussed how the flap curling may be a mechanism for substrate entry [Scott *et al.* 2000]. Piana *et al.* investigated the role of compensatory mutation in drug resistance and the role of conformation fluctuation on the enzymatic mechanism of HIV-pr [Piana *et al.* 2002a, Piana *et al.* 2002b]. Perryman *et al.* suggested that the flap dynamics of the double mutant V82F/I84V was significantly different than that of the wild type [Perryman *et al.* 2004]. Our previous work on the MD simulation study of I47V mutation of HIV-pr showed that along with a subtle change in the flap dynamics, change in a specific interaction between the side chain of the residue 47 in chain B and the ligand was likely the main reason for drug resistance [Bandyopadhyay and Meher 2006]. Toth *et al.* investigated flap opening using MD simulation and found that weak network of polar interactions are stabilizing the semi-open conformations [Toth *et al.* 2006]. Ode *et al.* examined the effect on non-active site mutation by MD simulation [Ode *et al.* 2006]. In an important work, Hornak *et al.* used implicit solvent with a low

TH-772_04610604

viscosity to model the flap opening [Hornak *et al.* 2006a]. Since, low viscosity allows the faster folding rates of proteins, transitions from close to open and back was seen for the first time in the simulation. Carlson group showed that the initial solvent distribution around HIV-pr and improper equilibration can influence the flap dynamics observed in simulations [Meagher *et al.* 2005]. However, computer simulation studies give various plausible models for the flap dynamics. Several simulations [Scott *et al.* 2000, Perryman *et al.* 2004] suggested one interesting model where the flaps are in curled conformation (flaps move inward, toward the active site) not corroborated by the NMR results. A recent NMR study on HIV-pr by Ishima *et al* suggests that the flaps weakly interact with each other even in the absence of ligands or substrates [Ishima *et al.* 2007]. They also concluded that their experimental findings can be explained without involving any curled conformation. Also, some simulations observed rapid flap opening (in ns), contradicting the NMR model [Scott *et al.* 2000]. It is likely that some of the observations came from the computer simulations may be influenced by the details of the simulation protocols. The initial solvent distribution is already discussed by Carlson's group. In the present work, the effect of force fields on the conformational dynamics especially the flap dynamics of HIV-pr is considered. Since, it is likely that the number of works and even longer simulation on HIV-pr would be achievable in the coming years, it is imperative to study the effects of simulation details on the flap dynamics of HIV-pr. To the best of our knowledge, there is no systematic study on the effect of force-field on the conformational dynamics of HIV-pr. Further, the results from such a study will be useful for future investigation of proteins with flexible binding sites. Significant differences in MD trajectories may come from several factors, such as the energy function used in the simulation (force field), the underlying integrator for time propagation, initial solvent distribution, initial velocity distribution, convergence of the trajectories to name a few. In the present work, we have focused on the force field effect by using the same program package (AMBER) with three different AMBER force fields namely ff99, ff99SB, and ff03 [Cornell *et al.* 1995, Hornak *et al.* 2006c, Duan *et al.* 2003]. The TIP3P [Jorgensen *et al.* 1983] water model was used with all the force fields. All other simulation details were same for all three trajectories. 10 ns MD simulation for each of the force fields was performed initially. Afterwards simulations with ff99SB and ff03 force fields were

TH-772_04610604

continued up to 30 ns. Another trajectory of 30 ns was run for ff99SB and ff03 starting from a different initial velocity distribution. The ff99SB force field differs from the ff99 force field by having better backbone dihedral parameters. In the ff03 force field, the charges are determined by performing quantum chemical calculation of peptides in condensed phase environment and the main chain torsion parameters are also different from ff99 and ff99SB.

While there are several structural and dynamical quantities can be calculated from MD simulation, it is not easy to validate those without any experimental results. In the present work, we have focused primarily on the N-H S^2 order parameters calculated from the simulations and compared it to the available NMR order parameters. Various structural parameters were also examined to understand the force field dependence of the conformational dynamics. One major source of discrepancy between NMR S^2 and calculated S^2 values is the extent of conformational sampling in the MD simulation. The full convergence of conformation sampling of large biomolecular systems with explicit solvent is difficult to get [Lyman *et al.* 2006]. Thus no attempt is made in the current work to achieve full convergence. Rather the focus of this work is to observe the differences in result from MD simulation with different force fields with simulation length of tens of nanosecond, the routine length of currently published MD simulations. Two separate trajectories were run in order to assess the quality of the results. The results of the simulation indicate that the N-H S^2 order parameters calculated from the ff99SB and ff03 are reasonably close to the NMR S^2 (S^2 NMR) values for most of the residues.

The ff99 values deviate maximum from the S^2 NMR values. Between ff99SB and ff03, although there are similarities in general, significant differences remain for several residues. These are the residues in the loop region, many of these being glycine. A similar finding on a different protein, GB3, is reported [Trbovic *et al.* 2008]. A major consequence of the different fluctuation of these loop residues is the larger active site cavity observed in the ff03 simulation. Examination of hydrogen bonding (H-bonding) of the loop residues show that the H-bonding pattern differs for some of the residues. It appears that multiple factors, such as H-bonding description, different torsion parameters

TH-772_04610604

in the two force fields, and the interplay between different terms in the force field are contributing to the differential mobility of the loop residues. This work shows that the complex dynamics of HIV-pr can be quite sensitive to the force field difference. This partly explains the discrepancy seen in previous simulations of HIV-pr.

One important consequence of the sensitivity of the binding site flexibility to the simulation parameters is how the conformational dynamics of wild type and different mutant forms of HIV-pr depend on the simulation details. Another consequence of our result is in the ensemble based docking approach against flexible receptors. In this approach, potential inhibitors are docked to an ensemble of protein structure, very often obtained from MD simulation [Perryman *et al.* 2006]. However, if the average geometry and flexibility of the binding site differs between simulations with different parameters, the binding of ligands may be influenced by that. More generally, many of the conclusions regarding biological function (such as binding, enzymatic action) are drawn from MD simulations by observing the differences in dynamics. However, the current work shows that careful examination with different simulation parameters is required before making such kind of conclusion.

6.2. Computational details:

All three MD simulations were started using a crystal structure of HIV-pr (pdb ID: 1HHP) [Spinelli *et al.* 1991] of resolution 2.7 Å with semi-open conformation. The Leap module of the AMBER 8 program package [Case *et al.* 2004] was used to prepare the system for simulation. The TIP3P [Jorgensen *et al.* 1983] water model was used in all the simulations. The system was immersed in a water box of size 83.5 x 62.5 x 68.9 Å3 containing more than 8000 water molecules. The net positive charge on the system was neutralized through the addition of chloride ions.

The charged-states of both the catalytic Asp-25 and Asp-124 were kept deprotonated in the simulations. The electrostatic interactions were calculated with the particle mesh ewald (PME) method [Essman *et al.*1995]. Constant temperature and pressure conditions in the simulation were achieved by coupling the system to a Berendsen's thermostat and

TH-772_04610604

barostat [Berendsen *et al.*1984]. Bonds involving the hydrogen atoms were constrained to their equilibrium position with the SHAKE algorithm. Order parameters (S^2) were calculated from a plateau region of the N-H internuclear vector autocorrelation function.

The system was minimized in two phases. In the first phase, the system was minimized giving restraints (30 kcal/mol/Å2) to protein and crystallographic waters for 500 steps with subsequent second phase minimization of the whole system. Then the system was heated to 300K over 25 ps with a 1 fs time step. The protein atoms were restrained with force constant of 30 kcal/mol/Å2 at the NVT ensemble. After that the force constant was reduced by 10 kcal/mol/ Å2 in each step to reach the unrestrained structure in three steps of 10 ps each. The system was then switched over to the NPT ensemble and equilibrated without any restraints for 180 ps. The system was equilibrated in total of 235 ps. The time step for MD simulation for the production run was 2 fs. All three trajectories were initially run for 10 ns. The ff99SB and ff03 simulations were later extended to 30ns. One more trajectory of 30ns was run with ff99SB and ff03 force fields with a different initial velocity distribution.

6.3. Results and Discussions:

6.3.1: RMSD of the Cα atoms:

Figure 6.1 shows the RMSD values of the Cα atoms calculated from the three simulations, which ensure stable trajectories. The ff99 simulation data is up to 10 ns, while the ff99SB and ff03 simulations are up to 30 ns. The ff99 force field shows the maximum RMSD among the three force fields with values fluctuating mostly between 2.0 and 2.5 Å. RMSD values fluctuate generally below 2 Å for ff99SB and ff03 force fields. There is a sharp increase of RMSD in the ff03 trajectory in the last 3 ns, suggesting a major change in conformation.

TH-772_04610604

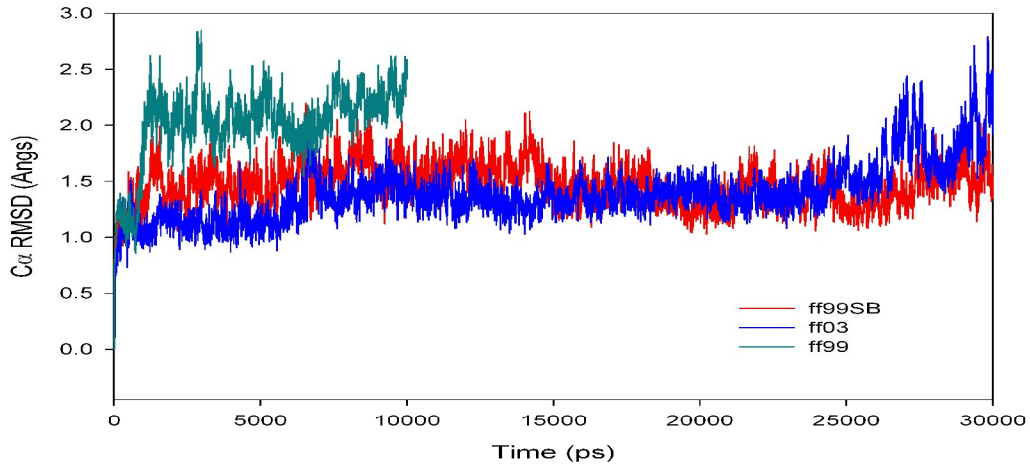


Figure 6.1: *RMSD values for the Cα atoms for the three trajectories with three different force fields.*

6.3.2: Generalized N-H order parameter (S^2):

The calculated S^2 values (S^2 MD) for the N-H bond vector averaged over the two chains of HIV-pr from the three force fields are compared with the S^2 NMR. In general, most of the S^2 MD values are lower than the S^2 NMR values as found by the previous workers [Hornak *et al.* 2006]. Figure 6.2(a) shows the comparison of S^2 MD calculated with three force fields with S^2 NMR. It is clearly seen that the ff99 values differ most from the S^2 NMR. The ff99 calculated S^2 values are lowest and hence makes HIV-pr most flexible among the three force fields considered. It shows six regions of HIV-pr highly flexible (low values of S^2), not corroborated by the NMR results. The ff99SB values show similar trend as the S^2 NMR with maximum differences comes in the residues 38-40 (flap elbow) and in the flap region. The maximum deviation between the ff99 and ff99SB results is observed in the residues Leu10, Gly16-17 (fulcrum region), Thr31, flap region (49-52), Gly68 and Thr80. The experimental values are almost always higher than the ff99SB values except for the residues Thr31, Ile50, Ile54, Ala71 and Leu90.

TH-772_04610604

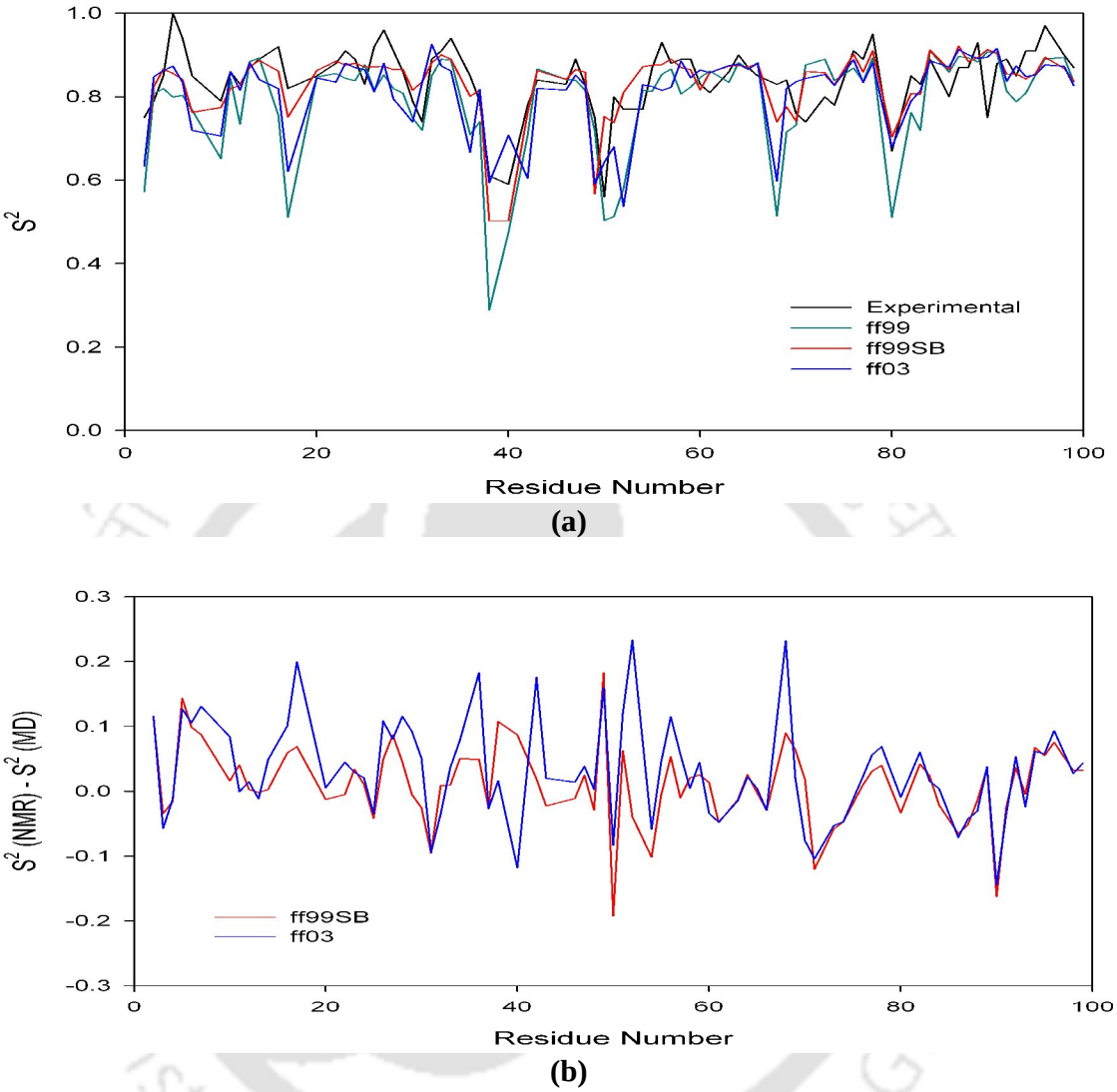


Figure 6.2: *(a) Comparison of experimental NMR S^2 N-H order parameter values with the calculated values from ff99, ff99SB and ff03 force field simulations. (b) The difference between ff99SB and ff03 calculated order parameters from the NMR S^2 values.*

The difference between the S^2 calculated using ff99SB and ff03 is depicted in Figure 6.2(b), which shows the difference between S^2 MD and S^2 NMR for these two force fields. The ff99SB values are in general higher than the ff03 values with the exception of residues 40 and 70. This is similar to the findings of Trbovic and coworkers [Trbovic *et al.* 2008]. The major difference between the ff99SB and ff03 values come for the residues Gly17, Met36, Gly40, Trp42, Ile50, Gly52, and Gly68. All of these residues are part of

TH-772_04610604

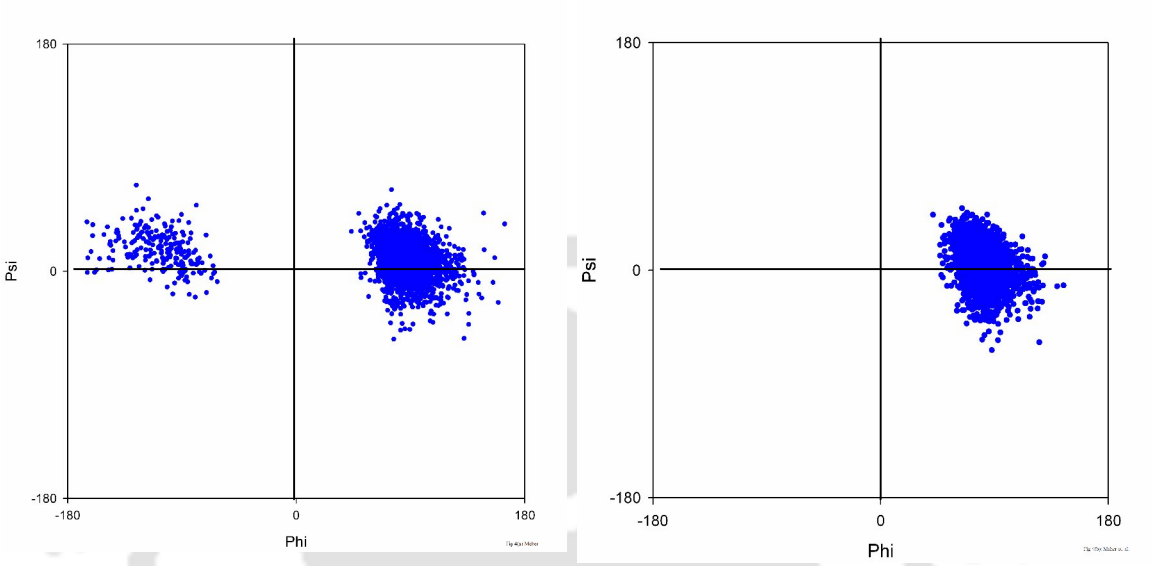
different loop regions. Gly17 is in the bend of the beta-hairpin of the fulcrum. Met36, Gly40 and Trp42 belong to the flap elbow loop. Residues Ile50 and Gly52 are part of the flap and Gly68 belongs to the part of the mobile loop 68-71 in the cantilever region [see Figure 1.7(a)]. Among the glycine, Gly68 is a special case, since this residue is part of the mobile loop 68-71, with histidine in position 71, which is likely to be affected by the proton exchange, not captured in our simulation [Piana *et al.* 2002a]. This in turn could change the dynamics of Gly68. One of the most important results from the present work is the higher mobility of many of the loop residues in the ff03 force field simulation. Except for the active site region (containing the catalytic triad and the active site wall), all the loops show difference in S^2 calculated by ff99SB and ff03 force fields. It is to be noted that the S^2 NMR values were derived from an autocatalysis-resistant mutant of HIV pr, where the following mutations were carried out, Gln7-Lys, Leu33-Ile, and Leu63-Ile. This could one of the reasons for the difference between S^2 NMR and S^2 MD. Chemical exchange is found to be important for the terminal residues (4-6, 97-99) and the flap residues in the NMR experiments [Ishima *et al.* 1999]. Although the time scale of chemical exchange is not accounted for in our MD simulation, the differences that are observed between ff99SB and ff03 are found for other loops, where chemical exchange is not significant. Thus the difference between ff99SB and ff03 cannot be due to the omission of chemical exchange in the MD simulation**.** To cross-check the results, S^2 MD values are calculated from the second trajectory of 30ns for both ff03 and ff99SB. In the second trajectory the S^2 MD of the flap residues did not converge. For other loop residues, ff03 derived S^2 values are mostly lower as seen in the first trajectory. Therefore, although the absolute values differ, the relative differences between ff99SB and ff03 exist in the second trajectory.

6.3.3. Ramachandran Plot:

The differences in the Ramachandran angles of these residues were also checked and found there are considerable differences in the Φ and Ψ angles. As an example the Ramachandran plot of residue Gly17 is shown in Figure 6.3. It can be seen that the ff03 simulation samples two regions of Φ, Ψ space (Φ -180 to -45 and again -45 to +45), while ff99SB samples only one region of the Φ, Ψ space. Examination of Ramachandran

TH-772_04610604

plots of other six residues also shows that the sampling of secondary structure by ff99SB and ff03 are different. Since, the ff99 force field is describing the protein dynamics poorly, for the rest of this manuscript; we will be reporting the difference between the ff99SB and ff03 force fields only.



(a) **(b)**

Figure 6.3: *Ramachandran Plot for Glycine17 for (a) ff03 (b) ff99SB force fields simulation.*

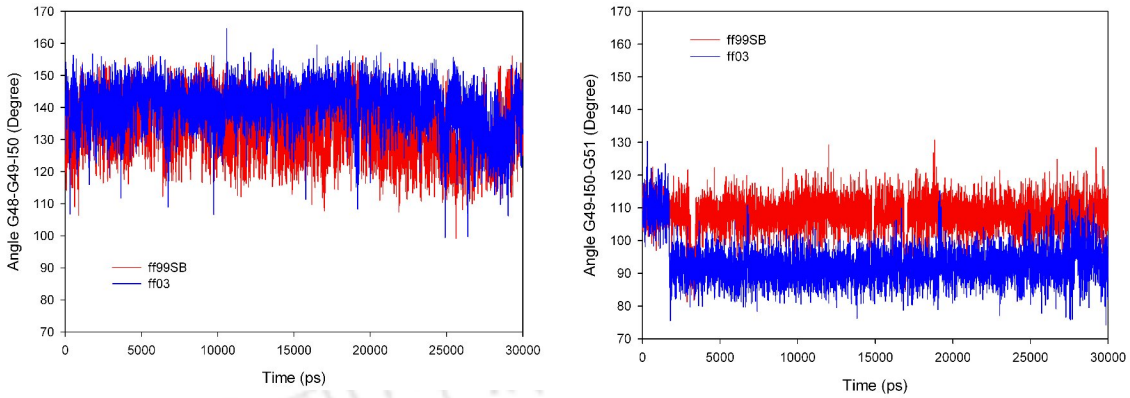
6.3.4 Flap movement

Understanding the flap movement is an important issue pertaining to the conformational dynamics of HIV-pr. In the present work, flap movement was examined in detail for the ff99SB and ff03 trajectories.

6.3.4.1 Flap curling and flap-flap distance:

The examination of the relevant angles involving the residues in the flap region shows that the flap curling is not significant in either of the ff03 and or the ff99SB trajectories [Figure 6.4 (a) & (b)]. The flap-flap distance (i.e. the distance between Ile50@CA-Ile149@CA) is found to be mostly overlapping between the two trajectories, indicating that the flap-flap distance is not sensitive to the differences between ff99SB and ff03 force fields [Figure 6.5].

TH-772_04610604



(a) **(b)**

Figure 6.4: *TriCa angle in the flap region (a) Angle Gly48-Gly49-Ile50 (b) Angle Gly49-Ile50-Gly51.*

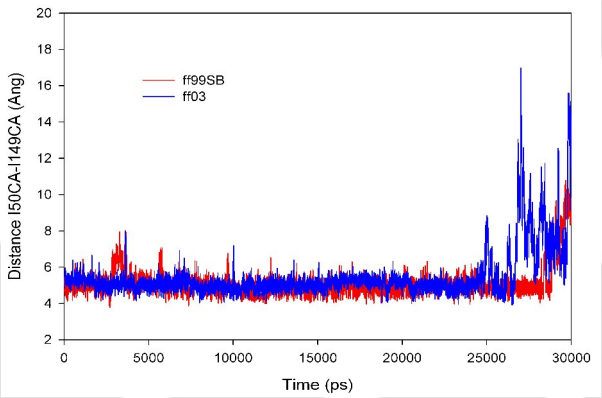


Figure 6.5: *Flap-flap distance from the ff99SB and ff03 force field simulations.*

6.3.4.2. Flap-active site distance:

The flap-active site distance (the distance between Ile50@CA-Asp25@CA and the corresponding distance in chain B) shows significant difference between these two force fields. The flap-active site distance for both chains A and B of HIV-pr are shown in Figure 6.6 for the two force fields. For the chain A (Figure 6.6a), it can be seen that, the ff03 distance is fluctuating between 15 to 20 Å, while the ff99SB distance is fluctuating between 12- 16 Å except for the last 3 ns of the simulation. The average value for this distance is 17.5 Å with standard deviation of 1.1 Å for the ff03 force field. For the ff99SB force field the average distance is 14.0 Å with standard deviation of 1.3 Å.

TH-772_04610604

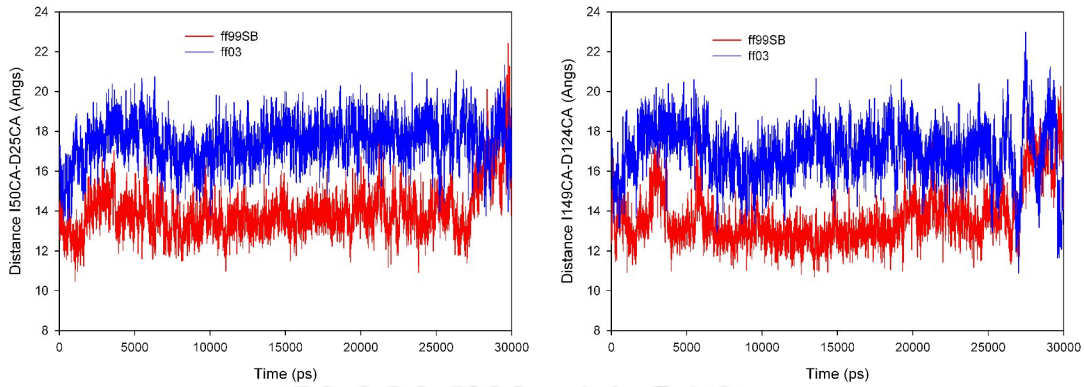


Figure 6.6: *Flap-active site distance for (a) Chain A and (b) Chain B from the ff99SB and ff03 force field simulations.*

The chain B distance (Figure 6.6b) also show similar features. The average flap-active site distance in chain B is found to be 13.7 Å (1.4 Å as std. dev.) for ff99SB and 16.9 Å (1.4 Å std. dev.) for ff03. The flap-active site distance is a measure of flap opening. Since, this distance on average is 3 Å longer for both chains in ff03 simulation, the opening of the cavity is more according to the ff03 simulation. Figure 6.7 shows superimposition of two representative structures from ff03 and ff99SB force fields showing the larger cavity size in the ff03 simulation.

A closer inspection of the movement of different loops reveals that the differential motion of flap elbow (containing residue 17) and the fulcrum (containing residue 40) are contributing most to the difference seen in the active site-flap distance. Flap elbow and fulcrum movement is pushing the two flaps upwards, thus the flap-flap distance does not change but the flap-active site distance changes significantly. Movies of MD simulations are given in the soft copy material. We have investigated possible reasons for the differential mobility in the loop regions. The ff03 and ff99SB differ both in their charges as well in their dihedral parameters. Both these parameters can influence the relative fluctuation. Also, the interaction with TIP3P water molecules may also contribute to their differences. The difference in H-bonding patterns involving the residues of interest and interaction of water molecules with these seven loop residues have been investigated systematically.

TH-772_04610604

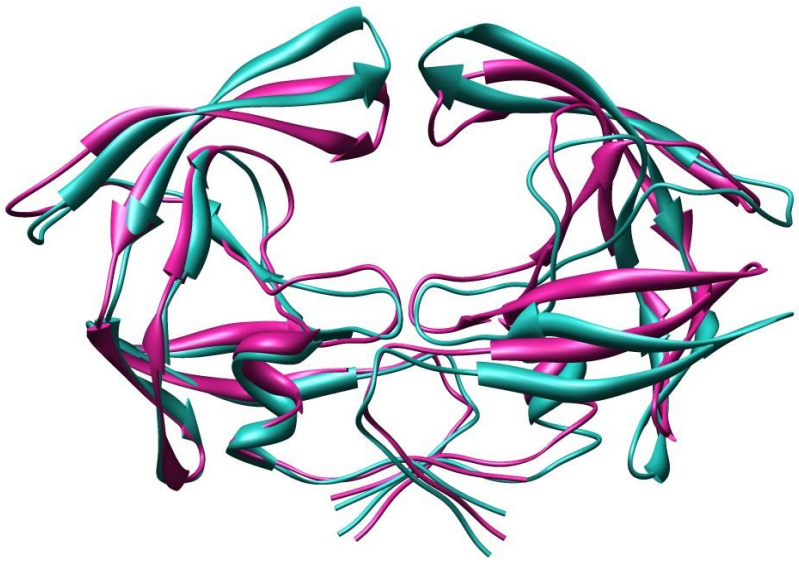


Figure 6.7: *Two representative structures obtained from ff03 (green) and ff99SB (magenta) simulations.*

6.3.4.3 Analysis of interactions

The H-bonding with water molecules for these seven loop residues did not show any significant difference. The intra-molecular H-bonding of these residues shows that for Met36 CO and Leu38 NH H-bond, there is 11% occupancy according to the ff03 force field, while according to the ff99SB force field the occupancy is 19%. For Met36NH and Glu34CO H-bond, the occupancy is 1% according to ff99SB and 13% according to ff03. Other major difference came for the H-bond involving Gly52CO and Gly49NH. Here the occupancy is 53% in ff99SB and 38% in ff03 force fields. While for Gly52NH and Gly49CO H-bond, the occupancies are 6 % and 21 % for ff99SB and ff03 respectively. Gly68NH and Ile66CO H-bond has occupancies 1% and 12% for the ff99SB and ff03 trajectories respectively. Table 6.1 shows the all observed H-bonds involving these seven residues. From this table it can be said that although H-bonding difference exists between ff03 and ff99SB for some residues, intra-molecular H-bonding alone is not sufficient to explain the differential fluctuation of the loop residues. It is likely that torsion parameters are also contributing to the differences, since most differences are in the highly flexible glycines. Although it is not easy to say the exact reason for the differences between these force fields, it seems that it is a combination of several factors, different description of H-

TH-772_04610604

bonding, different charge sets and different torsion parameters are contributing to the difference.

Table 6.1: Hydrogen bonded distances (in Angstrom) involving the loop residues showing different fluctuation between ff03 and ff99SB simulation and adjacent residues with their percentage of occupancy in ff99SB and ff03 trajectories.

	Ff99SB		ff03	
Residue@Atom	Occupancy	Distance	Occupancy	Distance
Met36@O...Gly38@N	19.0	2.9	11.8	2.9
Met135@O...Gly137@N	12.4	2.9	37.3	2.9
Gly40@N...Gly38@O	5.5	2.9	0.05	2.9
Gly139@N...Gly137@O	2.0	2.9	0.1	2.8
Ile50@O...Gly150@N	-	-	2.2	2.9
Gly52@O...Gly49@N	53.5	2.9	38.4	2.9
Gly52@N...Gly49@O	6.4	2.9	21.2	2.9
Gly151@O...Gly148@N	38.1	2.9	28.4	2.9
Gly151@N...Gly148@O	11.5	2.9	6.2	2.9
Gly68@N...Ile66@O	1.1	2.8	12.4	2.8
Gly68@O...Ile66@N	0.01	2.8	0.13	2.8
Gly167@N...Ile165@O	1.7	2.8	15.3	2.8
His69@N...Ile66@O	58.5	2.9	45.1	2.9
His168@N...Ile165@O	54.4	2.9	45.4	2.9

One caveat to our results, as mentioned before, is the insufficient sampling in the MD simulations. However, it is encouraging that qualitatively most of the S^2 MD values follow the trend seen in the S^2 NMR values. Also, the difference between ff99SB and ff03 follows the same trend as seen by the other authors [Trbovic *et al.* 2008]. Our results should not be taken to decide which one is a better force field between ff03 and ff99SB. However, our results do show the visible differences in the mobility of the loop residues with different force fields.

TH-772_04610604

This work shows how the subtle differences in the force fields can influence the protein dynamics. Also, the observation that active site size variation is directly linked to the fluctuation of the flap elbow and fulcrum can be used to design allosteric ligands. This is also found by McCammon group, with their simulation of HIV-pr with ff99 force field [Perryman *et al.* 2004]. The sensitivity of loop dynamics with force field is likely to be important not only for HIV-pr but also for other flexible proteins such as protein kinases. The outcome of this work would be important for proteins with flexible binding sites, which may show different fluctuation depending on the different force field parameters. One important case where this is a major problem is in the ensemble based docking approach for proteins with flexible binding site, where ligands are docked to the snapshots of MD simulation [Perryman *et al.* 2006]. If the average geometry and fluctuation of the binding site differ significantly on simulation parameters then this may as well influence the ligand binding and the relative ranking of ligands. Another major issue in HIV-pr dynamics is whether the different mutants have similar conformational dynamics as the wild type. It is suggested that the mutations away from the active site modulate the conformation dynamics in such a way that it reduces inhibitor binding affinity. Already there are several published works on the difference of dynamics between wild type and various mutants of HIV-pr using MD simulation. It would be interesting to check the effect of force field on the mutant flexibility as compared to the wild type for some of these.

6.4. Conclusion:

The specific objective of this work is to understand the effect of force field difference on the flexibility of HIV-pr and its significance in *in silico* drug design against HIV-pr.

1) However, in a more general sense, this work also shed light on the difficulty in modeling dynamics of proteins with flexible binding site and *in silico* drug design against flexible receptors.

TH-772_04610604

2) Simulation results show that ff99 force field's description of protein dynamics is poor. The ff99SB and ff03 force fields' overall description is similar but there are differences in dynamics of various loop residues. Many of these residues are glycines present in the loops.

3) The calculated S^2 values for these residues (except one) are significantly lower in the ff03 compared to ff99SB (and deviate more from the NMR S^2 values) indicating greater flexibility. Two different 30ns MD trajectories were used to check the ff03 and ff99SB S^2 values. These fluctuations lead to a larger active site cavity in the ff03 force field as a consequence of longer active site-flap distance (~3 Å for both chains).

4) The interactions determining the loop fluctuation indicate that it is likely that a combination of several factors like H-bonding, electrostatics, torsion parameters are contributing to this difference. The difficulty of modeling loops in MD simulation can be important for proteins with flexible binding sites.

TH-772_04610604

CHAPTER 7

Concluding Remarks

TH-772_04610604

Concluding Remarks:

7.1. Summary:

The main theme of the current thesis is the investigation of conformational dynamics of HIV-pr by MD simulations. Both biological and technical aspects (i.e. the dependence of the dynamics on the details of simulation) of the dynamics have been investigated in this thesis work.

For the biological aspect, first the effect of a mutation in the flap region on the dynamics of HIV-pr and its effect on ligand binding were investigated. The mutation I47V, which is relatively rare but showed up clinically for an experimental inhibitor JE-2147, was investigated. Four MD simulations of liganded and unliganded HIV-pr in their WT and mutant forms have been performed. The results of this study illustrated that the loss of one $–CH_2$ group for I47V mutant increases the side chain mobility of mutant Val47 in chain B as compared to the Ile47 in the WT protease. This may have a role in conferring the increased flexibility of flaps indirectly and decreased affinity for ligand binding originating resistance in HIV-pr. This is owing to the loss of optimized packing of the inhibitor to the residue 47 in chain B of the mutant when compared with WT enzyme. Positioning a larger group at the P2' position of JE-2147 might provide recovery of the loss of –CH2 group while going mutant from Ile to Val and may increase the binding properties between the ligand-receptor to minimize the resistance. There have been a large number of similar works on the effect of mutation by other researchers. It has been found that mutations in the flap region generally affect the flap mobility leading to either flap opening or closing mechanism and subsequently the binding affinity for inhibitors.

For instance, some mutations (such as E35D, G48V/L90M, I50V, F53L, I54V, V82F/I84V) [Meiselbach *et al.* 2007, Maschera *et al.* 1996, Liu *et al.* 2005, Ishima *et al.* 1999, Perryman *et al.* 2004] enhance the flap dynamics and direct the opening of the flaps while others (K45I, M46I, M46I/G51D) reduce the flap dynamics and direct closing of flaps [Mahalingam *et al.* 2001, Piana *et al.* 2002b, Lauria *et al.* 2007]. The mutations

TH-772_04610604

affecting the flap dynamics in HIV-pr can be categorized into two different types as: (1) flap dynamics enhancer mutations and (2) flap dynamics repressor mutations. By comparing the effect of I47V mutant with the previously published works on different mutant proteases, a general conclusion can be made that, mutant I47V facilitates higher mobility of the flaps (as the side chain becomes shorter relative to WT) and comes in the category of flap dynamics enhancer mutations. In specific, I47V guides the increased movement of Val47 side chain enhancing the flap dynamics indirectly. Also, directly it affects the VDW for the residue Ile47 and steric interactions for the residue Gly48.

Further, we studied the effects of change in physiological parameter like pressure on the conformation and dynamics of HIV-pr. From this study, we conclude that under high pressure condition, there is a general decrease in the protein's structural degrees of freedom. The results also demonstrate the increase in compactness of structure under high pressure that affects the secondary structure of the protein with reasonable changes in α-helix and β-sheets to turns and bends. The difference in Solvent Accessible Surface and the change in secondary structures with increase in pressure show that pressure acts as a unique conformer selector consequently reducing the structural variability. We could also see that the active site region of the protein was shrunk to a greater extent with pressure. This may affect considerably the binding affinity of drugs to the active site region. Further investigation is required to throw light on this.

The second part of the thesis focused on the effects of different simulation setups on the conformation and dynamics of HIV-pr. In particular, the effect of different force fields (both polarizable and non-polarizable) on the flap dynamics was investigated. In one study, the comparisons of one polarizable and other non-polarizable AMBER force field (ff02 and ff99 respectively) were carried out for HIV-pr dynamics. The outcome of this simulation showed that the polarizable simulation makes the protein more rigid compared to non-polarizable simulation which confers from its higher S^2 order parameters of the N-H bond vector and lower B-factors. The flap-active site distance is more in the ff99 force field ensuring a larger active site cavity opening. The most interesting result of this study is the differential movement of water molecules around the catalytic charged residue

TH-772_04610604

Asp25. It is noticed that there is more number of waters in the ff99 force field around the charged residue. Water movement around polar and hydrophobic residues is almost similar for the two force fields.

In another study, the effect of different AMBER non-polarizable force fields (ff99, ff99SB and ff03) on the conformational dynamics of HIV-pr was studied. Out of these three force fields, ff99SB and ff03 are two of the most widely used AMBER force fields. Simulation results show that ff99 force field's description of protein dynamics is poor. The ff99SB and ff03 force fields' overall description is similar but there are differences in dynamics of various loop residues, mostly glycines in the loop regions. The calculated S^2 values for these residues are considerably lower in the ff03 compared to ff99SB indicating greater flexibility. These fluctuations lead to a larger active site cavity in the ff03 force field as a result of longer active site-flap distance. The interactions determining the loop fluctuation indicate that it is likely that a combination of several factors like H-bonding, electrostatics, torsion parameters are contributing to this difference. Comparative study of the HIV-pr dynamics by multiple force fields sheds light on the difficulty in modeling dynamics of proteins with flexible binding site and *in silico* drug design against flexible receptors. So, the complex dynamics of HIV-pr can be sensitive enough to the force field difference. Hence a careful examination with different simulation parameters is required to conclude regarding the biological functions drawn from MD simulation studies.

7.2. Scope of future works:

The importance of mutation induced conformational dynamics (especially the flap dynamics) of HIV-pr has given wide opportunities to explore the molecular basis of drug resistance with greater insight. Furthermore, the effects of different simulation protocols and force fields on protein conformation and dynamics has given an idea to select the correct simulation setups and force fields before carrying out MD simulation of biomolecules.

TH-772_04610604

These are the potential directions that can be explored on the basis of present work.

- Information regarding the positioning of a larger group at the P2' position of JE-2147 will be helpful in drug designing. One can proceed for the structural development of JE-2147 *in vitro* and *in silico* as well.
- Polarization influences rigidity in proteins. For proteins with highly flexible or highly rigid domains, the inclusion of polarization in the system setup is expected to have significant outcomes.
- Also, polarization influences the differential motion of water molecules around a charged residue, which can be helpful in analyzing the proteins containing huge number of charged residues in its important regions like the catalytic sites.
- Difference in force fields illustrates the floppy behavior of loop residues with respect to the currently available standard force fields, which can be effective in development of force fields with greater stability for loop residues.

.

TH-772_04610604

Bibliography

TH-772_04610604

Bibliography:

- Akasaka, K. Tezuka, T. and Yamada, H. (**1997**) Pressure-induced changes in the folded structure of lysozyme. *J. Mol. Biol.* 271: 671-678.
- Alder, B. J. and Wainwright T. E. (**1957**) Phase Transition for a hard sphere system. *J. Chem. Phys.* 27: 1208-1209.
- Alder, B. J. and Wainwright T. E. (**1959**) Studies in Molecular Dynamics. I. General Method. *J. Chem. Phys.* 31: 459-466.
- Arkhipov, A. Freddolino, P. L. Imada, K. Namba, K. and Schulten, K. (**2006**) Coarse-Grained Molecular Dynamics simulations of a rotating bacterial flagellum. *Biophys. J.* 91: 4589-4597.
- Armstrong, R. N. (**1998**) Mechanistic imperatives for the evolution of glutathione transferases. *Curr Opin Chem Biol*. 2: 618-623.
- Baca, M. and Kent, S. B. H. (**1993**) Catalytic contribution of flap-substrate hydrogen bonds in "HIV-1 protease" explored by chemical synthesis. *Proc. Natl. Acad. Sci.* 90: 11638-11642.
- Bagossi, P. Cheng, Y.-S. E. Oroszlan, S. and Tözsér, J. (**1996**) Activity of linked HIV-1 proteinase dimers containing mutations in the active site region. *Protein Eng.* 9: 997-1003.
- Baldwin, E. Bhat, T. N. Gulnik, S. Liu, B. Topol, I. A. Kiso, Y. Mimoto, T. Mitsuya, H. and Erickson, J. W. (**1995**) Structure of HIV-1 protease with KNI-272, a tight-binding transition-state analog containing allophenylnorstatine. *Structure* 3: 581-590.
- Baldwin, R. L. (**1986**) Temperature dependence of the hydrophobic interaction in protein folding. *Proc. Natl. Acad. Sci.* 83: 8069-8072.
- Bandyopadhyay, P. and Meher, B. R. (**2006**) Drug resistance of HIV-1 protease against JE-2147: I47V mutation investigated by molecular dynamics simulation. *Chem. Biol. Drug Des.* 67: 155-161.
- Bayly, C. L. Cieplak, P. Cornell, W. D. and Kollman, P. A. (**1993**) A well-behaved electrostatic potential based method using charge restraints for determining atom-centered charges: The RESP model. *J. Phys. Chem.* 97: 10269-.

TH-772_04610604

- Behler, J. Martonak, R. Donadio, D. and Parrinello, M. (**2008**) Metadynamics simulations of the high-pressure phases of silicon employing a high-dimensional neural network potential. *Phys. Rev. Lett.* 100: 185501-185504.
- Ben-Naim, A. (**1978**) Hydrophobic Interactions. *Physics and Chemistry of Liquids.* 7: 375-386.
- Berendsen, H. J. C. Postma, J. P. M. Van Gusteren, W. F. Dinola, A. and Haak, J. R. (**1984**) Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* 81: 3684-3690.
- Berendsen, H. J. C and Hayward, S. (**2000**) Collective protein dynamics in relation to function. *Curr. Opin. Struct. Biol.* 10: 165-169.
- Block, S. M. (**1998**) Leading the procession: new insights into kinesin motors. *J. Cell Biol*. 140: 1281-1284.
- Boden, D. and Markowitz, M. (**1998**) Resistance to immunodeficiency virus type 1 protease inhibitors. *Antimicrob. Agents Chemother.* 42: 2775–2783.
- Borgstahl, K. Ng, G. E. Ren, K. Pradervand, Z. Burke, C. Srajer, P. M. Teng, V. T. Schildkamp, Y. McRee, W. Mo_at, D. E. and Getzo, E. K. (**1997**) Structure of a protein photocycle intermediate by millisecond time-resolved crystallography. *Science* 275: 1471-1475.
- Bray, D. D. Slattery, N. and Russel, C. S. (**1984**) Guanidinium-carboxylate interaction: methylguanidinium formate. *Int. J. Peptide Protein. Res.* 24: 414- 418.
- Burkert, U. and Allinger, N. L. (**1982**) Molecular Mechanics. ACS Monograph 177. *American Chemical Society*: Washington D. C.
- Byrne, M. P. Manuel, R. L. Lowe, L. G. and Stites, W. E. (**1995**) Energetic contribution of side chain hydrogen bonding to the stability of staphylococcalnuclease. *Biochemistry* 34: 13949-13960.
- Caldwell, J. W. and Kollman, P. A. (**1995**) *J. Phys. Chem.* 99: 6208-6219.
- Canalia, M. and Malliavin, T. E. (**2004**) Molecular Dynamics simulation of HPr under hydrostatic pressure. *Biopolymers* 74: 377-388.
- Case, D. A. *et al.* (**2002** and **2004**) AMBER 7 and 8, University of California, San Francisco, USA.

TH-772_04610604

- Cavanagh, J. and Venters, R. A. (**2001)** Protein dynamic studies move to a new time slot. *Nature* 8: 912-914.
- Chanda, B. Asamoah, O. K. Blunck, R. Roux, B. and Bezanilla, F. (**2005**) Gating charge displacement in voltage-gated ion channels involves limited transmembrane movement. *Nature* 436: 852-856.
- Changeux, J. P. (**1993**) Allosteric proteins: from regulatory enzymes to receptors. *Bioessays* 15: 625-634.
- Chatfield, D. C. and Brooks, B. R. (**1995**) HIV-1 Protease cleavage mechanism elucidated with Molecular Dynamics simulation. *J. Am. Chem. Soc.* 117: 5561-5572.
- Chen, X. and Tropsha, A. (**1995**) Relative binding free energies of peptide inhibitors of HIV-1 Protease: The influence of the active site protonation state. *J. Med. Chem.* 38: 42-48.
- Cieplak, P. J. C. and Kollman, P. A. (**2001**) Molecular mechanical models for organic and biological systems going beyond the atom centered two body additive approximations: aqueous solution free energies of methanol and N-methyl acetamide, nucleic acid base, and amide hydrogen bonding and chloroform/water partition coefficients of the nucleic acid bases. *J. Comp. Chem.* 22: 1048-1057.
- Collins, J. R. Burt, S. K. and Erickson, J. W. (**1995**) Flap opening in HIV-1 protease simulated by activated molecular dynamics. *Nat. Struct. Biol.* 2: 334-338.
- Cornell, W. D. Cieplak, P. J. C. Bayly, I. Gould, I. R. Merz, K. M. Ferguson, D. M. Spellmeyer, D. C. Fox, T. Caldwell, J. W. and Kollman, P. A. (**1995**) A second generation force field for the simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* 117: 5179-5197.
- Creighton, T. E. (**1993**) *Proteins structures and Molecular properties,* 2nd edition, W. H. Freeman publication.
- Daniel, R. M. Dunn, R.V. Finney, J. L. and Smith, J. C. (**2003)** The role of dynamics in enzyme activity. *Annu. Rev. Biophys. Biomol. Struct*. 32: 69-92.
- Daura, X. Jaun, B. Seebach, D. Van Gunsteren, W. F. and Mark, A. E. (**1998**) Reversible peptide folding in solution by molecular dynamics simulation. *J. Mol. Biol.* 5: 925-932.

TH-772_04610604

- Davies, D. R. Parlan, E. A. and Sheriff, S. (**1990**) Antibody-antigen complexes. *Annu. Rev. Biochem.* 59: 439-473.
- Davies, D. R. (**1990**) The structure and function of the aspartic proteinases. *Annu Rev Biophys. Biophys. Chem*. 19: 189-215.
- Davies, D. R. and Chacko, S. (**1993**) Antibody structure. *Acc. Chem. Res.* 26: 42-427.
- DeLano, W. L. The *PyMOL* Molecular Graphics System (**2002**) DeLano Scientific, San Carlos, CA, USA. http://www.pymol.org
- Dewar, M. J. S. Zoebisch, E. G. Healy, E. F. and Stewart, J. P. (**1985**) Development and use of quantum mechanical molecular models. 76. AM1: a new general purpose quantum mechanical molecular model. *J. Am. Chem. Soc.* 107: 3902-3909.
- Duan, Y. and Kollman, P. A. (**1998**) Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution. *Science* 5389: 740-744.
- Duan, Y. Wu, C. Chowdhury, S. Lee, M. C. Xiong, G. Zhang, W. Yang, R. Cieplak, P. Luo, R. Lee, T. Caldwell, J. Wang, J. and Kollman, P. A. (**2003**) A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations, *J. Comp. Chem.* 24: 1999-2012.
- Eisenmesser, E. Z. Akke, M. Bosco, D. A. and Kern, D. (**2002**) Enzyme dynamics during catalysis. *Science* 295: 1520-1523.
- Eisenmesser, E. Z. Millet, O. Labeikovsky, W. Korzhnev, D. M. Wolf-Watz, M. Bosco, D. A. Skalicky, J. J. Kay, L. E. and Kern, D. (**2005)** Intrinsic dynamics of an enzyme underlies catalysis. *Nature* 438: 117-121.
- Eliel, E. L. Allinger, N. L. Angyal, S. J. and Morrison, G. A. (**1965**) Conformational Analysis. Wiley-Interscience: New York.
- Erickson, J. W. Neidhart, D. J. VanDrie, J. Kempf, D. J. Wang, X. C. Norbeck, D. W. Plattner, J. J. Rittenhouse, J. W. Turon, M. Wideburg, N. Kohlbrenner, W. E. Simmer, R. Helfrich, R. Paul, D. A. and Knigge, M. (**1990**) Design, activity, and 2.8 Å crystal structure of a C2 symmetric inhibitor complexed to HIV-1 protease. *Science* 249: 527-533.
- Eriksson, A. E. Baase, W. A. Zhang, X. J. Heinz, D. W. Blaber, M. Baldwin, E. P. and Matthews, B. W. (**1992**) Response of a protein structure to cavity-creating mutations and its relation to the hydrophobic effect. *Science* 255: 178-183.

TH-772_04610604

- Essmann, U. Perera, L. Berkowitz, M. L. Darden, T. Lee, H. and Pedersen, L. G. (**1995**) A smooth particle mesh Ewald method" *J. Chem. Phys.* 19: 8577-8593.
- Ewald, P. (**1921**) Die Berechnung optischer und elektrostatischer Gitterpotentiale" *Ann. Phys.* 64: 253-287.
- Fitzgerald, P. M. D. and Springer, J. P. (**1991**) Structure and function of retroviral proteases. *Annu. Rev. Biophys. Biophys.Chem.* 20: 299-320.
- Freedberg, D. I. Wang, Y. X. Stahl, S. J. Kaufman, J. D. Wingfield, P. T. Kiso, Y. and Torchia, D. A. (**1998**) Flexibility and Function in HIV Protease: Dynamics of the HIV-1 Protease Bound to the Asymmetric Inhibitor Kynostatin 272 (KNI-272). *J. Am. Chem. Soc.* 120: 7916.
- Freedberg, D. I. Ishima, R. Jacob, J. Wang, Y.-X. Kustanovich, I. Louis, J. M. and Torchia, D. A. (**2002**) Rapid structural fluctuations of the free HIV protease flaps in solution: Relationship to crystal structures and comparison with predictions of dynamics calculations. *Protein Sci.* 11: 221-232.
- Frenkel, D. and Smit, B. (**2001**) Understanding Molecular Simulation (Academic Press; 2nd edition)
- Geeves, M. A. and Holmes, K. C. (**1999**) Structural mechanism of muscle contraction. *Annu. Rv. Biochem.* 68: 687-728.
- Gelfand, V. and Bershadsky, A. D. (**1991**) Microtubule dynamics: mechanism, regulation, and function. *Annu. Rev. Cell. Biol.* 07: 93-116.
- Geller, M. Miller, M. Swansom, S. M. and Maizel, J. (**1997**) Analysis of the structure of HIV-1 protease complexed with a hexapeptide inhibitor. Part II: Molecular dynamic studies of the active site region. *Proteins* 27: 195-203.
- Ghosh, A. K. Chapsal, B. D. Weber, I. T. and Mitsuya, H. (**2008**) Design of HIV-1 protease inhibitors targeting protein backbone: An effective strategy for combating drug resistance. *Acc. Chem. Res.* 41: 78-86.
- Goldman, Y. E. (**1998**) Wag the tail: structural dynamics of actomysin. *Cell* 93: 1- 4.
- Gruner, S. M. (**2004**) Soft materials and biomaterials under pressure. *High-Pressure Crystallography*, edited by A. Katrusiak and P. McMillan, Kluwer Academic Publisher, Netherlands.

TH-772_04610604

- Grzesiek, S. Bax, A. Nicholson, L. K. Yamazaki, T. Wingfield, P. T. Stahl, S. J. Eyermann, C. J. Torchia, D. A. Hodge, C. N. Lam, P. Y. S. Jadhav, P. K. and Chang, C. H. (**1994**) NMR evidence for the displacement of a conserved interior water molecule in HIV Protease by a non-peptide cyclic Urea-based inhibitor. *J. Am. Chem. Soc.* 116: 1581-1582.
- Guex, N. and Peitsch, M. C. (**1997**) SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. *Electrophoresis* 18:2714-2723.
- Gulnik, S. and Erickson, J. W. (**2000**) HIV protease: enzyme functions and drug resistance, *Vitamins and Hormones* 58: 213-256.
- Guo, Z. Brooks, C. L. III and Boczko, E. M. (**1997**) Exploring the folding free energy surface of a three-helix bundle protein. *Proc. Natl. Acad. Sci.* 19: 10161-10166.
- Hamelberg, D and McCammon, J. A. (**2005**) Fast peptidyl cis-trans isomerization within the flexible Gly-rich flaps of HIV-1 Protease. *J. Am. Chem. Soc.* 127: 13778-13779.
- Hamm, H. E. (**1998**) The many faces of G-protein signaling. *J. Biol. Chem.* 273: 669-672.
- Hansmann, U. H. E. and Okamoto, Y. (**1999**) New Monte Carlo algorithms for protein folding. *Curr. Opin. Struct. Biol.* 9: 177-183.
- Harrison, R. W. and Weber, I. T. (**1994**) Molecular dynamics simulations of HIV-1 protease with peptide substrate**.** *Protein Eng.* 7: 1353-1363.
- Harte, W. E. and Beveridge, D. L. (**1993**) Prediction of the protonation state of the active site aspartyl residues in HIV-1 protease-inhibitor complexes via molecular dynamics simulation. *J. Am. Chem. Soc.* 115: 3883-3886.
- Hayward, S. and Berendsen, H. J. C. (**1998**) Systematic analysis of domain motions in proteins from conformational change: new results on citrate synthase and T4 lysozyme. *Proteins* 30: 144-154.
- Helmerich, E. J. and Hofmann, K. P. (**1996**) Structure and function of proteins in G-protein coupled signal transfer. *Biochim. Biophys. Acta* 1286: 285-322.
- Heremans, K. and Smeller, L. (**1998**) Protein structure and dynamics at high pressure. *Biochim. Biophys. Acta* 1386: 353-370.

TH-772_04610604

- Hermans, Jr. J. and Rialdi, G. (**1965**) Heat of ionization and denaturation of Sperm-Whale myoglobin determined with a microcalorimeter. *Biochemistry* 4: 1277-1281.
- Hertogs, K. Bloor, S. Kemp, S. D. Van den Eynde, C. Alcorn, T. M. Pauwels, R. Van Houtte, M. Staszewski, S. Miller, V. and Larder, B. A. (**2000**) Phenotypic and genotypic analysis of clinical HIV-1 isolates reveals extensive protease inhibitor cross-resistance: a survey of over 6000 samples. *AIDS* 14: 1203-1210.
- Hillson, J. Onuchic, N. N. and García, A. E. (**1999**) Pressure-induced protein folding/unfolding kinetics. *Proc. Nat. Acad. Sci.* 96: 14848-14853.
- Hockney, R. W. (**1970**) The potential calculation and some applications. *Methods in Computational Physics* 9: 136-211.
- Hodge, C. N. Lam, P. Y. S. Eyermann, C. J. Jadhav, P. K. Ru, Y. Fernandez, C. H. De Lucca, G. V. Chang, C. H. Kaltenbach, III R. F. Holler, E. R. Woerner, F. Daneker, W. F. Emmet, G. Calabrese, J. C. and Aldrich, P. E. (**1998**) Calculated and experimental low-energy conformations of cyclic Urea HIV Protease inhibitors. *J. Am. Chem. Soc.*120: 4570-4581.
- Holloway, M. K. Wai, J. M. Halgren, T. A. Fitzgerald, P. M. D. Vacca, J. P. Dorsey, B. D. Levin, R. B. Thompson, W. J. and Chen, L. J. (**1995**) A priori prediction of activity for HIV-1 protease inhibitors employing energy minimization in the active site. *J. Med. Chem.* 38: 305-.
- Hong, L. Hartsuck, J. A. Foundling, S. Ermolieff, J. and Tang, J. (**1998**) Active-site mobility in human immunodeficiency virus, type 1, protease as demonstrated by crystal structure of A28S mutant. *Protein Sci.* 7: 300-305.
- Honig, B. and Yang, A. S. (**1995**) Free energy balance in protein folding. *Adv. Protein Chem.* 46: 27-58.
- Hornak, V. Okur, A. Rizzo, R. C. and Simmerling, C. (**2006a**) HIV-1 Protease flaps spontaneously close to the correct structure in simulations following manual placement of an inhibitor into the open state. *J. Am. Chem. Soc.* 128: 2812 -2813.
- Hornak, V. Okur, A. Robert, C. R. and Simmerling, C. (**2006b**) HIV-1 protease flaps spontaneously open and reclose in molecular dynamics simulations. *Proc. Nat. Acad. Sci.* 103: 915-920.

TH-772_04610604

- Hornak, V. Abel, R. Okur, A. Strockbine, B. Roitberg, A. and Simmerling, C. (**2006c**) Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins* 65: 712-725.
- Hummer, G. Garde, S. García, A. E. Paulaitis, M. E. and Pratt, L. R. (**1998**) The pressure dependence of hydrophobic interactions is consistent with the observed pressure denaturation of proteins. *Proc. Natl. Acad. Sci.* 95: 1552-1555.
- Humphrey, W. Dalke, A. and Schulten, K. (**1996**) VMD: Visual Molecular Dynamics. *J. Mol. Graph.* 14: 33-38.
- Hyland, L. J. Tomaszek, T. A. Roberts, G. D. Carr, S. A. Maagard, V. W. Bryan, H. L. Fakhoury, S. A. Moore, M. L. Minnich, M. D. Culp, J. S. DesJarlais, R. L. and Meek, T. D. (**1991a**) Human immunodeficiency virus-1 protease. 1. Initial velocity studies and kinetic characterization of reaction intermediates by oxygen-18 isotope exchange. *Biochemistry* 30: 8441-8453.
- Hyland, L. J. Tomaszek, T. A. and Meek, T. D. (**1991b**) Human immunodeficiency virus-1 protease. 2. Use of pH rate studies and solvent kinetic isotope effects to elucidate details of chemical mechanism. *Biochemistry* 30: 8454-8463.
- Ippolito, J. A. Alexander, R. S. and Christianson, D. W. (**1990**) Hydrogen bond stereochemistry in protein structure and function. *J. Mol. Biol.* 215: 457- 471.
- Ishima, R. Wingfield, P. T. Stahl, S. J. Kaufman, J. D. and Torchia, D. A. (**1998**) Using amide ^{1}H and ^{15}N transverse relaxation to detect millisecond time-scale motions in perdeuterated proteins: Application to HIV-1 Protease. *J. Am. Chem. Soc.* 120: 10534-10542.
- Ishima, R. Freedberg, D. I. Wang, Y. Louis, J. M. and Torchia, D. A. (**1999**) Flap opening and dimmer-interface flexibility in the free and inhibitor-bound HIV protease, and their implications for function. *Structure* 7: 1047-1055.
- Ishima, R. and Louis, J. M. (**2007**) A diverse view of protein dynamics from NMR studies of HIV-1 protease flaps. *Proteins* 70: 1408-1415.
- Jeyabalan, M. P. Nalivaika, E. and Schiffer, C. A. (**2000**) How does a symmetric dimmer recognize an asymmetric substrate? A substrate complex of HIV-1 protease. *J. Mol. Biol.* 301: 1207-1220.

TH-772_04610604

- Jeyabalan, M. Nalivaika, E. A. and Schiffer, C. A. (**2002**) Substrate shape determines specificity of recognition for HIV-1 protease: Analysis of crystal structures of six substrate complexes. *Structure* 10: 369-381.
- Jorgensen, W. J. Chandrasekhar, J. Madura, R. I. and Klein, M. (**1983**) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79: 926-935.
- Kabsch, W. and Sander, C. (**1983**) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 22: 2577-637.
- Kaminski, G. A. Stern, H. A. Berne, B. J. Friesner, R. A. Cao, Y. X. Murphy, R. B. Zhou, R. and Halgren, T. A. (**2002**) Development of a polarizable force field for proteins via *ab initio* quantum chemistry: First generation model and gas phase tests. *J. Comput. Chem.* 23: 1515-1531.
- Karplus, M. and McCammon, J. A. (**2002**) Molecular dynamics simulations of biomolecules. *Nat. Struct. Biol.* 9: 646-652.
- Katoh, E. Louis, J. M. Yamazaki, T. Gronenborn, A. M. Torchia, D. A. and Ishima, R. (**2003**) A solution NMR study of the binding kinetics and the internal dynamics of an HIV-1 protease–substrate complex. *Protein Sci.* 12: 1376-1385.
- Katoh, E. Yamazaki, T. Kiso, Y. Wingfield, P. T. Stahl, S. J. Kaufman, J. D. and Torchia, D. A. (**1999**) Determination of the rate of monomer interchange in a ligand-bound homodimeric protein from NOESY cross peaks: Application to the HIV Protease/KNI-529 complex. *J. Am. Chem. Soc.* 121: 2607-2608.
- Kauzmann, W. (**1959**) Some factors in the interpretation of protein denaturation. *Adv. Protein Chem.* 2: 1-63.
- Keinnan, S. and Avnir, D. (**2000**) Quantitative symmetry in structure-activity correlations: The near C_2 symmetry of inhibitor/HIV Protease complexes. *J. Am. Chem. Soc.* 122: 4378-4384.
- Kellis, J. T. J. Nyberg, K. and Fersht, A. R. (**1989**) Energetics of complementary side-chain packing in a protein hydrophobic core. *Biochemistry* 28: 4914- 4922.
- Kellis, J. T. J. Nyberg, K. Sali, D. and Fersht, A. R. (**1988**) Contribution of hydrophobic interactions to protein stability. *Nature* 333: 784-786.
- Kharakoz, D. P. (**2000**) Protein compressibility, dynamics, and pressure. *Biophys. J.* 79: 511-525.

TH-772_04610604

- Kim, B. Young, T. Harder, E. Friesner, R. A. and Berne, B. J. (**2005**) *J. Phys. Chem. B* 109: 16529
- Kitahara, R. Yokoyama, S. and Akasaka, K. (**2005**) NMR snapshots of a fluctuating protein structure: Ubiquitin at 30 bar–3 Kbar. *J. Mol. Biol.* 347: 277-285.
- Knipe, David M. *et al.* Fields Virology. Fourth Edition Lippincot Williamsand Wilkins. June (**2001**). 971-972.
- Kohl, N. E. Emini, E. A. Schleif, W. A. Davis, L. J. Heimbach, J. C. Dixon, R. A. Scolnick, E. M. and Sigal, I. S. (**1988**) Active human immunodeficiency virus protease is required for viral infectivity. *Proc. Natl. Acad. Sci.* 85: 4686-4690.
- Kovalesky, D. Dubyna, V. Mark, A. E. and Kornelyuk, A. (**2005**) A molecular dynamics study of the structural stability of HIV-1 protease under physiological conditions: The role of Na+ ions in stabilizing the active site. *Proteins* 58: 450-458.
- Kozal, M. J. Shah, N. Shen, N. Yang, R. Fucini, R. Merigan, T. C. Richman, D. D. Morris, D. Hubbell, E. Chee, M. and Gingeras, T. R. (**1996**) Extensive polymorphisms observed in HIV-1 clade B protease gene using high-density oligonucleotide arrays. *Nat. Med.* 2: 753-759.
- Kungl, A. J. Visser, N. V. Van Hoek, A. Visser, A. J. W. G. Billich, A. Schilk, A. Gstach, H. and Auer, M. (**1998)** Time-Resolved Fluorescence Anisotropy of HIV-1 protease inhibitor complexes correlates with inhibitory activity. *Biochemistry* 37: 2778-2786.
- Laskowski, R. A. MacArthur, M. W. Moss, D. S. and Thornton, J. M. (**1993**) *PROCHECK*: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* 26: 283-291.
- Lapatto, R. Blundell, T. H. Overington, A. J. Wilderspin, A. Wood, S. Merson, J. R. Whittle, P. J. Danley, D. E. and Geoghegan, K. F. (**1989**) X-ray analysis of HIV-1 proteinase at 2.7 Å resolution confirms structural homology among retroviral enzymes. *Nature* 342: 299-302.
- Lauria, A. Ippolito, M. and Almerico, A. M. (**2007**) Molecular Dynamics studies on HIV-1 protease: a comparision of the flap motions between wild type protease and the M46I/G51D double mutant. *J. Mol. Model*. 13: 1151-1156.

TH-772_04610604

- Leach, A. R. (**1996**) Molecular Modelling: Principles and Selected Applications. Edimburg Gate, Harlow, Pearson Education Ltd.
- Lee, H. Darden, T. A. and Pedersen, L. G. (**1996**) An *ab initio* Quantum Mechanical model for the catalytic mechanism of HIV-1 Protease. *J. Am. Chem. Soc.* 118: 3946-3950.
- Lee, T. Le, V.-D. Lim, D. Lin, Y.-C. Morris, G. M. Wong, A. L. Olson, A. J. Elder, J. H. and Wong, C.-H. (**1999**) Development of a new type of protease inhibitors, efficacious against FIV and HIV variants. *J. Am. Chem. Soc.* 121: 1145-1155.
- Lennard-Jones, J. E. (**1931**) Cohesion. *Proceedings of the Physical Society* 43: 461-482.
- Levinthal, C. (**1968**) Are there pathways for protein folding? *Journal de Chimie Physique* 65: 44-45.
- Levitt, M. (**1982**) Protein conformation, dynamics and folding by computer simulation. *Annu. Rev. Biophys. Bioeng.* 11: 251-271.
- Liu, F. Kovalevsky, A. Y. Louis, J. M. Boross, P. I. Wang, Y. F. Harrison, R. W. and Weber, I. T. (**2005**) Mechanism of drug resistance revealed by the crystal structure of the unliganded HIV-1 protease with F53L mutation. *J. Mol. Biol.* 358: 1191-1199.
- Liu, H. Müller-Plathe, F. and Van Gusteren, W. F. (**1996**) A combined Quantum/Classical Molecular Dynamics study of the catalytic mechanism of HIV Protease. *J. Mol. Biol.* 261: 454-469.
- Loeb, D. D. Swanstrom, R. Everitt, L. Manchester, M. Stamper, S. E. and Hutchinson, C. A. (**1989**) Complete mutagenesis of the HIV-1 protease. *Nature* 340: 397-400.
- Louis, J. M. Dyda, F. Nashed, N. T. Kimmel, A. R. and Davies, D. R. (**1998**) Hydrophilic peptides derived from the transframe region of Gag-Pol inhibit the HIV-1 protease. *Biochemistry* 37: 2105-2110.
- Luo, X. Kato, R. and Collins, J. R. (**1998**) Dynamic flexibility of protein-inhibitor complexes: A study of the HIV-1 Protease/KNI-272 complex. *J. Am. Chem. Soc.*120: 12410-12418.
- Lyman, E. Zuckerman, D. M. (**2006**) Ensemble-based convergence analysis of biomolecular trajectories. *Biophys. J.* 91: 164-72.

TH-772_04610604

- MacKerell Jr., A. D. Bashford, D. Bellott, M. Dunbrack Jr., R. L. Evanseck, J. D. Field, M. J. Fischer, S. Gao, J. Guo, H. Ha, S. Joseph-McCarthy, D. Kuchnir, L. Kuczera, K. Lau, F. T. K. Mattos, C. Michnick, S. Ngo, T. Nguyen, D. T. Prodhom, B. Reiher, I. Roux, W. E Schlenkrich, B. Smith, M. Stote, J. C. Straub, R. Watanabe, J. Wiorkiewicz-Kuczera, M. Yin, J. D. and Karplus, M. (**1998**) All-atom empirical potential for molecular modeling and dynamics studies of potential for molecular modeling and dynamics studies of proteins. *J. Phys.Chem. B* 102: 3586-3616.
- Mahalingam, B. Louis, J. M. Hung, J. Harrison, R. W. and Weber, I. T. (**2001**). Structural implications of drug-resistant mutants of HIV-1 protease: High-resolution crystal structures of the mutant protease/substrate analogue complexes. *Proteins* 43: 455-464.
- Maschera, B. Darby, G. Palu, G. Wright, L. L. Tisdale, M. Myers, R. Blair, E. D. and Fufine, E. S. (**1996**). Human immunodeficiency virus: Mutations in the viral protease that confer resistance to saquinavir increase the dissociation rate constant of the protease-saquinavir complex. *J. Biol. Chem.* 271: 33231-33235.
- Matsumura, M. Becktel, W. J. and Matthews, B. W. (**1988**) Hydrophobic stabilization in T4 lysozyme determined directly by multiple substitutions of Ile 3. *Nature* 334: 406-410.
- Matthew, J. B. (**1985**) Electrostatic effects in proteins. *Ann. Rev. Biophys. Bioeng.* 14: 387- 417.
- McCammon, J. A. Gelin, B. R. and Karplus, M. (**1977**). Dynamics of folded proteins. *Nature* 267: 585-590.
- McCarthy, A. N. and Grigera, J. R. (**2006a**) Effect of pressure on the conformation of proteins. A molecular dynamics simulation of lysozyme. *J. Mol. Gr. Mod*. 24: 254-261.
- McCarthy, A. N. and Grigera, J. R. (**2006b**) Pressure denaturation of apomyoglobin: A molecular dynamics simulation study. *Biochim. Biophys. Acta* 1764: 506-515.
- McQuade, T. J. Tomasselli, A. Liu, G. L. Karacostas, V. Moss, B. Sawyer, T. K. Heinrikson, R. L. and Tarpley, W. G. (**1990**) A synthetic HIV-1 protease inhibitor with antiviral activity arrests HIV-like particle maturation. *Science* 247: 454-456.

TH-772_04610604

- Meagher, K. L. and Carlson, H. A. (**2005**) Solvation influences flap collapse in HIV-protease. *Proteins* 58: 119-125.
- Meiselbach, H. Horn, A. H. C. Harrer, T. and Sticht, H. (**2007**) Insights into amprenavir resistance in E35D HIV-1 protease mutation from molecular dynamics and binding free-energy calculations. *J. Mol. Mod.* 13: 297-304.
- Miller, M. Rao, J. M. Mohana, J. K. Leis, J. and Wlodawer, A. (**1989a**) Crystal structure of a retroviral protease proves relationship to aspartic protease family. *Nature* 337: 576-579.
- Miller, M. Schneider, J. Sathyanarayana, B. K. Toth, M. V. Marshall, G. R. Clawson, L. Selk, L. M. Kent, S. B. H. and Wlodawer, A. (**1989b**) Structure of complex of synthetic HIV-1 protease with a substrate-based inhibitor at 2.3 A resolution. *Science* 246: 1149-1152.
- Monod, J. Wyman, J. and Changeux, J. P. (**1965**) On the nature of allosteric transitions: a plausible model. *J. Mol. Biol.* 12: 88-118.
- Mu, Y. and Stock, G. (**2002**). Conformational dynamics of Trialanine in water: A Molecular Dynamics study. *J. Phys. Chem. B* 20: 5294-5301.
- Myers, J. K. and Pace, C. N. (**1996**) Hydrogen bonding stabilizes globular proteins. *Biophys. J.* 71: 2033-2039.
- Navia, M. A. Fitzgerald, P. M. D. McKeever, B. M. Leu, C.-T. Heimbach, J. C. Herber, W. K. Sigal, I. S. Darke, P. L. and Springer, J. P. (**1989**) Three-dimensional structure of aspartyl protease from human immunodeficiency virus HIV-1. *Nature* 337: 615-620.
- Nelson, D. L. Cox, M. M. (**2000**) Lehninger Principles of Biochemistry. Third Edition, Macmillan Worth Publishers.
- Nicholson, L. K. Yamazaki, T. Torchia, D. A. Grzesiek, S. Bax, A. Stahlm, S. J. Kaufman, J. D. Wingfield, P. T. Lam, P. Y. Jadhav, P. K. and et al. (**1995**) Flexibility and function in HIV-1 protease. *Nat. Struct. Biol.* 2: 274-280.
- Ode, H. Neva, S. Hata, M. Sugiura, W. and Hoshino, T. (**2006**) Computational simulations of HIV-1 Proteases-multi-drug resistance due to non-active site mutation L90M. *J. Am. Chem. Soc.* 128: 7887-7895.

TH-772_04610604

- Ohtaka, H. Schon, A. and Freire, E. (**2003**) Multidrug resistance to HIV-1 protease inhibition requires cooperative coupling between distal mutations. *Biochemistry* 42: 13659-13666.
- Okimoto, N. Tsukui, T. Hata, M. Hoshino, T. and Tsuda, M. (**1999**) Hydrolysis mechanism of the Phenylalanine−Proline peptide bond specific to HIV-1 Protease: Investigation by the *ab initio* Molecular Orbital Method. *J. Am. Chem. Soc.* 121: 7349-7354.
- Okimoto, N. Tsukui, T. Kitayama, K. Hata, M. Hoshimo, T. and Tsuda, M. (**2000**) Molecular Dynamics study of HIV-1 Protease-Substrate complex: Roles of the water molecules at the loop structures of the active site. *J. Am. Chem. Soc.* 122: 5613–5622.
- Paci, E. and Marchi, M. (**1996**) Intrinsic compressibility and volume compression in solvated proteins by molecular dynamics simulation at high pressure. *Proc. Natl. Acad. Sci.* 93: 11609-11614.
- Paci, E. (**2002**) High pressure simulations of biomolecules. *Biochim. Biophys. Acta*, 1595: 185-200.
- Perryman, A. L. McCammon, J. H. and Lin, J. A. (**2004**) HIV-1 protease resistance 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.* 13: 1108-1123.
- Perryman, A. L. Lin, J. and McCammon, J. A. (**2006**) Optimization and computational evaluation of a series of potential active site inhibitors of the V82F/I84V drug-resistant mutant of HIV-1 Protease: an application of the Relaxed Complex Method of structure-based drug design. *Chem. Biol. Drug Des*. 67: 336-345.
- Pettersen, E. F. Goddard, T. D. Huang, C. C. Couch, G. S. Greenblatt, D. M., Meng E. C. and Ferrin, T. E. (**2004**) UCSF Chimera - A visualization system for exploratory research and analysis**.** *J. Comput. Chem.* 25: 1605-1612.
- Pettit, S. C. Lindquist, J. N. Kaplan, A. H. and Swanstrom, R. (**2005**) Processing sites in the human immunodeficiency virus type 1 (HIV-1) Gag-Pro-Pol precursor are cleaved by the viral protease at different rates. *Retrovirology* 2: 66-71.

TH-772_04610604

- Pettit, S. C. Sheng, N. Tritch, R. Erickson-Viitanen, S. and Swanstrom, R. (**1998**) The regulation of sequential processing of HIV-1 Gag by the viral protease. *Adv. Exp. Med. Biol.* 436: 15-25.
- Piana, S. Carloni, P. and Parrinello, M. (**2002a**) Role of conformational fluctuations in the enzymatic reaction of HIV-1 protease. *J. Mol. Biol.* 319: 567-583.
- Piana, S. Carloni, P. and Rothlisberger, U. (**2002b**) Drug resistance in HIV-1 protease: Flexibility-assisted mechanism of compensatory mutations. *Protein Sci.* 11: 2393-2402.
- Polgár, L. Szeltner, Z. and Boros, I. (**1994**) Substrate-dependent mechanisms in the catalysis of Human Immunodeficiency Virus Protease. *Biochemistry* 33: 9351-9357.
- Poorman, R. A. Tomasselli, A. G. Heinrikson, R. L. and Kezdy, F. L. (**1991**) A cumulative specificity model for proteases from human immunodeficiency virus types 1 and 2, inferred from statistical analysis of an extended substrate data base. *J. Biol. Chem.* 266: 14554-14561.
- Privalov, P. L. and Gill, S. J. (**1989**) The hydrophobic effect: a reappraisal. *Pure. Appl. Chem.* 61: 1097-1104.
- Rahman, A. (**1964**). Correlations in the Motion of Atoms in Liquid Argon. *Phys. Rev. A* 136: 405-411.
- Ramachandran, G. N. and Sasiskharan, V. (**1968**) Conformation of Polypeptides and Proteins. *Adv. Protein Chem.* 23: 283-437.
- Rao, B. G. Tilton, R. F. and Singh, U. C. (**1992**) Free energy perturbation studies on inhibitor binding to HIV-1 proteinase. *J. Am. Chem. Soc.* 114: 4447-4452.
- Rasmussen, B. F. Stock, A. M. Ringe, D. and Petsko, G. A. (**1992**) Crystalline ribonuclease A loses function below the dynamical transition at 220 K. *Nature* 357: 423 -424.
- Refaee, M. Tezuka, T. Akasaka, K. and Williamson, M. P. (**2003**) Pressure-dependent changes in the solution structure of Hen Egg-white Lysozyme. *J .Mol. Biol.* 327: 857-865.
- Ren, P. and Ponder, J. W. (**2003**) Polarizable atomic multipole water model for Molecular Mechanics simulation. *J. Phys. Chem. B* 107: 5933-5947.

TH-772_04610604

- Reiling, K. K. Endres, N. F. Dauber, D. S. Craik, C. S. and Stroud, R. M. (**2002**) Anisotropic dynamics of the JE-2147-HIV Protease Complex: Drug resistance and Thermodynamic Binding mode examined in a 1.09 Å Structure. *Biochemistry* 41: 4582-4594.
- Richards, F. M. (**1977**) Areas, volumes, packing and protein structure. *Annu. Rev. Biophys. Bioeng.* 6.
- Rick, S. W. Stuart, S. J. and Berne, B. J. (**1994**) Dynamical fluctuating charge force fields: application to liquid water. *J. Chem. Phys.* 101: 6141-6156.
- Ridky, T. W. Kikonyogo, A. Leis, J. Gulnik, S. Copeland, T. Erickson, J. W. Wlodawer, A. Kurinov, I. Harrison, R. W. and Weber, I. T. (**1998**) Drug-resistant HIV-1 Proteases identify enzyme residues important for substrate selection and catalytic rate. *Biochemistry* 37: 13835-13845.
- Roder, H. and Shastry, M. C. R. (**1999)** Methods for exploring early events in protein folding. *Curr. Opin. Struct. Biol*. 9: 620-626.
- Rose, R. Craik, C. S. Douglas, N. L. and Stroud, R. M. (**1996**) Three-Dimensional structures of HIV 1 and SIV Protease product complexes. *Biochemistry* 35: 12933-12944.
- Rose, R. B. Craik, C. S. and Stroud, R. M. (**1998**) Domain flexibility in retroviral proteases: Structural implications for drug resistance mutations. *Biochemistry* 37: 2607–2621.
- Ryckaert, J-. P. Ciccotti, G. and Berendsen, H. J. C. (**1977**). Numerical integration of the Cartesian equations of motion of a system with constraints: Molecular Dynamics of *n*-Alkanes. *J. Comput. Phys.* 23: 327–341.
- Sadiq, S. K. Wan, S. and Coveney, P. V. (**2007**) Insights into a mutation-assisted lateral drug escape mechanism from the HIV-1 Protease active site. *Biochemistry* 46: 14865-14877.
- Sanbonmatsu, K. Y. Joseph, S. and Tung, C. (**2005**) Simulating movement of tRNA into the ribosome during decoding. *Proc. Natl. Acad. Sci.* 102: 15854-15859**.**
- Sandberg, W. S. and Terwillinger, T. C. (**1991**) Energetics of repacking a protein interior. *Proc. Natl. Acad. Sci.* 88: 1706-1710.

TH-772_04610604

- Sayle, R. A. and Milner Whilte, E. J. (**1995**) RASMOL: Biomolecular graphics for all. *Trends Biochem. Sci.* 20: 374-382.
- Schaefer, M. Bartels, C. and Karplus, M. (**1998**) Solution conformations and thermodynamics of structured peptides: molecular dynamics simulation with an implicit solvation model. *J. Mol. Biol.* 3: 835-848.
- Scott, W. R. and Schiffer, C. A. (**2000**) Curling of flap tips in HIV-1 protease as a mechanism for substrate entry and tolerance of drug resistance. *Struct. Fold. Des.* 8: 1259-1265.
- Seelmeier, S. Schmidt, H. Turk, B. and von der Helm, K. (**1988**) Human immunodeficiency virus has an aspartic-type protease that can be inhibited by pepstatin A. *Proc. Natl. Acad. Sci.* 85: 6612-6616.
- Seibold, S. A and Cukier, R. I. (**2007**) A molecular dynamics study comparing a wild-type with a multiple drug resistant HIV protease: Differences in flap and aspartate-25 cavity dimensions. *Proteins* 69: 551-565.
- Serrano, L. Kellis, J. T. J. Cann, P. Matoushek, A. and Fersht, A. R. (**1992**) The folding of an enzyme. II. Substructure of barnase and the contribution of different interactions to protein stability. *J. Mol. Biol.* 224: 783-804.
- Shafer, R. W. Chuang, T. K. Hsu, P. White, C. B. and Katzenstein, D. A. (**1999**) Sequence and drug susceptibility of subtype C protease from human immunodeficiency virus type I seroconverters in Zimbawe. *AIDS Res. Hum. Retroviruses* 15: 65-69.
- Shaffer, L. and Verkhivker, G. M. (**1998**) Predicting structural effects in HIV-1 protease mutant complexes with flexible ligand docking and protein side-chain optimization. *Proteins* 33: 295-310.
- Shirley, B. A Stanssens, P. Hahn, U. and Pace, C. N. (**1992**) Contribution of hydrogen bonding to the conformational stability of ribonuclease T1. *Biochemistry* 31: 725-732.
- Shortle, D. Stites, W. E. and Meeker, A. K. (**1990**) Contributions of the large hydrophobic amino acids to the stability of staphylococcal nuclease. *Biochemistry* 29: 8033-8041.
- Silva, A. M. Cachau, R. E. Sham, H. L. and Erickson, J. W. (**1996**) Inhibition and catalytic mechanism of HIV-1 aspartic protease. *J. Mol. Biol.* 255: 321-346.

TH-772_04610604

- Singh, G. and Senapati, S. (**2008**) Molecular Dynamics simulations of ligand-induced flap closing in HIV-1 Protease approach X-ray resolution: Establishing the role of bound water in the flap closing mechanism. *Biochemistry* 47: 10657-10664.
- Skiba, N. P. and Hamm, H. E. (**1998**) How Gsα activates adenylyl cyclase. *Nat. Struct. Biol.* 05: 88-92.
- Smith, R. Brereton, I. M. Chai, R. Y. and Kent, S. B. H. (**1996**) Ionization states of the catalytic residues in HIV-1 protease. *Nat. Struct. Biol.* 3: 946-950.
- Spinelli, S. Liu, Q. Z. Alzari, P. M. Hirel, P. H. and Poljak, R. J. (**1991**) The three-dimensional structure of the aspartyl protease from the HIV-1 isolate BRU. *Biochimie.* 73: 1391-1396.
- Stillinger, F. H. and Rahman, A. (**1974**). Improved simulation ofliquid water by Molecular Dyanamics. *J. Chem. Phys.* 60: 1545-1557.
- Stillinger, F. H. and Weber, T. A. (**1988**) Molecular dynamics simulation for chemically reactive substances. Fluorine. *J. Chem. Phys.* 8: 5123-5133.
- Stryer L. (**1998**) *Biochemistry*. 3rd Ed. New York: W. H. Freeman and co.Genick, USA.
- Suzuki, T. Murakami, T. Iino, R. Suzuki, J. Ono, S. Shirakihara, Y. and Yoshida, M. (**2003**) F_0F_1-ATPase/Synthase is geared to the synthesis mode by conformational rearrangement of ε-subunit in response to proton motive force and ADP/ATP balance. *J. Biol. Chem.* 278: 46840- 46846.
- Swain, A. Miller, M. M. Green, J. Rich, D. H. Schneider, J. Kent, S. B. and Wlodawer, A. (**1990**) X-Ray crystallography Structure of a complex between a synthetic protease of Human Immunodeficiency virus 1 and a substrate based Hydroxyethylamine inhibitor. *Proc. Natl. Acad. Sci.* 87: 8805-8809.
- Tawa, G. J. Topol, I. A. Burt, S. K. and Erickson, J. W. (**1998**) Calculation of relative binding free energies of peptidic inhibitors to HIV-1 Protease and its I84V mutant. *J. Am. Chem. Soc.* 120: 8856-8863.
- Taylor, R. and Kennard, O. (**1984**) Hydrogen-bond geometry in organic crystals. *Acc. Chem. Res.* 17: 320-326.
- Thaisrivongs, S. Skulnick, H. I. Turner, S. R. Strohback, J.W. Tommasi, R. A. Johnson, P. D. Aristoff, P. A. Judge, T. M. Gammill, R. B. and Morris, J. K. (**1996**)

TH-772_04610604

Structure-based design of HIV protease inhibitors: sulfonamide-containing 5,6 dihydro-4-hydroxy-2-pyrones as non-peptide inhibitors. *J. Med. Chem.* 39: 4349.

- Todd, J. M. Semo, N. and Freire, E. (**1998**) The structural stability of the HIV-1 protease. *J. Mol. Biol.* 283: 475-488.
- Todd, J. M. and Freire, E. (**1999**) The effect of inhibitor binding on the structural stability and cooperativity of the HIV-1 protease**.** *Proteins* 36: 147-156.
- Torrie, G. M. and Valleau, J. P. (**1977**). Nonphysical sampling distributions in Monte-Carlo free energy estimation: Umbrella Sampling. *J. Comput. Phys.* 23: 187-199.
- Toth, G. and Borics, A. (**2006**) Flap opening mechanism of HIV-1 protease. *J. Mol. Gr. Mod.* 24: 465-474.
- Tozzini, V. Trylska, J. Chang, C. and McCammon, J. A. (**2007**) Flap opening dynamics in HIV-1 protease explored with a coarse-grained model. *J. Struct. Biol.* 157: 606-615.
- Trbovic, N. Kim, B. Friesner, R. A. and Palmer, A. G. (**2008**) Structural analysis of protein dynamics by MD simulations and NMR spin-relaxation. *Proteins* 71: 684-694.
- Tripos, International, (SYBYL 7.3), 1699 South Hanley Rd., St. Louis, Missouri, 63144, USA
- Tropsha, A. and Hermans (**1992**) Application of free energy simulations to the binding of a transition-state-analogue inhibitor to HTV protease. *Protein Eng.* 5: 29-33.
- Trylska, J. Antosiewicz, J. Geller, M. Hodge, C. N. Klabe, R. M. Head, M. S. and Gilson, M. K. (**1999**) Thermodynamic linkage between the binding of protons and inhibitors to HIV-1 protease. *Protein Sci.* 8: 180-195.
- Trzesniak, D. Lins, R. D. and van Gunsteren, W. F. (**2006**) Protein under pressure: Molecular Dynamics simulation of the Arc repressor. *Proteins,* 65:136-144.
- UNAIDS (**2008**): www.unaids.org.
- Urayama, P. Phillips, Jr. G. N. and Gruner, S. M. (**2002**) Probing substates in sperm whale myoglobin using high-pressure crystallography. *Structure* 10:51-60.

TH-772_04610604

- Van der Spoel, D. Lindahl, E. Hess, B. van Buuren, A. R. Apol, E. Meulenhoff, P. J. Tieleman, D. P. Sijbers, A. L. T. M. Feenstra, K. A. van Drunen, R. and Berendsen, H. J. C. (**2004**) Gromacs User Manual version 3.2.
- Velasquez-Campoy, A. Todd, J. M. and Freire, E. (**2000**) HIV-1 protease inhibitors: enthalpic versus entropic optimization of the binding affinity. *Biochemistry* 39: 2201-2207.
- Venturini, A. López-Ortiz, F. and Alvarez, J. M. G. (**1998**) Theoretical proposal of a catalytic mechanism for the HIV-1 protease involving an enzyme-bound tetrahedral intermediate. *J. Am. Chem. Soc.* 120: 1110-1111.
- Verlet, L. (**1967**). Computer "Experiments" on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules. *Physical Rev.* 159: 98 -103.
- Wang, W. and Kollman, P. A. (**2000**) Free energy calculations on dimer stability of the HIV protease using molecular dynamics and a continuum solvent model. *J. Mol. Biol.* 303: 567-582.
- Wang, Y. X. Freedberg, D. I. Grzesiek, S. Torchia, D. A. Wingfield, P.T. Kaufman, J. D. Stahl, S. J. Chang, C. H. and Hodge, C. N. (**1996a**) Mapping hydration water molecules in the HIV-1 protease/DMP323 complex in solution by NMR spectroscopy. *Biochemistry* 35: 12694-12704.
- Wang, Y. X. Freedberg, D. I. Yamazaki, T. Wingfield, P. T. Stahl, S. J. Kaufman, J. D. Kiso, Y. and Torchia, D. A. (**1996b**) Solution NMR evidence that the HIV-1 Protease catalytic aspartyl groups have different ionization states in the complex formed with the asymmetric drug KNI-272. *Biochemistry* 35: 9945-9950.
- Warshall, A. and Russel, S. T. (**1984**) Calculations of electrostatic interactions in biological systems and in solutions. *Quart. Rev. Biophys.* 17: 283- 422.
- Weber, I. T. Miller, M. Jaskolski, M. Leis, J. Skalka, A. M. and Wlodawer, A. (**1989**) Molecular modeling of the HIV-1 protease and its substrate binding site. *Science* 243: 928-931.
- Weber T. A. and Stillinger, F. H. (**1984**). The effect of density on the inherent structure in liquids. *J. Chem. Phys.* 80: 2742-2746.
- Wittayanarakul, K. Aruksakunwong, O. Saen-oon, S. Chantratita, W. Parasuk, V. Sompurnpisut, P. and Hannongbua, S. (**2005**) Insights into saquinavir resistance in

TH-772_04610604

the G48V HIV-1 protease: Quantum calculations and MD simulations. *Biophys. J.* 88: 867-879.

- Wlodawer, A. Miller, M. Jaskolski, M. Sathyanarayana, B. K. Baldwin, E. Weber, I. T. Selk, L. M. Clawson, L. Schneider, J. and Kent, S. B. (**1989**) Conserved folding in retroviral proteases: crystal structure of a synthetic HIV-1 protease. *Science* 245: 616–621.
- Wlodawer, A. and Erickson, J. W. (**1993**) Structure-based inhibitors of HIV-1 protease. *Annu. Rev. Biochem.* 62: 543-585.
- Wlodawer, A. and Vondrasek, J. (**1998**) INHIBITORS OF HIV-1 PROTEASE: A majorsuccess of structure-assisted drug design. *Ann. Rev. Biophys. Biomol. Struct.* 27: 249-284.
- Wroblowski, B. Diaz, J. F. Heremans, K. and Engelborghs, Y. (**1996**) Molecular mechanisms of pressure induced conformational changes in BPTI. *Proteins* 25: 446-455.
- Yamagata, Y. Kubota, M. Sumikawa, Y. Funahashi, J. Takano, K. Fujii, S. and Yutani, K. (**1998**) Contribution of hydrogen bonds to the conformational stability of human lysozyme: Calorimetry and X-ray analysis of six Tyr-Phe mutants. *Biochemistry* 37: 9355-9362.
- Yamato, T. Higo, J. Seno, Y. and Gō, N. (**1993**) Conformational deformation in deoxymyoglobin by hydrostatic pressure. *Proteins* 16: 327-340.
- Yamazaki, T. Nicholson, L. K. Torchia, D. A. Wingfield, P. T. Stahl, S. J. Kaufman, J. D. Eyermann, C. J. Hodge, C. N. Lam, P. Y. S. Ru, Y. Jadhav, P. K. Chang, C. H. and Weber, P. C. (**1994a**) 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.* 116: 10791-10792.
- Yamazaki, T. Nicholson, L. K. Torchia, D. A. Stahl, S. J. Kaufman, J. D. Wingfield, P. T. Domaille, P. J. and Campbell-Burk, S. (**1994b**) Secondary structure and signal assignments of human-immunodeficiency-virus-1 protease complexed to a novel, structure-based inhibitor. *Eur. J. Biochem.* 219: 707-712.
- Yamazaki, T. Hinck, A. P. Wang, Y. X. Nicholson, L. K. Torchia, D. A. Wingfield, P. Stahl, S. J. Kaufman, J. D. Chang, C. H. Domaille, P. J. and Lam, P. Y. (**1996**).

TH-772_04610604

Three dimensional solution structure of the HIV-1 protease complexed with DMP323, a novel cyclic urea-type inhibitor, determined by nuclear magnetic resonance spectroscopy. *Protein Sci.* 5: 495-506.

- Yong, J. A. T. (**1998**) Much more than just Gag, Pol and Env! (Book Review) *Trends in Genetics* 14: 1- 475.
- Yoshimura, K. *et al.* (**1999**) JE-2147: A dipeptide protease inhibitor (PI) that potently inhibits multi-PI-resistant HIV-1. *Proc. Natl. Acad. Sci.* 96: 8675-8680.
- Yu, M. H. Weissman, J. S. and Kim, P. S. (**1995**) Contribution of individual side-chains to the stability of BPTI examined by alanine-scanning mutagenesis. *J. Mol. Biol.* 249: 388-397.
- Yutani, K. Ogasahara, K. Tsujita, T. and Sugino, Y. (**1987**) Dependence of conformational stability on hydrophobicity of the amino acid residue in a series of variant proteins substituted at a unique position of tryptophan synthase alpha subunit. *Proc. Natl. Acad. Sci.* 84: 4441- 4444.

TH-772_04610604

Publications

TH-772_04610604

List of Publications:

1) Pradipta Bandyopadhyay & **B.R. Meher**. Drug resistance of HIV-1 protease against JE-2147: I47V mutation investigated by molecular dynamics simulation (**2006**). *Chemical Biology and Drug Design*, 67: 155-161.

2) **B.R. Meher**, M.V. Satish Kumar & Pradipta Bandyopadhyay. Molecular dynamics simulation of HIV-1 protease with polarizable and non-polarizable force fields (**2009**). (Special issue in Simulation) *Indian Journal of Physics*, 83: 81-90.

3) **B.R. Meher**, M.V. Satish Kumar & Kausik Sen. Pressure effects on protein conformation: A Molecular Dynamics simulation study of HIV-1 protease. (*Under revision*)

4) **B.R. Meher**, Smriti Sharma, M.V. Satish Kumar & Pradipta Bandyopadhyay Conformational dynamics of HIV-1 protease: A comparative study with multiple AMBER force fields. (*Communicated*)

International Conference Presentations:

1) **B.R. Meher**, M. V. Satish Kumar & Pradipta Bandyopadhyay Effect of Simulation protocol and force field on the flap dynamics of HIV-1 protease. *5th International Conference on Bioinformatics (INCOB-**2006**)*, held on 18th -20th December, 2006 at New Delhi, India.

2) **B.R. Meher**, M. V. Satish Kumar & Pradipta Bandyopadhyay. Flap dynamics of HIV-1 protease: A comparative study with multiple AMBER force fields. *International Conference on Bioinformatics (BIOCONVENE-**2007**)*, held on 16th -22nd December, 2007 at Hyderabad, India.

3) **B.R. Meher**, M. V. Satish Kumar & Kausik Sen. Pressure induced conformational dynamics of HIV-1 protease: A Molecular Dynamics simulation study. *International Conference on Information Technology (ICIT-**2008**)*, held on 17th-20th December, 2008 at Bhubaneswar, Orissa, India.

TH-772_04610604

List of Publications:

1) Pradipta Bandyopadhyay & **B.R. Meher**. Drug resistance of HIV-1 protease against JE-2147: I47V mutation investigated by molecular dynamics simulation (**2006**). *Chemical Biology and Drug Design*, 67: 155-161.

2) **B.R. Meher**, M.V. Satish Kumar & Kausik Sen. Pressure induced conformational dynamics of HIV-1 protease A Molecular Dynamics simulation study. (**2008**) *IEEE CS Press* 118-122.

3) **B.R. Meher**, M.V. Satish Kumar & Pradipta Bandyopadhyay. Molecular dynamics simulation of HIV-1 protease with polarizable and non-polarizable force fields (**2009**). (Special issue in Simulation) *Indian Journal of Physics*, 83: 81-90.

4) **B.R. Meher**, Smriti Sharma, M.V. Satish Kumar & Pradipta Bandyopadhyay Conformational dynamics of HIV-1 protease: A comparative study with multiple AMBER force fields. (*Communicated*)

International Conference Presentations:

1) **B.R. Meher**, M. V. Satish Kumar & Pradipta Bandyopadhyay Effect of Simulation protocol and force field on the flap dynamics of HIV-1 protease. *5th International Conference on Bioinformatics (INCOB-**2006**)*, held on 18th -20th December, 2006 at New Delhi, India.

2) **B.R. Meher**, M. V. Satish Kumar & Pradipta Bandyopadhyay. Flap dynamics of HIV-1 protease: A comparative study with multiple AMBER force fields. *International Conference on Bioinformatics (BIOCONVENE-**2007**)*, held on 16th -22nd December, 2007 at Hyderabad, India.

3) **B.R. Meher**, M. V. Satish Kumar & Kausik Sen. Pressure induced conformational dynamics of HIV-1 protease: A Molecular Dynamics simulation study. *International Conference on Information Technology (ICIT-**2008**)*, held on 17th-20th December, 2008 at Bhubaneswar, Orissa, India.

TH-772_04610604

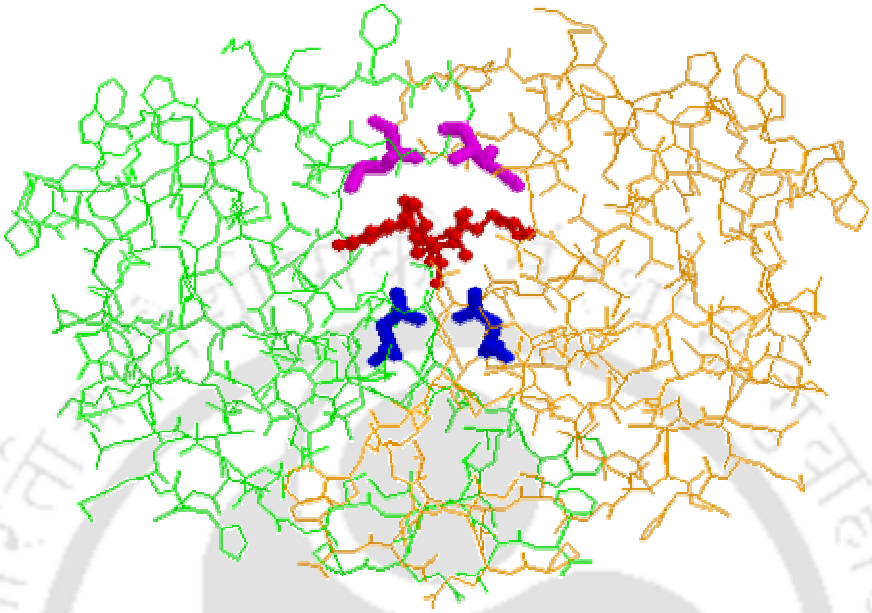


Figure: 3S1. *Structure of JE-2147 complexed HIV-1 protease showing the local fluctuation sites. The protease monomers are colored orange and green. JE-2147 is shown in red. The paired active site Asp residues (Asp25, Asp124) are shown in blue, and the flap tip residues are shown in purple (Ile 50,Ile149). The residues highlighted in blue and purple are the locations of residues involved in local fluctuations.*

Indian Institute of Technology Guwahati

The physiological relevance of study of protein under high pressure can be understood from the discussion below.

The application of pressure to a protein solution provides a way to perturb the structure of the protein and its interactions with solvent in a continuous and controlled way. It is expected that the application of pressure will reduce the packing defects in a protein which will lead to more compact conformations and perhaps to reduced mobility of the atoms within the protein also. MD simulations represent a powerful approach to study the relationship between protein conformation and atomic mobility as has been demonstrated in several studies at ambient pressures [McCammon and Harvey, 1988].

It can be found several evidences from the literature on account of studying proteins at high pressure. In biophysical chemistry, pressure has long been used as an environmental variable to explore the interactions of proteins with ligands and to study the protein dynamics, conformational equilibria and other properties of the native state of the proteins [Weber and Drickamer, 1983; Heremans, 1982; Jannaschs *et al.*, 1987]. Studies of pressure effects on proteins have also revealed that proteins unfold at high pressures [Zipp and Kauzzmann, 1973; Li *et al.* 1976; Carrier *et al.*, 1990]. Pressure induced unfolding has been detected both by optical methods and by hydrogen exchange [Wong 1991]. X-ray crystallography study of a protein at modestly elevated pressure (1kbar) has been reported which provides information about the anisotropic response of the protein to applied pressure [Kundrot and Richards, 1987]. Pressure studies of proteins have also been used to probe the physical properties of the protein hydration layer [Kundrot and Richards, 1988]. Based on the literature evidence, we have accomplished MD simulations of HIV-pr at high pressure (3kbar) and compared the results with a corresponding simulation at low pressure (1bar).

TH-772_04610604

References:

Carrier, D., Mantsch, H., & Wong, P. T. T. (**1990**) Pressure-induced reversible changes in secondary structure of poly(L-lysine): An IR spectroscopic study. *Biopolymers.* 29:837-844.

Heremans, K. (**1982**) *Annu. Rev. of Biophys. Bioeng.* 11: 1-54.

Jannasch, H. W., Marquis, R. E., & Zimmerman, A. M, Eds. (**1987**) *Current Perspectives in High Pressure Biology,* Adenine Press, New York.

Kundrot, C. E., & Richards, F. M. (**1987**) Crystal structure of hen egg-white lysozyme at a hydrostatic pressure of 1000 atmospheres. *J.* Mol. *Biol.* 193:157-170.

Kundrot, *C.* E., & Richards, F. M. (**1988**) Effect of hydrostatic pressure on the solvent in crystals of hen egg-white lysozyme. *J. Mol. Biol.* 200: 401-410.

Li, T., Hook, J. W., Drickamer, H. G., & Weber, G. (**1976**). Plurality of pressure-denatured forms in chymotrypsinogen and lysozyme *Biochemistry* 15: 5571-5580.

McCammon, J. A., and S. C. Harvey. (**1988**). Dynamics of proteins and nucleic acids. Cambridge University Press, Cambridge, MA.

Weber, G. and H.G. Drickamer, (**1983**) The effect of high pressure upon proteins and other biomolecules. *Q. Rev. Biophys.* 16: 89–112.

Wong, P. T. T. (**1991**) Pressure effect on hydrogen isotope exchange kinetics in chymotrypsinogen investigated by FT-IR spectroscopy. *Can.* J. *Chem.* 69:1699-1704.

Zipp, A., and Kauzmann, W. (**1976**) Pressure denaturation of Metmyoglobin. *Biochemistry* 12: 4217-4228.

TH-772_04610604