A DISSERTATION ON

Protein-Ligand Free Energy Calculation Through Enhanced Sampling Method

SUBMITTED TO THE
DEPARTMENT OF BIOENGINEERING
FACULTY OF ENGINEERING
INTEGRAL UNIVERSITY, LUCKNOW

IN PARTIAL FULFILMENT
FOR THE
DEGREE OF MASTERS OF TECHNOLOGY
IN BIOINFORMATICS

BY

Azra Shuaib
M.Tech Bioinformatics (IV Semester)
Roll No: 2001081002

UNDER THE SUPERVISION OF

Dr. Harshwardhan H. Katkar
Assistant Professor
Indian Institute of Technology Kanpur

INTEGRAL UNIVERSITY, DASAULI, KURSI ROAD

LUCKNOW- 226026

DECLARATION FORM

I, **Azra Shuaib**, a student **M. Tech Bioinformatics** (II Year/ IV Semester), Integral University have completed my six months dissertation work entitled **"Protein-Ligand Free Energy Calculation Through Enhanced Sampling Method"** successfully from **Indian Institute of Technology Kanpur** under the able guidance of **Dr. Harshwardhan H. Katkar.**

I, hereby, affirm that the work has been done by me in all aspects. I have sincerely prepared this project report and the results reported in this study are genuine and authentic.

Azra Shuaib

Date-

Course Coordinator

Dr. Mohammad Kalim Ahmad Khan

INTEGRAL UNIVERSITY

Established Under the Integral University Act 2004 (U.P. Act No.9 of 2004)
Approved by University Grant Commission
Phone No.: +91(0522) 2890812, 2890730, 3296117, 6451039, Fax No.: 0522-2890809
Kursi Road, Lucknow-226026 Uttar Pradesh (INDIA)

CERTIFICATE BY INTERNAL ADVISOR

This is to certify that Azra Shuaib, a student of **M. Tech Bioinformatics** (II Year/ IV Semester), Integral University has completed her six months dissertation work entitled **"Protein-Ligand Free Energy Calculation Through Enhanced Sampling Method" successfully**. She has completed this work from **Indian Institute of Technology Kanpur** under the guidance of Dr. Harshwardhan H. Katkar. The dissertation was a compulsory part of her **M. Tech Bioinformatics.**

I wish her good luck and bright future.

Dr. Salman Akhtar
Associate Professor
Department of Bioengineering
Faculty of Engineering

INTEGRAL UNIVERSITY

Established Under the Integral University Act 2004 (U.P. Act No.9 of 2004)
Approved by University Grant Commission
Phone No.: +91(0522) 2890812, 2890730, 3296117, 6451039, Fax No.: 0522-2890809
Kursi Road, Lucknow-226026 Uttar Pradesh (INDIA)

TO WHOM IT MAY CONCERN

This is to certify that Azra Shuaib, a student of **M.Tech Bioinformatics** (II Year/ IV Semester), Integral University has completed her six months dissertation work entitled **"Protein-Ligand Free Energy Calculation Through Enhanced Sampling Method"** successfully. She has completed this work from **Indian Institute of Technology Kanpur** under the guidance of Dr. Harshwardhan H. Katkar. The dissertation was a compulsory part of her **M. Tech Bioinformatics**

I wish her good luck and a bright future.

Dr. Alvina Farooqui
Head
Department of Bioengineering
Faculty of Engineering

CONTENTS

ACKNOWLEDGEMENT

I bow in reverence to the Almighty for blessing me with strong will power, patience and confidence, which helped me in completing the present work.

The satisfaction and euphoria that accompany the accomplishment of any work would be incomplete without the mention of people who made it possible and whose consistent guidance and encouragement crown all the efforts. This project was not only a technical endeavor but also an interesting learning experience.

First of all, I would like to thank my mentor Dr. Harshwardhan H. Katkar for his continuous motivation, help and guidance in materializing the resources and providing accurate and adequate assistance at various stages of the work. I am genuinely grateful to him for a smooth, erudite journey my research experience under him has been. Further, I would like to thank my lab mates Dr. Yogendra, Manisha, Priyanka, Dr. Akhil, Pratyush, Rajat, Akash, and Nishesh for creating a balanced working environment in the lab. Finally, I want to thank all of my friends at Integral University Lucknow who made my stay here one of the best experiences of my life till now.

I would also like to thank my loving parents with whose blessings I was able to achieve my goal successfully. I am thankful to my friends for having made everything possible by giving me strength and confidence to do this extraordinary work.

I gratefully acknowledge Chancellor, Prof. SW Akhtar, Pro chancellor Dr Syed Nadeem Akhtar, Vice Chancellor Prof Javed Mussarat, Registrar Dr. Haris Siddiqui and Controller of Examination, Dr Abdul Rahman Khan of Integral University Lucknow for their insight and constant encouragement. I would like to express my sincere thanks to Head Department of Bioengineering Dr Alvina Farooqui for instilling the confidence and for her constant encouragement. A note of thanks and appreciation also goes to all the faculty members and especially Dr. Ashish of the Department of Bioengineering for their support, help, and guidance.

I extend my heartfelt thanks to my Internal Advisor Salman Akhtar, Associate Professor, Department of Bioengineering for his sincere guidance and valuable suggestions at various steps of this project to bring out the best in me. I am thankful to my Course Coordinator Dr Mohammad Kalim Ahmad Khan for helping me to complete my task and answering my queries.

List of Abbreviations

S.No	Perticulars
SBDD	Structure Based Drug Design
SARS	Severe Accute Respiratory Syndrome
M^{Pro}	Main Protease
MERS	Middle East Respiratory Syndrome
SM	Shape Matching
IC	Incremental Construction
DG	Distance Geometry
EP	Evolutionary Programming
GA	Genetic Algorithm
LGA	Lamarckian Genetic Algorithm
MD	Molecular Dynamics
MetaD	Metadynamics
GROMACS	GROningen MAchine for Chemical Simulations
FES	Free Energy Surface
CV	Collective variable

List of Tables

LIST OF FIGURES

`

1- INTRODUCTION

The essential involvement in the proteolytic processing of viral replicase polyprotein (Pillaiyar, T., Manickam el.al 2016) makes coronavirus protease, M^{pro} (nsp5, 3CLpro) a prime target for designing the leads against coronavirus Main protease. The challenging process of generating and optimizing a new drug candidate against the target M^{pro} can be sped up via computational drug discovery. Due to the sheer explosive growth of faster architectures and improved algorithms for large computations in a time-affordable way, the influence of computational structure-based drug design on drug discovery has increased in the last decade. While the primary goal is the development of molecules that can bind to the protein receptor strongly and selectively, the capability to estimate protein-ligand binding affinity with accuracy is critical for enhancing our knowledge of molecular recognition and rational drug design (D. L. Mobley and M. K. Gilson et. Al 2017, J. D. Chodera, D. L. Mobley, M. R. Shirts et. Al 2011). It is vital to carefully select and have salient information about druggable protein binding pockets from all the cavities that might dynamically appear or permanently be present on the protein surface. This work revealed a novel pocket, outside the active site (referred to as non-active site in the rest of this thesis) and evaluated its druggability by calculating its binding free energy with a pre-selected drug and the rate of binding of this drug with the active site vs the non-active site. To achieve this goal, docking studies were exploited, followed by Molecular dynamics simulations to refine the docking results. Enhanced sampling was later performed to estimate the binding free energy.

To evaluate and rationalize ligand-receptor docking, there are several rapidly growing numbers of methods available. There is a difference between quick, simplistic approaches and more precise, computationally costly methods. Approaches in the former group frequently imply a rigid receptor or ligand but are centered on effective potentials that represent the ligand-receptor interactions. The latter group techniques are significantly more costly computationally because it makes use of flexible ligands and receptors, implicit or explicit water solvation, and interaction Hamiltonians that are more complex.

Molecular docking is an extensively used technique in Computer-Aided Drug Design (Forli, Stefano. 2015), with the goal of dramatically lowering the cost of developing new medications. A molecular docking algorithm can be divided into two primary jobs: i) the search for potential ligand binding poses to the chosen protein target(s); ii) the rank of all

generated poses using a particular "scoring function" that should best reflect the complex's free energy of binding. Tasks i and ii can be executed in different ways, but they are both vulnerable to failure due to the inherent constraints of docking algorithms. The inability to sample conformations of the binding site within the existence of ligands and the challenges associated with accurately accounting for entropic and de-solvation effects in the scoring functions are two examples of these constraints [Forli, Stefano 2015, Mandal, Soma, and Sanat K. Mandal 2009, Chen, Yu-Chian 2015]. Recent advancements in computational power, flexible dockingand enhanced sampling methods, classical force-field accuracy, and simulation parameters have enabled precise and consistent estimation of protein-ligand binding free energies to guide small molecule drug discovery. Molecular dynamics (MD) and Monte Carlo (MC) simulations are two computationally intensive approaches for exploring the flexibility of biomolecular systems.

Molecular dynamics has gained much importance as a research methodology for studying dynamics of systems with millions of atoms. Free energy estimated using an MD simulation trajectory, however, is only reasonably accurate if the system is ergodic in the simulation's timescale and the simulation is prolonged enough for the system to explore all the energetically significant configurations. This isn't always the case, though, in systems based on real-world applications. The important configurations (e.g. metastable states) may be separated by large free-energy barriers that prevent exploring the transition between these configurations with limited computational resources, which is a common cause of the problem. In this situation, the system can only move from one metastable state to another if those infrequent fluctuations push the system over the free energy. A similar issue is faced when the system slowly diffuses in configurational space. Under these circumstances, it is impractical to use reasonable computer time to compute necessary information. In fact, to properly integrate the equations of motion in a typical MD simulation with an atomistic empirical force field, one must require a time step of the order of femtoseconds (10^{-15} seconds). This is necessary to appropriately resolve the system's fastest motion, which is often bond stretching and bond bending. Unfortunately, intriguing things frequently happen over much longer periods of time. For instance, a tiny denatured protein may need more than 10^{-4} seconds to explore the challenging energy environment and discover its folded state. Therefore, approximately 10^{11} MD time steps are required to study such a process. In the instance of a small protein (32000 atoms), a preliminary calculation of the computing cost of a single time step results in the startling estimate that 1 year must be dedicated on a

petaflop computing system to see a single folding event. An even longer computing time would be required if we had to simulate this system from beginning These arguments make it quite evident that MD has a limit on explorable time-frames. Numerous approaches have been suggested in recent years to overcome this limitation. These are called "enhanced sampling approaches." (Chipot, Christophe et.al 2007, Dellago, Christoph et.al 2009) These are often employed to construct the free energy landscape by limiting revisiting of sampled states. With the aid of accelerated sampling through metadynamics, an enhanced sampling method, biologically relevant conformations which are separated by high energy barriers can be explored efficiently. Free energy landscape of the system along a reaction coordinate can be estimated, using which, many relevant quantities such as energy barrier, binding free energy, probability of a given state, and the mechanism of the process can be computed.

Protein-ligand interaction commonly accompanies structural changes in protein which can be investigated in terms of an energy landscape. An energy landscape is a map that includes information about the “states” of the system. It is a projection of the plot of the free energy (to be termed as F) through the configuration space, whose point is defined by the coordinates of each atom (Chong et al., 2019). The funneled landscape concept has recently been used to the study of biomolecular aggregation and binding (Adamcik & Mezzenga, 2018; Hartl & Hayer-Hartl, 2009; Zheng et al., 2012).

Focusing on the molecular recognition of protein and ligand interaction, we present the energy landscape of M^{pro} and its ligand, Dihydroergocristine through enhanced sampling method, namely metadynamics. Two protein-ligand systems were generated by employing molecular docking. One of the systems was generated by identifying a potential biding sites on the protein (a.k.a non-Active site binding pocket) and possible conformations of the ligand binding to these sites/pockets. Another system was generated by identifyning potential ligand binding conformation to the Active site of the protein. FurtherF, we performed molecular dynamics simulation of the retrieved biologically functional unit of M^{pro} followed by the subsequent metadynamics studies of the protein-ligand systems to obtain the free energy landscape of the system. From the energy landscape, the binding free energy of the bound and the unbound states are calculated. The ligand’s binding rate to the active site vs the non-active site was also determined.

Instead of monitoring the positioning of each atom in the system, the collective variables which are the simpler quantities that are able to describe the chemical processes, are used

in various enhanced sampling techniques including metadynamics. We have performed untempered metadynamics simulation along distance between the drug and the binding pocket as the collective variable (CV), and applied the angle restraint to restrict sampling to unbound conformations in a certain direction from the protein, with the direction chosen using structural intuition. GROMACS version 2020.4 patched with PLUMED version 2.7.2 were used for all the simulations that were performed. The free energy landscape are computed in order to gain insights from the protein-ligand simulation.

Objectives of the study:

- Performing molecular docking studies to
    - identify the non-active site of M^{pro} based on binding score
    - generate protein-ligand complex and the most stable ligand conformation in the active site and the non-active site region of the M^{pro}
- Performing molecular dynamics studies to generate an equilibrated conformation of the system
- Performing metadynamics simulation of protein-ligand systems to
    - calculate the binding free energy
    - compare the rates of binding of the ligand at the active site and the non-active site

2- REVIEW OF LITERATURE

2.1 M^{pro}: An alluring target for designing drug for SARS Cov-2

Coronaviruses have indeed proven to be a significant challenge for designing drugs. The seventh coronavirus reported to infect humans is SARS-CoV-2; HKU1, NL63, OC43, and 229E are linked to moderate symptoms, while SARS-CoV, MERS-CoV, and SARS-CoV-2 can cause severe sickness (Corman, V. M et.al 2018).

It is a zoonotic positive-sense single-stranded RNA virus having genetic resemblances to bat coronaviruses and is transmissible to humans (Andersen, K. G., et.al 2020). The angiotensin converting enzyme 2 receptor is the major route through which the virus enters human cells (ACE2) (Shang, Jian, et al. 2020). Over 550 million COVID-19 cases have been recorded to date, and there have been over 6351801 fatalities globally (WHO 2022).

When compared to SARS and MERS-CoV, SARS-CoV-2 possesses a combination of high transmissibility, a longer incubation period, and a significantly narrow window between the onset of symptoms and maximal infectivity (Petersen, E. et al 2020). As a result, while having a low fatality rate, COVID-19 is proving to be far more difficult to eliminate and may therefore continue to be an epidemiological issue until treatment is discovered. Recently, the druggability of the SARS-CoV-2 spike protein has been investigated using molecular dynamics (MD) simulations (Casalino, L. et al. 2020). In the adaptive evolution of SARS coronaviruses, the SARS spike protein has been identified as the major target of selective pressure (Zhang, C.-Y., et.al 2006). However, the receptor-binding domain (RBD) of the spike protein is the most variable region in the coronavirus genome (Zhou, P. et al 2020, Wu, F. et al.) which makes M^{pro} a promising alternative target for designing drugs. As the sequence and structure of the M^{Pro} are closely linked to those from other beta coronaviruses, any inhibitor found for the SARS-CoV-2 M^{Pro} would very likely also act as an inhibitor of the subsequent evolution of this virus, in contrast to immunogenic techniques that use spike protein (Durrant, J. D et.al 2011). Protease inhibitors have been widely utilized in other countries to treat HIV-AIDS (Durrant, J. D et. Al 2011, Deeks, S. G. et al 1999), and hepatitis (Lamarre, D. et al. 2003). Molecular dynamics were preeminent in recognizing free states and the dynamic bond of the HIV-1 protease flaps (two glycine-rich β-hairpins) that enclose a large substrate-binding pocket accepted as a target for antiviral drugs (York, D., Darden 1993). Additionally, the first FDA-approved HIV integrase inhibitor was developed after MD simulations helped locate a mysterious trench

in the HIV integrase (raltegravir) (Durrant, J. D. & McCammon et.al 2011, Schames, J. R. et al).

The virology done on the virus states that the virus has two categories of proteins, structural and non-structural. Structural proteins like nucleocapsids protect the viral single-stranded positive RNA which is up to 33.5 kilobases in length. Other structure proteins like membrane glycoproteins help the virus to bind to ACE II receptors facilitating entry into the host cell (Li, X.; Geng, M.; Peng et.al 2020). The non-structural protein like $3CL^{pro}$ or M^{pro} is a key protease in promoting replication of the virus. It acts on polypeptides, PPs and PPa for further releasing of 16 nonstructural proteins. 3 CL^{pro} is a hugely similar orthologue (96% dentity) to SARS CoV. The structure deposited in the protein data bank (PDB ID: 6Y2F) states that the beta-barrel has two domains and one domain is in the alpha barrel. The active site is located in the midst of two beta-barrel domains. Previous information from reports propounded that 41, 140, 142-145, 161, 163, 166, and 172 majorly contribute to the active site of M^{pro} (Zhang, L., Lin, D., Sun, et.al 2020). Because of the paramount role of M^{pro} in the life cycle of virus and no close homologous relation to humans, it is an alluring target for designing an antiviral against the virus. (Pillaiyar, T., Manickam el.al 2016).

His41-Cys145 is a catalytic dyad that plays a major role in the proteolytic activity of both SARS-CoV and SARS-CoV $2M^{Pro}$ proteins (Ullrich, S. & Nitsche et.al 2020, Jin, Z. et al.). Deranging this dyad by introducing conformational changes impairs M^{pro} activity and successive viral replication and transcription. Abdel-Maksoud, Khaled, et al. 2020 targeted this dyad considering the allosteric mechanism linked to it. The cryptic pockets that were not detectable from the M^{Pro} crystal structure can be discovered as additional pockets for exploring M^{Pro} inhibition beside the active site (Sztain, T., Amaro et.al 2020). Gaussian accelerated molecular dynamics (Miao, Y., Feher et.al 2015) was employed in revealing those pockets.

The molecular docking technique is a crucial tool in computer-assisted drug design to forecast binding affinity and assess the interaction mode. Exploiting this ability of the molecular docking we employed molecular docking in the current study to accomplish a similar approach, results of which were further refined by enhanced sampling through metadynamics study. The idea was to generate the best ligand conformation and predict the

novel regions where drugs can bind. Subsequently, detailed molecular dynamics simulation was performed to analyze the pocket regions.

2.2 Molecular docking

In contemporary drug discovery studies, protein-protein and protein-ligand docking play a vital role in making predictions of the orientation of the ligands when it is attached to the protein receptor using electrostatic interaction and shape to quantify it. The sum of all the interactions like van der Waals, Coulombic, and formation of hydrogen bonds is quantified by a docking score which describes the potentiality of binding. The ligand is explored in a six-dimensional rotational or translational space in the most basic rigid-body systems to sit in the binding site. (Alberg and Schreiber 1993). The original "lock-and-key model" (Morrison, et.al 2006), which makes reference to the rigid docking of receptors and ligands to determine the best orientation for the "key" to open the "lock," is depicted in Figure 1A. The significance of geometric complementarity is emphasized by this model. In reality, however, the docking procedure is believed to be such that flexible that ligands and receptors must alter their conformation to fit one ranothe well. Thus “induced fit model" was proposed (Figure 1) (Koshland Jr et.al 2010).

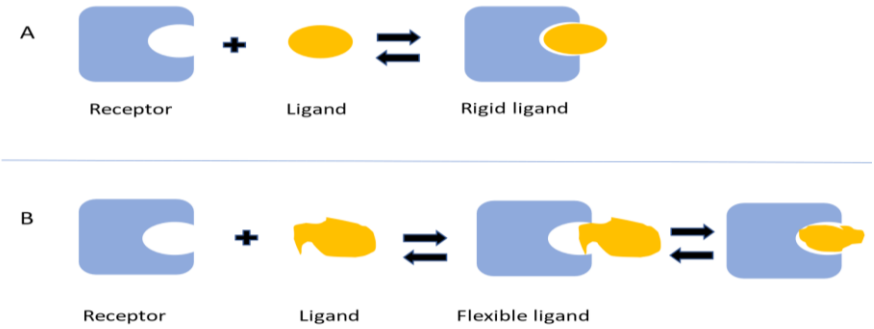


Figure 1: The two models of molecular docking. A: Lock and key model; B: Induced fit model

For a docking simulation to produce a successful result using limited computational resources, speed and accuracy of the algorithm are important. The major goal in developing a docking algorithm is to come up with a quick technique that can find a novel lead molecule (in virtual screening) or replicate an experimental conformation (for confirmation using experimental data) with the highest degree of accuracy, while using finite computational resources. There are many docking programs such as AUTODOCK (Morris, et.al 1998, Goodsell et al 1996), FLEXX (Rarey, M et.al 1996, Kramer et.al 1999), DOCK(Ewing et.al 2001), GOLD (Joy et.al 2006, Rarey et.al 1996), ZDOCK (Chen, et.al

2003), M-ZDOCK (Pierce et.al 2005), MSDOCK (Sauton et.al 2008), Surflex (Jain, et.al 2003), MCDOCK (Liu, M et.al 1999). Every docking programme uses a different search algorithm, such as Incremental Construction (IC) (Rarey et.al 1996, Kramer et.al 1999), Genetic Algorithm (GA) (Morris et.al 1998, Goodsell et.al 1996), Monte Carlo (MC) (Liu, et.al 1999), etc.

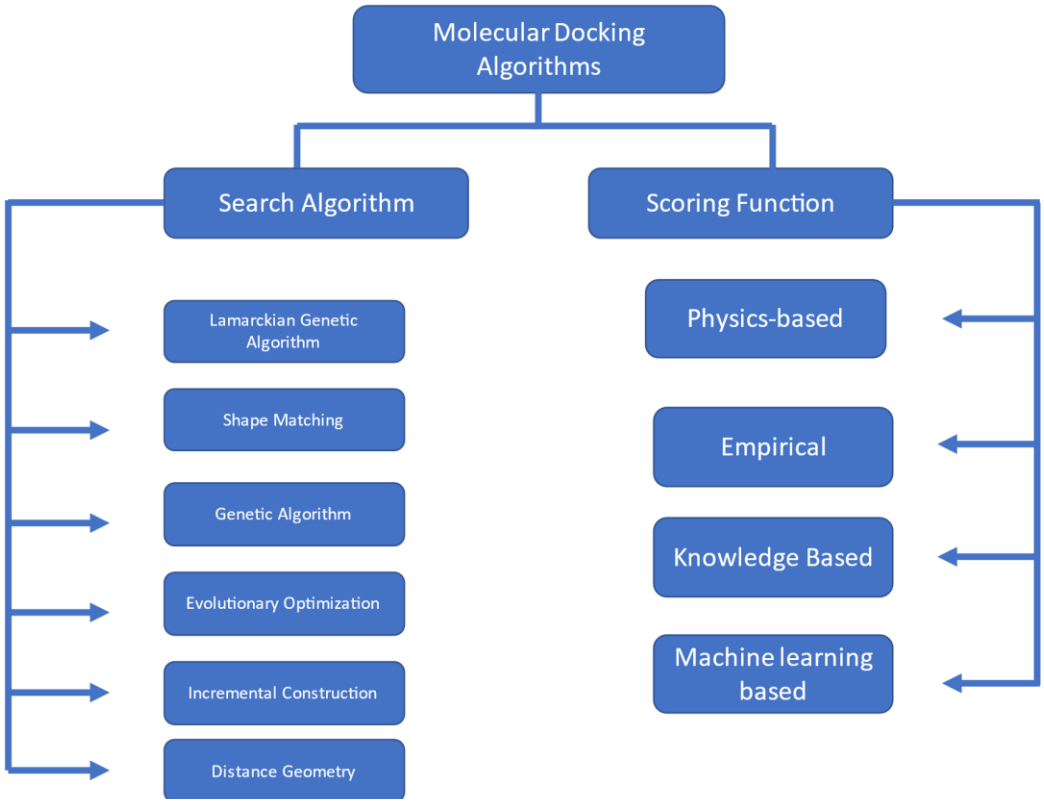


Figure 2: Some popular search and scoring algorithms used in molecular docking

2.2.1 Search Algorithms in Docking

The docking applications can be classified on the basis of their search algorithms. When taking into account the ligand and/or receptor's flexibility, docking methods can be divided into two main categories: Flexible and rigid-body docking. The rigid-body docking method, which primarily considers geometric complementarities among two molecules, and limits the specificity and accuracy of results by not accounting for the flexibility of either the ligand or the receptor. Yet, for several cases, it is able to identify ligand binding sites for proteins (Figure 3) and appropriate conformations of ligand when crosschecked with the crystallographic structure (Perozzo et.al 2002, Pereira et.al 2004, Olveria 2006, Schulze-Gahmen et.al 1996, Chrencik et al 2004, J.E.Timmers et.al 2008). Flexible docking algorithms, at a greater computational cost, can take into account multiple possible ligand

or receptor conformations, as well as conformational flexibility of both the molecules simultaneously. Commonly used algorithms and tools available for rigid-body and flexible docking is summarized in the following subsections.

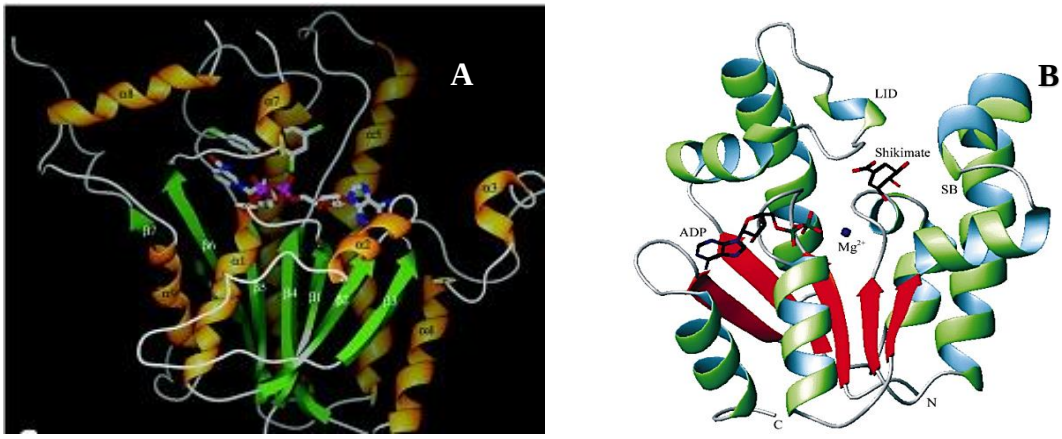


Figure 3: Examples of the successful rigid docking: Best ranked conformations of rigid body docking: A; Enoyl acyl carrier protein in complexed with triclosan (Perozzo, et.al 2002) B; Shikimate kinase in complexed with shikimic acid (Oliveira, et.al 2006)

2.2.1.1 The Fast Shape Matching (SM)

The geometric overlap between two molecules is considered by shape matching algorithms. Applications for rigid-body docking, like ZDOCK (Chen, R.; Li et.al 2003), typically use SM as their primary approach. ZDOCK integrates shape complementarity, electrostatics, and desolvation, parameters through the Fast Fourier Transform algorithm, using atomic coordinates as the only conformation for the docking procedure and taking into account the geometrical surface models. With an RMSD as low as 2.5 Angstroms, it has proven to be highly accurate for protein-protein docking simulations, reproducing three-dimensional experimental data in the CAPRI Challenge (Chen, et.al 2003).

SM approach is frequently used in flexible docking as a part of search strategies such as in SYSDOC , EUDOC, DOCK , and MS DOCK. The initial step in the DOCK process is to identify potential binding site locations where ligand atoms might be present. These areas are also known as "spherical centers."

2.2.1.2 Incremental Construction (IC)

This algorithm breaks the ligand into fragments that are eventually docked independently at the receptor site (Dias, R., de Azevedo et.al 2008) .) The pieces are fused together after the fragments are docked. The fragmentation allows the ligand flexibility to be taken into account. After being docked initially, rigid fragments function as "anchors" that are

subsequently joined by flexible parts (rotatable bonds) of the ligand. The IC algorithm, often referred to as anchor and-grow approaches, has been used in a number of docking applications, including DOCK, FLEXX, FLOG, and Surflex.

2.2.1.3 Distance Geometry (DG)

The information that can be represented through intra- and intermolecular distances is used in this search technique. These distances can be linked, making it possible to use them to calculate structures or conformations (Moré, J.J. et.al 1999). When compared to other algorithms, which often deal with a large number of constraints by imposing extra bounds or by deducing bounds from the provided bounds, the DG approach uses a smaller set of distance constraints.

2.2.1.4 Evolutionary Programming

To resolve problems related to the speed of search, EP algorithms employ computational models evolving natural selection processes. Although there have been many computational models developed, they all share the same idea of replicating the processes of selection, mutation, and reproduction, that depend on the "performance" of individuals (or "chromosomes") of the specie within of an "environment" (Eiben et.al 2003, Fogel, et.al 1998).

2.2.1.5 Genetic Algorithms (GA)

In the instance of molecular docking, GAs attempt to identify the closest conformation of the global energy minimum. GAs is a subclass of EP algorithms that are used to find the exact or approximate answers to search problems. In GA algorithm (Holland J. 1975), the crossover genetic operator is used to combine (mate) two chromosomes (parents) to create a new chromosome (offspring). This new chromosome may be superior to both of its parents if it adopts the most admirable characteristics from both the parents swapping the major region of the parents. In this process, a variety of intricate scoring systems are used while taking into consideration a number of parameters, including the rates of mutation and crossover events as well as the number of evolutionary cycles. GA is implemented by programs like GOLD, DOCK.

2.2.1.6 Lamarckian Genetic Algorithm (LGA)

Lamarckian genetic algorithm is applied by most popular docking program, AUTODOCK (Morris, et.al 1998, Goodsell et.al 1996). The LGA alters between the “phenotypic space” and “genotypic space”. "Genotypic space" and "phenotypic space" are alternated by the

LGA. In genotypic space, mutation and crossover take place, while the phenotypic space is explained by the energy function to be optimized. Energy minimization (local sampling), conceptually related to MC minimization, is carried out in phenotypic space following genotypic modifications to the population (global sampling).

2.2.2 Scoring Function

In general, docking algorithms forecast a number of ligand orientations (poses) inside the binding site. The accuracy of docking hits and orientation is influenced by a number of variables, including the effectiveness of the scoring methods employed for result analysis. There are several scoring functions proposed in the literature (Betzi et.al 2006,Yang et.al 2006, Krammer et.al. 2005, Velec et.al 2005, Wang et.al 1998, Wang et.al 2002 , Zhao et.al 2008, Muryshev et.al 2003 de Azevedo et.al 2008). Numerous scoring functions have been evaluated and compared for accuracy (Wang et.al 2004, Wang et.al 2003). Through a wide range of methods, scoring functions can predict the free energy of binding or the binding constant given the atomic coordinates of a binary complex as input. The Gibbs-Helmholtz equation (von Helmholtz, H. 1882) (eq 2.1) is used to determine the free energy of binding (Du, X., Li. Et.al 2016).

$$\Delta G = \Delta H - T\Delta S \qquad (2.1)$$

G stands for the free energy of binding, H for enthalpy, T for Kelvin temperature, and S for entropy. The following equation illustrates how the binding constant, K_b, and G can be related:

$$\Delta G = -RTlnK_b \qquad (2.2)$$

where K_b is the binding constant.

Scoring functions can be divided into three categories: force-field based, knowledge-based, empirical based, and scoring functions. Scoring functions based on force fields use non-bonded terms from classical mechanics force fields. Scoring functions are employed during docking operations to optimize the positioning of ligands. After the docking process is finished, scoring functions are employed to rank individual ligands discovered. The best affinity ligand will be predicted by this ranking technique.

Target flexibility is an important parameter to consider in the process of drug designing which if not taken care of can result in major artifacts. This makes target flexibility an

important discussion. Despite the enormous applications of docking studies, its inability to treat protein flexibility raises a concern as protein dynamics is a crucial element of intramolecular as well as signaling systems. Unable to sample biologically relevant conformation is just another shortcoming. Hence optimization of docking results is pivotal.

2.3 Target Flexibility: A Critical Consideration in Structure-based Drug Design to counter docking limitations.

Since about the time that Perutz, Kendrew, and colleagues revealed the first X-ray crystal structure of the hemoglobin and myoglobin proteins at near-atomic resolution (Bolton, et.al 1970, Fermi, et.al 1984, Kendrew, et.al 1958, Watson, et.al 1969), structure-based drug discovery has been a significant part of medicinal chemistry (Abraham, D. J. et.al 2007). Hemoglobin has two relatively distinct forms, "tense" and "relaxed," according to its oxygenation. However, in past years, a family of relaxed hemoglobin structures with various tertiary structure conformations has already been found. In fact, flexibility is a property of all protein systems that is commonly required for function (e.g., as in hemoglobin). Under native state conditions, proteins have an inherent capacity to go through functionally significant conformational transitions (Tsai, et.al 1999, Bahar, et.al 2007) across many different time and space scales. (Henzler-Wildman et.al 2007). Nuclear receptors are modular proteins whose biological function depends on a high degree of conformational flexibility. The majority of the pharmacology of nuclear receptor ligands has been described in terms of its ability to displace (or stabilize) the short α-helix segment (known as H12 or AF-2) located at the carboxy end of the receptor in its conformation in the protein "active" state (Krauss et.al 2006, Hellal-Levy et.al 2000, Hu, X et.al 2000). A surprising amount of structural variation is revealed by the available X-ray crystal structures of ligands binding and inhibiting nuclear receptor proteins like the farnesoid X-receptor (FXR) (Downes et.al 2003, Mi, L. Z et.al 2003). Additionally, the function of G-protein-coupled receptors (GPCRs), the biggest known superfamily among membrane proteins, and a significant element of intra- and intermolecular communication signalling systems depends on protein dynamics. Knowing the significance of protein flexibility in SBDD there was a need to optimize docking results by introducing protein flexibility with the help of molecular dynamics simulations.

2.4 Molecular dynamics Simulation

The earlier view of treating proteins as rigid structures have been superseded in recent years by a dynamic model in which the internal motions and resulting conformational changes

are simulated explicitly. In this context, molecular dynamics (MD) simulations have developed into a crucial computational tool for comprehending the physical underpinnings of the structure and function of biological macromolecules. In recent years, molecular dynamics (MD) simulations have had a significantly increased impact on molecular biology and drug development. These simulations capture the atomic level details and temporal behaviour of proteins and other biomolecules. The appeal of biomolecular modelling to experimentalists has expanded as a result of significant advancements in simulation speed, accessibility, and accuracy, as well as the abundance of experimental structural data. Simulations have been useful for discovering the structural origins of disease by unravelling the functional mechanisms of proteins and other biomolecules, and designing and optimizing small molecules, peptides, and proteins (Hollingsworth, S. Aet.al 2018)

Molecular dynamics simulation is the study of simulating the movements of a system of particles. Systems as small as an atom and a diatomic molecule undergoing a chemical reaction and as huge as a galaxy have both been subjected to it (Barnes et.sal 1989). The fundamental concept of an MD simulation is simple and illustrated in Fig.4. One may compute the force applied on every atom by all of the other atoms in a biomolecular system given their positions and an accurate atomistic force-field (for instance, a protein surrounded by water and possibly a lipid bilayer). Thus, it is possible to anticipate each atom's spatial position as a function of time using Newton's equations of motion. To be more specific, one calculates the forces on each atom repeatedly while stepping through time, updating each atom's position and velocity using those forces. In essence, the generated trajectory is a three-dimensional movie that depicts the system's atomic-level configuration at each point during the simulated time period.

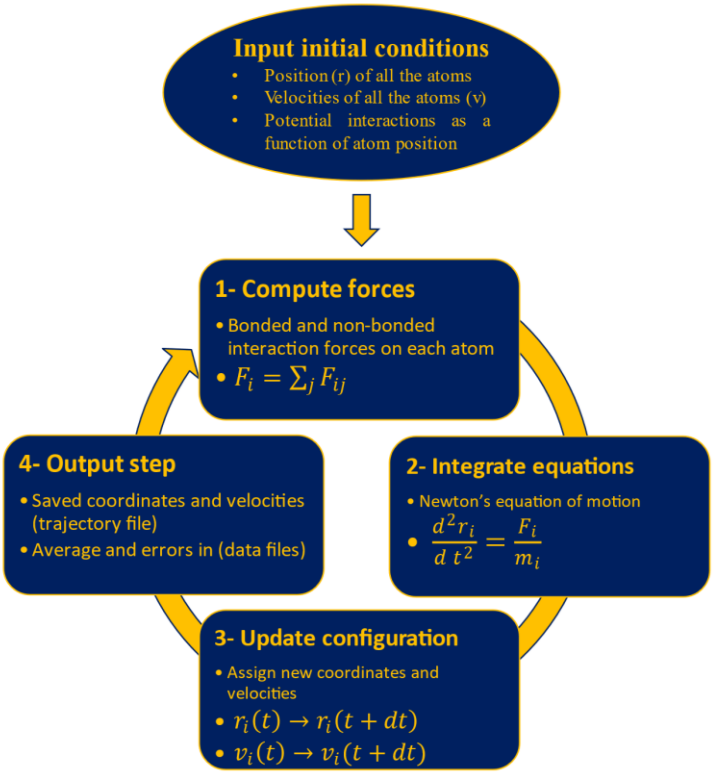


Figure 4: Global Algorithms behind MD

2.4.1 Principle

By integrating Newton's equations of motion, an MD simulation is a method for creating a dynamical trajectory for a system made up of N particles. We would need a set of initial conditions (particle positions and velocities), a decent model to depict the force exerted between the particles (whether from calculations using electronic structure or using the empirical force fields (discussed later)), and a definition of the boundary conditions to be used.

$$m_i \frac{d^2 r_i}{dt^2} = F_i = \frac{-\partial}{\partial r_i} U(r_1, r_2, \ldots, r_N), \qquad (2.3)$$

where the potential energy U (r1, r2, rN) depends on coordinates of the N particles.

2.4.1.1 Force fields

A force field (FF) is a numerical expression that illustrates the dependence of a system's energy on the coordinates of its particles. It is made up of a set of parameters and an analytical or tabulated representation of the interatomic potential energy, U (r1, r2, rN). The parameters are often determined either by trying to fit to experimental data or by performing semi-empirical quantum mechanical computations or ab initio computations.

Molecules are believed to be set of atoms held together by bonded interactions and the FF with a simplified model that is appropriate for the region being simulated substitutes the true potential (González, M. A. 2011). The equation number 2.4 shows what a typical expression for FF may look like.

$$U = \sum_{Bonds} \frac{1}{2} k_b (r - r_o)^2 + \sum_{Angles} \frac{1}{2} k_a (\theta - \theta_o)^2 + \sum_{Torsions} \frac{V_n}{2} [1 + cos(n\emptyset - \delta)] + \sum_{Improper} V_{imp} + \sum_{LJ} 4\epsilon_{ij} \left(\frac{\sigma_{ij}^{12}}{r_{ij}^{12}} - \frac{\sigma_{ij}^{6}}{r_{ij}^{6}} \right) + \sum_{elec} \frac{q_i q_j}{r_{ij}} \quad (2.4)$$

The first four terms represent intramolecular or local contributions (bond stretching, angle bending, dihedral and improper torsions) and the last two terms represent the intermolecular electrostatic and van der Waals interactions (in this case via a 12-6 Lennard-Jones potential).

2.4.1.1.1 Intramolecular terms

Bond stretching is frequently depicted using a simple harmonic function that regulates the length of covalent bonds, as given in equation (2.4). In most circumstances, a trigonometric potential is chosen, however a harmonic potential can also be used to express angle bending (2.5).

$$U_{bending} = \frac{1}{2} k_a (cos\theta - cos\theta_o) \quad (2.5)$$

To improve the fitting to vibrational spectra, additional terms are occasionally introduced. The Urey-Bradley potential is the most typical addition (H. C. Urey et.al 1931)

$$U^{UB} = \sum_{Angle} \frac{1}{2} k_{UB(s-s_o)^2,} \quad (2.6)$$

Where s denotes the distance between two external atoms making up the angle.

Any molecule with more than three atoms in a sequence allows the addition of a dihedral or torsion term. Torsional motions are essential to maintaining the proper level of rigidity in the molecule and replicating the significant conformational changes brought on by rotations about bonds. A cosine function, like the one in equation (2.4), is typically employed to express torsional energy.

Where n is the number of minima or maxima between 0 and 2π, Vn specifies the height of the potential barrier, and ϕ is the torsional angle, and δdefines the phase.

To maintain planarity an extra component that represents the constructive contributions of the out-of-plane motions to energy is added, typical expression for which is:

$$U_{imp} = \sum_{impropers} \frac{k_{imp}}{2}[1 + cos(2\omega - \pi)] \quad (2.7)$$

2.4.1.1.2 Intermolecular Interaction

The Van der Waals interaction arise from the repulsive and attractive forces between two atoms. The interaction between the induced dipoles produces an attractive component that changes as r^{-6}, while the repulsion is caused by the overlapping of the electron clouds of the two atoms. These interactions are frequently represented by the 12-6 Lennard-Jones (LJ) potential. For the LJ potential, the associated depth of the well for two atoms interacting i and j is provided by the geometric mean,$\epsilon_{ij} = \left(\epsilon \middle| i \epsilon_j\right)^{1/2}$, whereas the point at which the potential would be zero can be determined by either the geometric mean $\sigma_{ij} = (\sigma_i\sigma_j)^{1/2}$, or the arithmetic mean, $\sigma_{ij} = \frac{1}{2}\left(\sigma \middle| i + \sigma_j\right)$, depending on the FF selected. $\left(\frac{\sigma}{r}\right)^{12}$

$$V(r) = 4 \in \left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right] \quad (2.8)$$

Long Range interactions

We cannot utilise the straightforward truncation technique securely because the electrostatic potential supplied in equation (2.4) does not truncate with distance rapidly enough. The reaction field method, which assumes that all molecules exist in a continuum of dielectric properties beyond a specific cutoff, is a straightforward alternative for that. The charge distribution inside the cutoff sphere polarises the dielectric continuum, producing an extra electrical field further into cavity for a specific atom when all interactions with those particles are explicitly taken into account. The following equation yields this reaction field:

$$E_{rf} = \frac{1}{4\pi \in_o} \frac{2\left(\in_{rf} - 1\right)}{2 \in_{rf} + 1} \frac{1}{R_c^3} \sum_{j \in R} l_j \quad (2.9)$$

Where $\in_o$ representsthe dielectric constant in vacuum, $\in_{rf}$ is the dielectric constant of the system, R_c is the cutoff radius employed in Lennard Jones interactions.

2.4.1.2 Initial conditions

For every particle in the system, we must have its initial positions and velocities known. The positions in the case of a crystal will normally be available as a crystallographic file, and we can create a supercell by combining numerous unit cells. Each particle's velocity is determined at random from a Maxwellian distribution centred on the required temperatures, and then it is modified to zero the angular momentum and centre of mass velocity of the entire system.

2.4.1.3 Evaluation of forces

The calculation of force for a pairwise potential like Lennard-Jones potential is simple. In Lennard-Jones the force exerted by atom j on atom i is given by:

$$f_{ij} = \frac{\partial}{\partial r_{ij}}\left\{4 \in \left[\left(\frac{\sigma}{r_{ij}}\right)^{12} - \left(\frac{\sigma}{r_{ij}}\right)^{6}\right]\right\} = \frac{48\epsilon}{\sigma^2}\left[\left(\frac{\sigma}{r_{ij}}\right)^{14} - \frac{1}{2}\left(\frac{\sigma}{r_{ij}}\right)^{8}\right] r_{ij}, \tag{2.10}$$

$$f_i = \sum_{j \neq i} f_{ij}, \tag{2.11}$$

$$f_{ji} = -f_{ij} \tag{2.12}$$

Although the derivation is more laborious for the intramolecular potentials, one could get analytical expressions for the forces (Allen et.al 1987). Eventually, if the potential is presented as a table of values for U vs. r., the forces must be calculated using numerical differentiation.

2.4.1.4 Periodic boundary conditions

We could simulate our N-particle system in isolation, i.e. a vacuum-sealed environment. But most of the time, we are only concerned with a liquid or solid system's bulk properties, thus we must impose certain boundary conditions. An endless number of copies of the simulation boxes (cells) are present. Only N atoms that are explicitly present inside the main cell are taken into account, but as soon as one of them exits, an image particle comes in from the opposite side to take its place.

2.4.2 Thermodynamic ensemble

During the integration of equation 2.3, the constant number of particles *N*, volume of simulation box *V*, and total energy of system *E* are all kept constant. This indicates that the simulation trajectory will be generated in microcanonical or NVE ensemble However, modest drifts in the total energy may be brought on by numerical integration errors, force fluctuations, and inconsistent forces that are often produced by the cut-off used in truncating interaction potentials. Furthermore, since the total energy is constant but not the individual contributions from the kinetic and potential energy, a system that is out of equilibrium will approach into it as the temperature changes. This can be achieved by using various thermostats and barostats available. Here we will discuss the temperature and the pressure coupling we have used in our system to be simulated.

2.4.2.1 Nosé-Hoover temperature coupling

Nosé (Nosé, S. 1984) first suggested an extended-ensemble approach, which Hoover (Hoover, W. G. 1985) later modified. The friction force is proportional to the product of the velocities of the individual particles and a friction parameter ξ. The difference between the current kinetic energy and the reference temperature is used to calculate the time derivative for this friction parameter (also known as the heat bath variable), which is a completely dynamic quantity with its own momentum (p_ξ) and equation of motion. The global MD scheme's particle equations of motion are replaced in this version by:

$$\frac{d^2 r_i}{dt^2} = \frac{F_i}{m_i} - \frac{p_\xi}{Q}\frac{dr_i}{dt} \tag{2.13}$$

where the heat bath parameter's ξ equation of motion is:

$$\frac{dp_\xi}{dt} = (T - T_0) \tag{2.14}$$

2.4.2.2 Parrinello Rahman Pressure coupling

In theory, the Parrinello-Rahman (Nosé & Klein, 2006; Parrinello & Rahman, 1998) method provides the real NPT ensemble and is close to the Nosé-Hoover temperature coupling. Similar to the Nosé-Hoover coupling, the particle's equations of motion are likewise modified. The following is the modified Hamiltonian, which is conserved:

$$E_{pot} + E_{kin} + \sum_i P_{ii} V + \sum_{i,j} \frac{1}{2} W_{ij} \left(\frac{db_{ij}}{dt}\right)^2 \tag{2.15}$$

The Hamiltonian yields the following equations of motion for atoms:

$$\frac{d^2 r_i}{dt^2} = \frac{F_i}{m_i} - M\frac{dr_i}{dt} \tag{2.16}$$

Although this extra term gives the appearance of friction, it is fictitious and is actually a result of how the Parrinello-Rahman equations of motion are defined, where all particle coordinates are expressed relative to the box vectors.

2.4.3 Integration Algorithm

Equation (2.3) can be integrated numerically. Therefore, in order to advance over small time increments, we must discretize the trajectory and employ an integrator:

$$r_i(t_0) \rightarrow r_i(t_0 + \Delta t) \rightarrow r_i(t_0 + 2\Delta t) \rightarrow \cdots r_i(t_0 + n\Delta t) \tag{2.17}$$

Truncated Taylor-series expansion was used as immediate apparent solution. However, this algorithm is erratic and unreliable. Verlet proposed a better solution. The terms in $\Delta t, \Delta t^3$, etc. cancel out when Taylor expansion for +Δt and -Δt are added up, we get:

$$r_i(t_0 + \Delta t) = -r_i(t_0 + \Delta t) + 2r_i(t_0) + a_i(t_0)\Delta t^2 + O(\Delta t^4) \tag{2.18}$$

An equivalent algorithm to the Verlet integrator is Velocity Verlet producing the same trajectory.

$$r_i(t_0 + \Delta t) = r_i(t_0) + v_i(t_0)\Delta t + \frac{1}{2}a_i(t_0)\Delta t^2, \tag{2.19}$$

$$v_i(t_0 + \Delta t) = v_i(t_0) + \frac{1}{2}[a_i(t_0) + a_i(t_0 + \Delta t)]\Delta t. \tag{2.20}$$

A small-time step would enhance the accuracy of equation (2.3) and its numerical solution, but at the expense of necessitating a high number of time steps to produce a trajectory that is the required length. Alternatively on the other hand, a number for Δt that is too high will result in significant energy fluctuations or drifts, and the simulation may potentially become unstable. Therefore, getting the correct time step requires making a compromise.

2.4.4 Solvent models

Using the explicit solvent model, biomolecular MD simulation is carried out in a water environment that is representative of reality (Nguyen et al. 2014). As available, the solvent models SPC/E, TIP3P, TIP4P, and TIP5P (mainly water models) are utilised in MD simulation (Jorgensen and Tirado-Rives 2005) These water models have one or more of the

physical characteristics of water, such as the radial distribution function, density anomaly, and diffusivity, well optimised. The MD simulation system includes explicit water molecules in an effort to replicate the natural cell-like environment. The steepest descent approach is frequently used to eliminate faulty connections and improve problematic geometries (Kini and Evans 1991)

2.4.5 Energy-Minimization Methods in MD Simulations

For MD structural data, there are various energy minimization strategies. The low energy zones are located by employing a grid search approach and the zero-order method's energy function. The conjugate gradient, often known as the steepest descent method, is the first derivative approach when considering gradient as an energy function. Since the Hessian function is a second derived method, the Newton-Raphson algorithm employs it to find the energy minima (Kini and Evans 1991). Steepest descent and conjugate gradient algorithms are the two basic techniques for energy minimization.

2.4.7 MD Simualtion Tools

When employing the MD simulation method, a number of tools are available to examine the atomic-level alterations in the biomolecules (Khan et al. 2016). Some, like Desmond, have a graphical user interface, while others, like GROMACS and AMBER, operate via command lines. GROMACS (Pronk et al. 2013; Oostenbrink et al. 2004), (AMBER) (Case et al. 2005; Salomon-Ferrer et al. 2013), (CHARMM-GUI), and Nanoscale MD (NAMD) (Phillips et al. 2005),are some well-known and often used tools for MD simulation (Brooks et al. 2009). In order to execute these MD simulations, more powerful hardware and software are required.

2.4.8 Milestones in MD

Simulations using MD are not new. In the latter half of the 1950s, primitive gas simulations using MD were first conducted (Alder and Wainwright, 1957). In the late 1970s, the first MD simulation of a protein was carried out (McCammon et al., 1977), and the foundational work that made these simulations possible was one of the accomplishments honoured by the 2013 Nobel Prize in Chemistry (Levitt and Lifson,1969; Lifson and Warshel, 1968). Experimental structural biology articles have started to use simulations more regularly, where they are used to both interpret experimental findings and direct experimental work. This trend is especially noticeable in the field of neuroscience, where simulations have been employed to study proteins essential for neuronal signalling (Dawe et al., 2016; Delemotte

et al., 2011; Jensen et al., 2012; Shi et al., 2008, Dror et al., 2013;), to aid in the development of drugs that target the nervous system (Manglik et al., 2016; McCorvy et al (Kato et al., 2018; Takemoto et al., 2015), to shed light on the processes by which proteins aggregate that cause neurodegenerative diseases (Khandogin and Brooks, 2007; Wu and Shea, 2013), and to lay the groundwork for the development of better optogenetics tools (Kato et al., 2018; Takemoto et al., 2015).

In the field of AIDS research, there are examples of medications that have been created using molecular modelling and MD simulations (K.R. Jerome et.al 2005, Y. Chong et.al 2002). The most recent treatment is a triplet of two reverse transcriptase inhibitors and a protease inhibitor (such as indinavir) (e.g. Zidovudine, Sustiva). All medications have been created using the rational drug design process, which combines combinatorial chemistry, X-ray crystallography, molecular modelling, and high throughput (virtual) screening. Molecular modelling is a crucial component of the entire drug discovery process nowadays (E. De Clercq et.al 2002).

However, the timescale problem disables classical MD simulation to sample important conformation that is trapped in the local minima stays an issue. The time scale problem needs to be rectified by enhanced sampling methods like metadynamics.

2.5 Metadynamics

The Monte Carlo simulation method and molecular dynamics (MD) have profoundly influenced a wide range of disciplines, including pharmacology, biology, materials science, and astronomy. Nevertheless, despite their popularity, these simulation techniques have drawbacks that limit the range of their applications. The limited time scale that modern computer technology and sampling techniques encounter is a serious limitation. The time scale of the transition to the intended state and the time scale attainable with simulations frequently differ significantly in practise due to free energy barriers. An ingenious method like metadynamics can be used to accelerate the sampling.

Metadynamics is an effective method for enhanced sampling in molecular dynamics simulations and generating the free-energy landscape as a function of a chosen few collective variables (CVs). A history-dependent bias potential, which is adaptively built in the space of the CVs, accelerates the sampling in metadynamics simulations. The algorithm has undergone significant advancements recently, creating a system that is well organised, flexible, and precise and has found many successful applications across a variety of fields. Numerous metadynamics study examples are available (Barducci et al. 2011) to accurately estimate the binding free energy. The graph (Figure 5) showing number of publication including the terms metadynamics on system having protein was plotted by web of science to report the increasing trend of metadynamics in current research world.

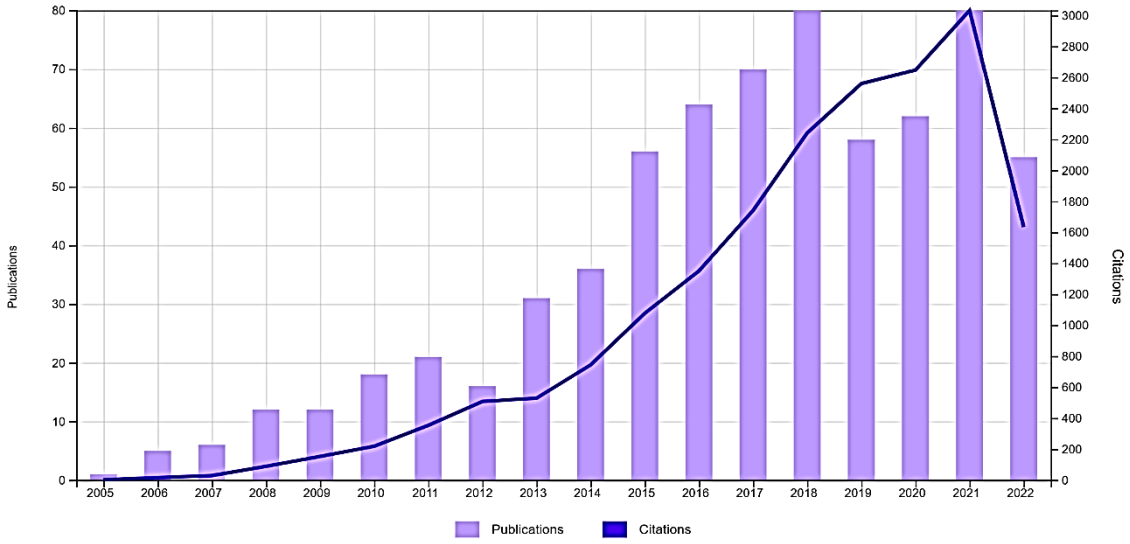


Figure 5: Growth in publications and citations of metadynamics for biomolecular systems with the term protein mentioned. Prepared by analysing via Web of Science. (https://www.webofknowledge.com) in August 2022

2.5.1 Central idea of metadynamics and its algorithm

In metadynamics, the system's Hamiltonian is supplemented by an external history-dependent bias potential that is a function of the CVs. Any explicit x-dependent function, such as an angle, a distance, a coordination number, or the potential energy, can be the CV. The probability distribution describes the equilibrium behaviour of these variables:

$$P(s) = \frac{exp\left(-(1/T)\right)F(s)}{\int dsexp\left(-\left(\frac{1}{T}\right)F(s)\right)} \qquad (2.21)$$

The system is prevented from returning to configurations that have previously been sampled by this potential, which can be expressed as a sum of Gaussians deposited along the system's track in the CVs space. Let S be a collection of d functions of the system's microscopic coordinates R:

$$S(R) = \big(S(R_1)\big) \ldots . \big(S_d(R)\big) \tag{2.22}$$

Metadynamics bias potential at time t, can be written as

$$V_G(S) = \sum_{k\tau<t} W(k\tau) exp \tag{2.23}$$

where W(kτ) is the height of the Gaussian, σ is the width of the Gaussian for the i^{th} CV, and τ is the Gaussian deposition stride. The system is pushed away from any local minimum and into exploring new regions of the phase space as a result of the metadynamics bias potential. Additionally, as a function of the CVs, the bias potential converges to minus the free energy in the long-time limit:

$$V(\vec{s}, t \to \infty) = -F(\vec{s}) + C \tag{2.24}$$

For the duration of a simulation, Gaussian kernels with constant height are inserted according to traditional metadynamics. The system is ultimately forced to explore high-free-energy regions as a result, and the estimate of the free energy derived from the bias potential oscillates around the real value.

The cost of a metadynamics calculation rises with the duration of the simulation because at each step, the values of an increasing number of Gaussian kernels must be evaluated. The bias can be stores on a grid to overcome this problem. The PLUMED input file specifies the required grid spacing (GRID SPACING) or the number of bins (GRID BIN) for each collective variable.

2.5.2 Collective Variables

CVs are stochastic functions of the atomic coordinates that offer a low-dimensional representation of the conformational space because they are typically fewer than the number of atomic coordinates. The key components of the physical behaviour that we are interested in should be able to be described by the collection of CVs. The selection of the CVs has a significant impact on the efficacy of metadynamics as well as many other techniques. In the most typical case, one typically repeats their protocol a few times in a

trial-and-error manner, using the outcomes of earlier simulations to improve older CVs and/or add new CVs to the earlier list. Therefore, active research areas include the creation of CVs that are effective at capturing the characteristics of general classes of reactions (Sega M. et.al 2009, Pietrucci F et.al 2009) or of techniques for improving a suboptimal reaction coordinate(Branduardi D et.al 2007).

2.5.2.1 Choice of CV

Similar to other techniques that rebuild the free energy in a set of generalised coordinates, the selection of the CVs has a significant impact on the accuracy of metadynamics. Experience has shown that selecting the best group of CVs is challenging but not impossible, even in complex instances. If this is done, metadynamics offers extremely rationale transition pathways and, more crucially, has the ability to find new, unforeseen stable and metastable states.

The CVs should ideally meet three criteria:

- Should be able to clearly distinguish between beginning, transitional, and final state.
- Should be able to describe slow events that are important to the process of interest.
- The number of slow events shouldn't be too large to take very long time to fill the free energy surface.

Finding a "good" group of CVs can be exceedingly challenging in many situations. For instance, this occurs when studying the folding of the β-hairpin, a short polypeptide that forms a β-turn with 20 residues. There was no two- or three-variable combination that could adequately characterise all of the slow degrees of freedom in this case.

Types of CV

Here is a brief description of a few variables that have been applied.

Geometry-related CVs, such as distances, angles, and dihedrals produced by atoms or groups of atoms, are the most basic type. In the study of chemical reactions and biophysical systems like protein-ligand recognition, these variables are often utilised.

Coordinates numbers.

It is the most general type of CV that can be defined as

$$S(r) = \sum_{i,j} f(r_{ij}) \tag{2.25}$$

$$f(r_{ij}) = \frac{1 - (r_{ij}/r_o)^n}{1 - (r_{ij}/r_o)^m} \tag{2.26}$$

where two sets of atoms are used to run the sums on i and j, for r>r0, the function f (r) is roughly 1, and for larger r, it goes to zero. The variables n and m can be adjusted to alter the function's smoothness and asymptotic behaviour because, for big r, $f(r) \sim 1/r^{m-n}$. The coordination number may be used, for example, to count the number of bonds that exist between different atomic species or to determine whether a bond exists between two atoms (Iannuzzi, M., Laio et.al 2003, Churakov, S. V., Iannuzzi et.al 2004, Ensing, B., Laio, A et.al 2004, Ensing, B., & Klein, et. Al 2005,).

ANGLE

The ANGLE keyword tells PLUMED to use a CV that is determined by the angles between the centres of mass of three groups of atoms. The LIST keyword is used to group the atoms and operate virtual atom calculation (center of mass). The basic syntax is mentioned in the subsection of methodology.

By using the SIN or COS keywords in the same way as for the TORSION is done, it is also possible to use the cosine or sine of the angle as a collective coordinate.

DISTANCE

The DISTANCE keyword instructs PLUMED to calculate distance between the two groups (center of mass) and apply the bias potential along the CV value. The syntax of the CV is similar as the syntax for angle which is mentioned the methodology section. Group of the atoms are defined using the keyword LIST and then the operation of center of mass (COM) calculation and distance calculation is performed by PLUMED. The term DIFFDIST can be used to represent the difference between two distances at any time. This might be useful in bond formation or bond breaking. The MINDIST keyword directs PLUMED to use the atoms' shortest distance as a CV.

Torsion

The TORSION keyword is used to direct the PLUMED to use dihedral angle as CV. Four atoms or, more generally, the positions of the centres of mass of four groupings of atoms can be used to establish this angle.The LIST keyword is used for properly defining these groups.

2.5.3 The Input files

PLUMED input is stored in a single file with the default name of plumed.dat that specifies the CVs, the kind of run to be executed, and the parameters for the production of the bias potential. PLUMED is activated on the command line by including the flag -plumed following the name of .dat input file. According to the functionality needed, the commands

in this file can be separated into two parts. The kind of simulation (metadynamics, umbrella and steering sampling, replica exchange methods), as well as the parameters required for the selected methodology, are defined in the first part of commands. The second part describes the degrees of freedom often addressed as collective variable on which the algorithms operate. Each line in the file describes some action that is to be performed by PLUMED. This could be a collective variable computation, sporadic trajectory analysis, or biassing of the dynamics. The first word in the line describes the particular action to be carried out then the following number of keywords provide more details on how the action should be performed. Clearer details of the plumed input file for our system is provided in the method section.

A .tpr file is also required for PLUMED to run a metadynamics simulation. This file includes the simulation's starting structure, all simulation parameters, and molecular topology.

3- MATERIALS AND METHODS

The section gives a detail note on the procedure that is followed to accomplish the study. The section can be divided in three major techniques that were used throughout to accomplish three different objectives. First was the system preparation with Molecular Docking studies. Second was protein's classical MD simulation studies to find out parameters and reference values like average RMSD for protein to compare with the Avg RMSD of protein-ligand systems after metadynamics run, and lastly Metadynamics studies to calculate Binding free energy of the protein-ligand system.

3.1 Formation of Protein-Ligand system through molecular docking studies

The target M^{pro} protein structures with PDB id 6YB7 and 6LU7 were retrieved from RCSB PDB. The 6YB7 structure had a resolution of 1.25 Å and was determined by X-RAY diffraction. (Owen, C. D., et al. 2020). The unligated active site structure of M^{pro} (6YB7) was selected to expose and test the other regions of M^{pro} where the ligand molecule can bind. The system formed with ligand present at the non-Active site region is referred as non-Active pocket system for the ease.

For the purpose of generating protein-ligand system in which the ligand is present at the active site, M^{pro} structure with PDB Id 6LU7 (Jin, Z., Du et.al 2020) which was found to be in complex with inhibitor N3 was taken from RCSB PDB (Berman, H. M., Westbrook et.al 200). The structure has resolution of 2.16Å and was determined by X-RAY diffraction. N3 in reference structure (6LY7) was bounded at the so called active of the main protease. The active site residues to decide the active site pocket was then extracted by using script option of the Discovery studio. The same procedure was adopted to determine the pockets of the non-Active pocket system after performing docking. The system where ligand was present is the active-site is referred as Active pocket system.

After the protein was retrieved, residue binding sites were determined and heteroatoms were deleted, the biologically functional unit of the protein was taken to dock with the potential ligand of choice with the help of PyRx software. The vina Wizard which runs through Autodock Vina was used for the purpose of making protein-ligand systems where ligand is docked at different locations according to docking score.

3.1.1 non-Active Pocket System

The docking was performed in three steps to get the conformations of the ligand molecule as well as the best stable complexes or protein region where ligand is fused at different positions in the protein for further molecular dynamics studies.

A- Blind docking studies were performed on M^{pro} so as to be able to access and explore all the geometrically compatible regions of the protein. The best non-active site region on the basis of docking score of the ligand was selected.

B- The ligand binding site atoms forming the novel pocket were determined from discovery studio.

C- The site-specific docking was performed on the region of the top conformation we got from previous blind docking results. This gave the best ligand conformation in the non-active site region. Docking results were used to build the protein-ligand system in which the ligand was docked at the non-Active site region in the protein.

3.1.2 Active Pocket System

The site-specific docking was performed on the active site region of the protein to obtain the best ligand conformation in the active site region. The results of docking studies were used to generate the protein-ligand system in which the ligand is docked in the active site region of the protein.

After the Protein-ligand systems were generated, subsequent molecular dynamics and metadynamics were performed to achieve the objectives of the study.

3.2 Molecular dynamics (MD) simulation of M^{pro}

Classical MD simulation was carried out for the protein in water to test the stability of the biologically functional unit after the MD simulation of 10 ns. This run was carried out to estimate the values of different parameters used in the molecular dynamics parameter (.mdp) later in metadynamics studies. To estimate the time scaling required to achieve the objectives of each MD run like minimization, equilibration, etc. CHARMM GUI server and GROMACS were used to acquire this part of the study. The Water box formation, solvation, ionization PBC set up were accomplished by CHRAMM-GUI to generate input files for GROMACS. The Energy Minimization, equilibration, production run, and analysis were performed using GROMACS.

3.2.1 Simulation Preparation

The crystal structure with unligated active site (6YB7) was used as starting point for this simulation. A web-based graphical user interface called CHARMM GUI was used to prepare the complex biomolecular systems ready for molecular simulations. CHARMM 36 force fields were used and the preparation was carried out in steps including generating topology of the system, box formation, solvation, and ionization. CHARMM force field equation can be given by the following equation:

$$V = \sum_{bonds} k_b(b - b_o)^2 + \sum_{angles} k_\theta(\theta - \theta_o)^2 + \sum_{dihedrals} k_\phi[1 + cos\,(n\phi - \delta)]$$
$$+ \sum_{impropers} k_\omega(\omega - \omega_o)^2 + \sum_{Urey-Bradley} k_u(u - u_o)^2$$
$$+ \sum_{non-bonded} \epsilon\left[\left(\frac{R_{min_{ij}}}{r_{ij}}\right)^{12} - \left(\frac{R_{min_{ij}}}{r_{ij}}\right)^{6}\right] + \frac{q_i q_j}{\epsilon r_{ij}}$$

First term in the equation describes the energy corresponding to bond stretches, k_b being the bond force constant, $(b\text{-}b_o)$ is the distance it atom has moved from equilibrium, second is bond angle equilibrium $k_\theta$is angle force constant and angle from equilibrium is represented by $(\theta - \theta_o)$, fourth term is for improper where k_ω is force constant and $\omega - \omega_o$ is bending out of plane. Fifth term represents Urey-Bradley component with k_u representing force constant and u is the distance in harmonic potential between 1, 3 atoms.The last two terms represnsts non bonded ingteractions between pairs of atoms i,j. With a typical 12-6 Lennard-Jones potential and a Coulombic potential, the van der Waals (VDW) energy and electrostatic energy are computed.

CHARMM GUI

A web-based graphical user interface called CHARMM-GUI, found at http://www.charmm-gui.org, was created to make it easier and more uniform to utilize both basic and sophisticated simulation techniques in CHARMM. With its many facilities for analyzing and manipulating atomic coordinates and dynamics trajectories, CHARMM is a popular academic research program for macromolecular mechanics and dynamics. A molecular model system can be built and validated in an interactive manner on the web, and if a fault is discovered through visual inspection, one can return to the previous setup and regenerate the entire system from scratch. Input files are produced by CHARMM-GUI for a variety of applications, including CHARMM, GROMACS, AMBER, NAMD,

GENESIS, LAMMPS, OpenMM, Desmond, and CHARMM/OpenMM. Since its creation in 2006, CHARMM-GUI has been widely used for a variety of tasks and now includes a number of unique modules made to set up a wide variety of simulations (Jo, S., Kim et.al 2008). We used the solution builder module of the input generator for solvation, ionization as well as generation of inputs for the GROMACS for our system with CHARMM GUI. The CHARMM-GUI generated the initial structure, topology, force fields, as well as molecular dynamics parameter files with some basic values that were modified later according to our system needs. By offering input files for all popular simulation programs that adhere to widely known standards and well-validated methodologies, CHARMM-GUI also increases the reproducibility and quality of biomolecular simulations.

3.2.1.1 Box Formation, solvation, and ionization

CHARMM GUI generated the topology file for the systems, the rectangle-shaped water box of edge distance 10 was formed and solvated with water using explicit solvation, TIP3 water model.

Neutralizing the system was carried out for obtaining accurate electrostatic values during the simulation. Under periodic boundary condition (PBC) the system was neutralize with Monte Carlo algorithm adding 83 POT and 79 CLA ions were placed by CHARMMM GUI in a concentration 0.15 to neutralize the system. The box crystal type was cubic with dimensions along x,y,z axis are 98X98X98 and with crystal angle 90. Charmm36m force field option was selected for generation of input files for GROMACS.

3.2.2 GROMACS

One of the most popular open-source and free chemistry software programs is GROMACS (Abraham, M. J et.al 2015), which is mostly used for dynamical simulations of biomolecules. It is a freely available tool, and a brief tutorial of this tool can be accessed by this link (http:// www.mdtutorials.com/gmx/) (Pronk et al. 2013).

It offers a wide range of computation styles, preparation tools, and analysis software. Energy minimization, and simulated annealing calculations, normal-mode analysis, are all supported, as are simulations with leap-frog Verlet, velocity Verlet, Brownian, and stochastic dynamics. There are numerous methods for controlling pressure and/or temperature. The 15 variations of AMBER, CHARMM, GROMOS, and OPLS are validated and incorporated, and all frequently used molecular mechanics force fields are usable. With nearly two million lines of code, GROMACS has developed into a very huge software project. We direct the interested reader to the earlier articles published for a

thorough discussion of the historical development and several algorithms incorporated in the engine (Berendsen et al., 1995a; Hess et al., 2008a, 2008b; Lindahl et al., 2001; van der Spoel et al., 2005; Abraham et al., 2015; Berendsen et al., 1995b; Páll et al., 2015; Pronk et al., 2013).

3.2.2.1 Energy minimization

Before beginning the dynamics of the electroneutral system generated all the steric clashes and inappropriate geometry in the system was resolved with the energy minimization step. It was run for 6000 steps, keeping emtotal =100 and steep descent integrator was used for the process.

3.2.2.2 Equilibration

Through an assessment of equilibration technique's effects on the stability of Mpro, we contributed to the efforts being made to increase the reliability of MD simulations. Equilibration was performed to bring system to the desired temperature and pressure we want to simulate our system. Parrinello–Rahman barostat and the Nose–Hoover thermostat were used in pressure and temperature coupling were used to bring system to 303.15K and 1 mPa. Equilibration step was accomplished in two steps, constant number of particles volume and temperature (NVT) and constant number of particles pressure and temperature (NPT). Temperature, pressure and density profile were tracked during the respective time scales of the processes.

3.2.2.3 Production run

The system was well-equilibrated at the appropriate temperature and pressure after the two equilibration phases were finished. Position restrictions were released, and production MD was run to collect data. Post the completion of the production run, analysis was performed with RMSD, RMSF, and radius of gyration.

3.3 Free energy calculation using Metadynamics

Basic steps followed for the generation of free energy is represented with the help of a schematic (Figure 6).

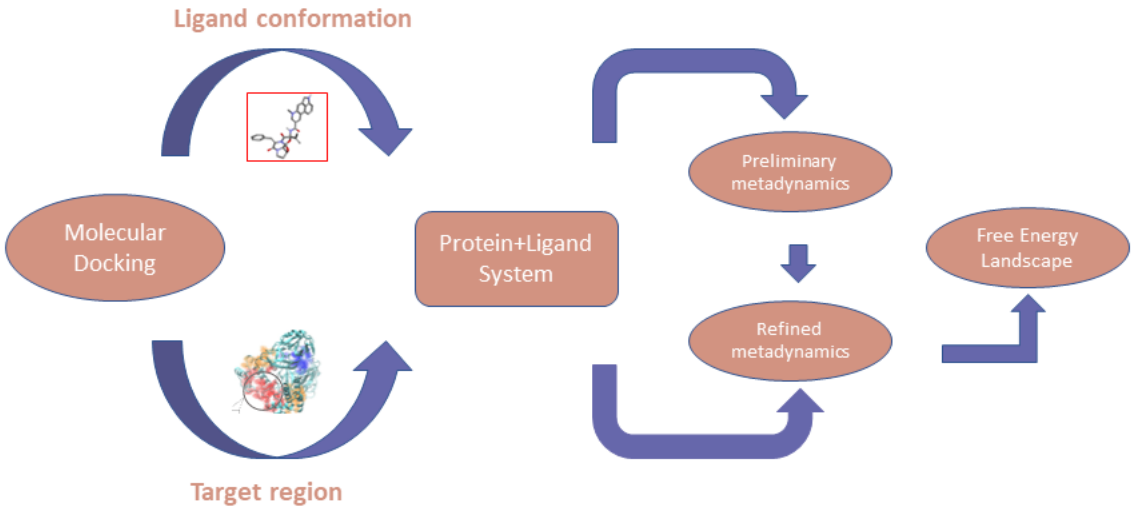


Figure 6:A schematic of approach followed for FES estimation

For both the systems, two biased simulations were performed using plumed library patched with GROMACS software suit.

1. The Unrestrained metadynamics run
2. The Restrained metadynamics run

Simulation of 450 ns was performed for restrained metadynamics for the Active pocket system and 300ns metadynamics run was performed for non-Active system while the unrestrained simulation ran for less than 300ns in non-Active pocket system while for the Active Pocket system it ran for 150ns only to save the computational costs. The results of unrestrained metadynamics run were analyzed to modify the parameters for the final restrained metadynamics run the details of which are discussed below. The procedure mentioned is consistently followed for both the Active and the non-Active pocket systems to have a fair comparison.

3.3.1 PLUMED

PLUMED is a free, open-source, community-developed library that offers numerous techniques, such as:

- Free energy methods
- enhanced sampling techniques
- tools for analyzing the massive volumes of data generated by molecular dynamics (MD) simulations

These methods can be combined with a wide range of collective variables to describe intricate processes in biology, chemistry, physics, and material science. Calculating CVs from atomic positions is the most fundamental task performed by executable and C++ library provided by PLUMED. The advantage of having this feature both in a standalone application and in a library is that it enables one to compute CVs during an MD simulation. The ability of PLUMED to apply additional forces to the CVs is a second crucial feature since it enables users to use free-energy techniques like metadynamics, umbrella sampling, and steered MD. PLUMED is also able to analyze trajectory data in different ways. This can be achieved during post processing of the trajectory file, like promptly during simulation or like a conventional analysis tool.

3.3.1.1 PLUMED input file

A single input file is used by PLUMED, and each line contains instructions for PLUMED perform something. A simple example of input file can be: -

c1: COM ATOMS= 1-10

c2: COM ATOMS= 10-20

c3: COM ATOMS= 20-30

a1: ANGLE ATOMS = c1, c2, c3

metad: METAD ARG=a1 PACE=100 HEIGHT=0.5 SIGMA=0.5

l: LOWER_WALL ARG=a1 AT=2.792 KAPPA=10000

PRINT ARG=a1 FILE=COLVAR STRIDE=500

Each line in the aforementioned input file gives PLUMED instructions to generate a new object called as an "Action." Every conceivable action the user instructs PLUMED to perform, including the calculation of a simulation bias potential (e.g., METAD or UPPER WALL), a center of mass (COM), a CV (DISTANCE, ANGLE, or TORSION), the writing out of some data (PRINT), can be cast as an Action object. The majority part, these Actions use some input—typically the positions of certain atoms in the system (for the first center of mass in the example above, the input is the positions of atoms 1-10) and use the instantaneous value of a CV to determine a new CV or bias potential.

After the straight forward patching and including PLUMED in the MD engine is done the following command was ran to process a plumed file in GROMACS to give outputs.

$$gmx\ mdrun - s\ topol.tpr - nsteps\ 5000000 - plumed\ plumed.dat$$

A brief about the .tpr file is discussed below.

3.3.1.2 Preparation of .tpr file

PLUMED needs a .tpr file which is a binary file containing all the necessary information about the system like topology, coordinates, force fields and other parameters. It executes the command after fetching the information about the system from .tpr file. For the generation of .tpr file for metadynamics run following steps were performed. These are same as we have discussed in the section 2 of methodology.

- The system was minimized
    - The required files for the energy minimization simulation were prepared by CHARMM-GUI and the energy minimization was performed in GROMACS
- Equilibrated by performing NVT and NPT simulations
    - The systems were equilibrated at temperature 303. 15K and pressure 1 MPa
- The .gro file of the last frame of the trajectory file generated at the end of NPT simulation and the parameter file were used as inputs to generate a .tpr file. We used the below mentioned command to generate the .tpr file for both our Active and non-Active pocket systems.

$$gmx\ grompp - f\ metaD.mdp - c\ npt.gro - t\ npt.cpt - p\ topol.top \\ - o\ md.tpr\ - n\ index.ndx$$

3.3.2 Unrestrained metadynamics run

Unrestrained metadynamics was performed for less than 300 ns. The output files at the end were analyzed to calculate prerequisite CV that were used for the final restrained metadynamics run. Prerequisite CV are the inputs that are available to the PLUMED without using trajectory to calculate them or as an output from the Actions in the input file before it.

3.3.2.1 The PLUMED input for unrestrained run

The input file can be divided in three sections. First section has commands telling PLUMED to take ACTION to calculate collective variables that were the Center of mass of pocket, the COM of ligand and the distance between them. The second section commands PLUMED to apply restrains on the CV calculation by introducing upper and lower boundaries at certain CV values. The third section directs PLUMED to bias along distance as CV and print the values of the CV in COLVAR file with the given stride and

the information of the gaussian kernels deposited were given as output file called HILLS. The two metadynamics runs were performed in order to refine the plumed input file.

The grid size was chosen such that the GRID_MIN and GRID_MAX values where 0 Å and 20 Å respectively. The value for GRID_MIN was taken to be 0 as the distance between center of mass of the pocket and the ligand cannot go negative. The GRID_MAX was taken 20Å because the interest was to let the ligand move enough away from the pocket in the solvent so that the direction and the path of unbinding of the ligand from its pocket can conveniently be visualized later during analysis. Number of bins were taken to be 200.

The restraining potential was applied by introducing upper wall at 19 and lower wall at 1 CV value. When the CV value exceeds this range system experience a force constant of value1000 as a result of which it is pushed back to the accessible range of phase space.

The metadynamics run was executed for less than 300ns. The resulting COLVAR file was plotted by using gnuplot. The Gaussian kernels that were deposited throughout the simulation and saved in the HILLS file were added together using sum hills by the following command.

$$plumed\ sum\_hills\ --hills\ HILLS$$

The aforementioned command creates a file named fes.dat in which a regular grid is used to calculate the free-energy surface as a function of distance CV.

3.3.3 Restrained metadynamics run

On the basis of the results of the previous metadynamics runs, the necessary modifications were made in the input file to refine the results by adding angle restraining.

3.3.3.1 PLUMED input for restrained metadynamics run

The angle restraining was added on the three virtual points. Details on these virtual points are discussed in this section. The rest values and ACTIONS were kept same as was in the input file of unrestrained metadynamics run.

Determination of virtual points for angle calculation

Visualization of the .xtc file that was generated at the end of the unrestrained run was done using VMD (Humphrey et.al 1996) software. Upon visualizing the trajectory, three molecules lying nearly colinear on the path and direction of binding and unbinding of the ligand from the pocket into the solvent were decided. One of the three selected molecule was ligand itself the other two molecule were from the protein with index numbers 6626

(Ala 129) and 5965 (Val 86) in the non-Active pocket system while in the Active pocket system the indexes of these two molecules were 8824 (Gln 8824) and 7901(Asp 216). Using Tk console, a console script was written to calculate center of mass of the neighboring atoms within 7Å radius of these three nearly colinear lying molecules. The center of mass coordinates was also determined to calculate angle between the three points manually. This confirms that the angle between these three points should be in the range 160^o – 180^o in the bound and the unbound frame of the output trajectory file of unrestrained metadynamics run. Following example of TK console can deliver some idea.

$$set\ sel[atomselect\ top\ name\ CA\ and\ within\ 7 of\ index\ 6626]$$
$$measure\ center\ \$sel\ weight\ mass$$

- The first command selected a group of α carbon (CA) atoms within 7 Å of radius of the index number 6626 by using keyword atomselect and store the value in variable *sel.*
- Second command performed action of computing center of mass of the values store in variable *sel.*

The residues found within the 7 Å radius of the index number 6626 were found to be GLY 109, Thr 111, Phe 112, Gln 127, Cys 128, Ala 129, Met 130, Arg 131, Ile 136, Lys 137, Glys 138, Glu 290 .

Once the indexes of the molecules within 7 Å of range was determined in both the systems, their indexes of α carbon were provided in the plumed input file to calculate the center of mass and then angle between these virtual points were calculated by PLUMED. Below is the modified part of the unrestrained input file.

```
#COM calculation
C1: COM ATOMS=9443-9517
c2: COM ATOMS=7836,7846,7881,7895,7902,7914,7938
c3: COM ATOMS=8166,8180,8199,8215,8225,8242,8264,8285,8299,8320,8337
#angle calculation
a1: ANGLE ATOMS=c1, c2,c3
lwall: LOWER_WALLS ARG=a1 AT=2.8 KAPPA=100000 EXP=2 EPS=1 OFFSET=0
```

C1 has the indexes of ligand molecule while c2 and c3 have the indexes of the α-carbons of the atoms that are neighbors with the nearly colinear lying molecules.

The angle restraining was then applied by the introduction of the lower wall at 160^{o} so that the sampling within the given range is only allowed.

On completion of the simulation, the COLVAR file was plotted in a graph and HILLS file was used to generate Free Energy Landscape. To examine the convergence of metadynamics run, the estimate of Free energy as a function of time was determined and plotted. The accuracy of the FES reconstruction depends on the selection of the metadynamics parameters, namely the height and width of the Gaussians, the number of molecular dynamics steps between each metadynamics step (T), and the overall length of the metadynamics trajectory. The untempered metadynamics was performed which implies a constant height of the Gaussian equal to be 0.01 kcal/mol. The width of the gaussian was selected to be 0.5 and the gaussians were made to be deposited at the pace of 500 and the data of the hills deposited was collected in the HILLS file. The collective variable was printed at the stride of 100.

3.3.3.2 Generation of Free energy surface

The CV values near the upper and lower boundary were found to be not sampled upon investigating the COLVAR file with the help of histogram. This happened as a result of force acting on the boundaries to push the system back into the range to enable the sampling only in the desired range of CV as well as in the interested region. As a result, those small CV values were trimmed off and were excluded during FES generation and further calculation of dissociation constant.

4 RESULTS AND DISCUSSION

In this section, results of docking studies are discussed along with a snapshot of ligand conformation and the two protein targets (active and non-active) along with Vina score of the two systems. Subsequently, classical MD simulation results are discussed. Refined free energies of interaction of the drug at active and non-active sites are presented towards the end of this section.

4.1 Molecular docking

From the molecular docking studies, we accomplish two objectives for the non-active site and the active site docking of the ligand molecule with the target protein:

A. The best conformation of the ligand was generated for both the systems (Figure 7)
B. 3 top non-active site regions were detected on the basis of best docking score for non-active site region (Figure 8)

4.1.1 Conformation of the ligand and binding region

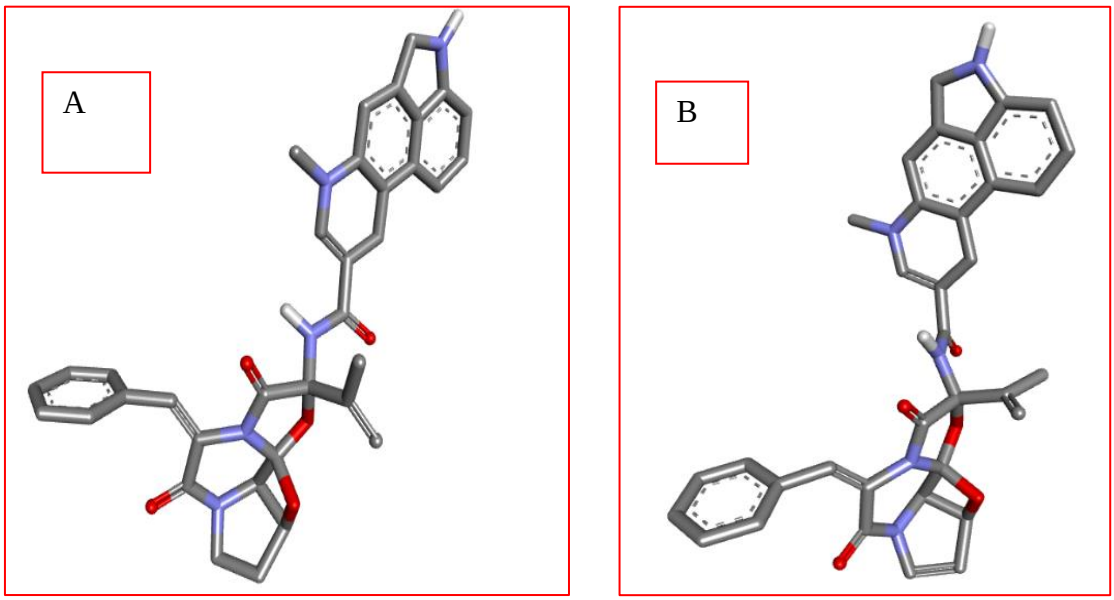


Figure 7: The best-chosen conformation of the ligand on the basis of the vina score, to perform the further study; A: shows ligand docked at non-Active site (shown as red region; B: Ligand conformation in the Active site region

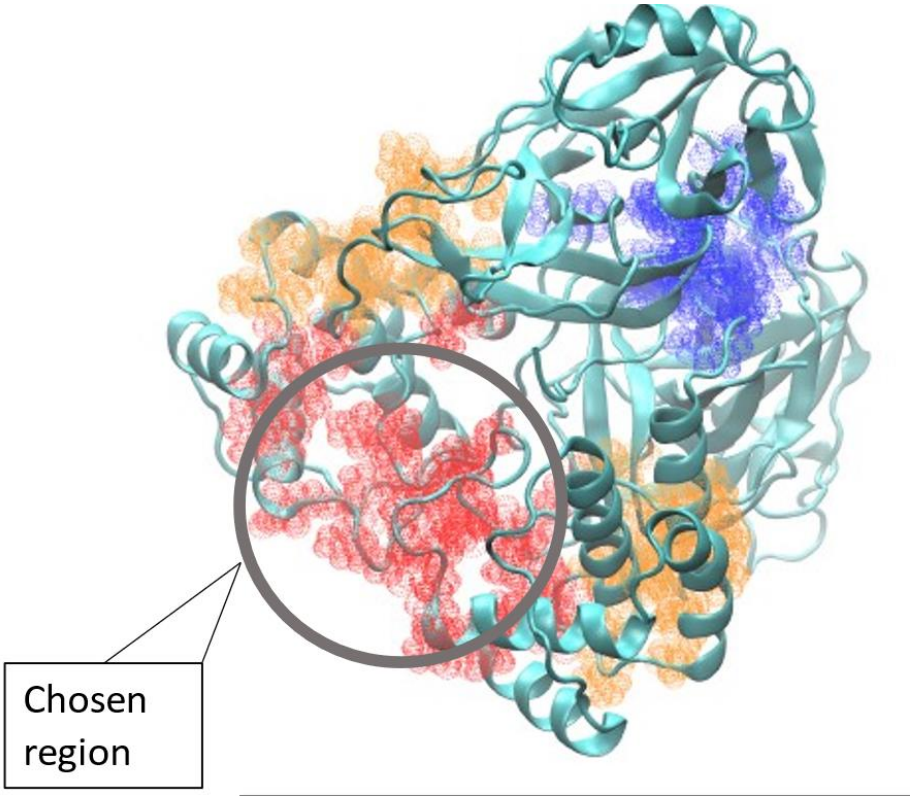


Figure 8: The three best chosen BE regions in the protein based on the docking score. Red region was found to show the best ligand conformation blue region was found to be the second-best conformation and the yellow one was third in vina score rank

4.1.2 Investigation of the pocket region

The investigation of the substrate binding residue for non-Active site region was performed with discovery studio script and we discovered the novel pocket comprising of the following residues.

Phe 3, Arg 4, Lys137, Thr 199, Tyr 237, Tyr 239, Leu 272, Asn 274, Gly 275, Met 276 Asn 277, Gly 278, Thr 280, Gly 283, Ala 285, Leu 286, Leu 287.

The pocket had few residues belonging to chain A, few found to belong to chain B while 3 of the above residues were present in both the chains indicating that the non-Active pocket system is interacting better for the nonce.

As the ligand was found interacting in the protein-protein interface, it increases the hope of being molecular target or as the protein-protein interface is emerging for the development of therapeutic drug (Bai et al., 2016).

Upon visualizing the substrate binding pocket for Active pocket system,the following residues were found. All the residues belong to either of the chains in the Active pocket system.

Ala 2, Val 3, Leu 4, Thr 24, Thr 25, Leu 27, His 41, Met 49, Phe 140, Leu 141, Asn 142, Ser 144, Gly 143, Cys 145, His 163, His 164, Met 165, Glu 166, Pro 168, Asp 187, His 172, Gln 189, Thr 190, Ala 191, Gln 192

The visuals extracted from discovery studio showing few ligand binding site atoms for the non-Active and Active and the intermolecular interaction it shown in F igure 9.

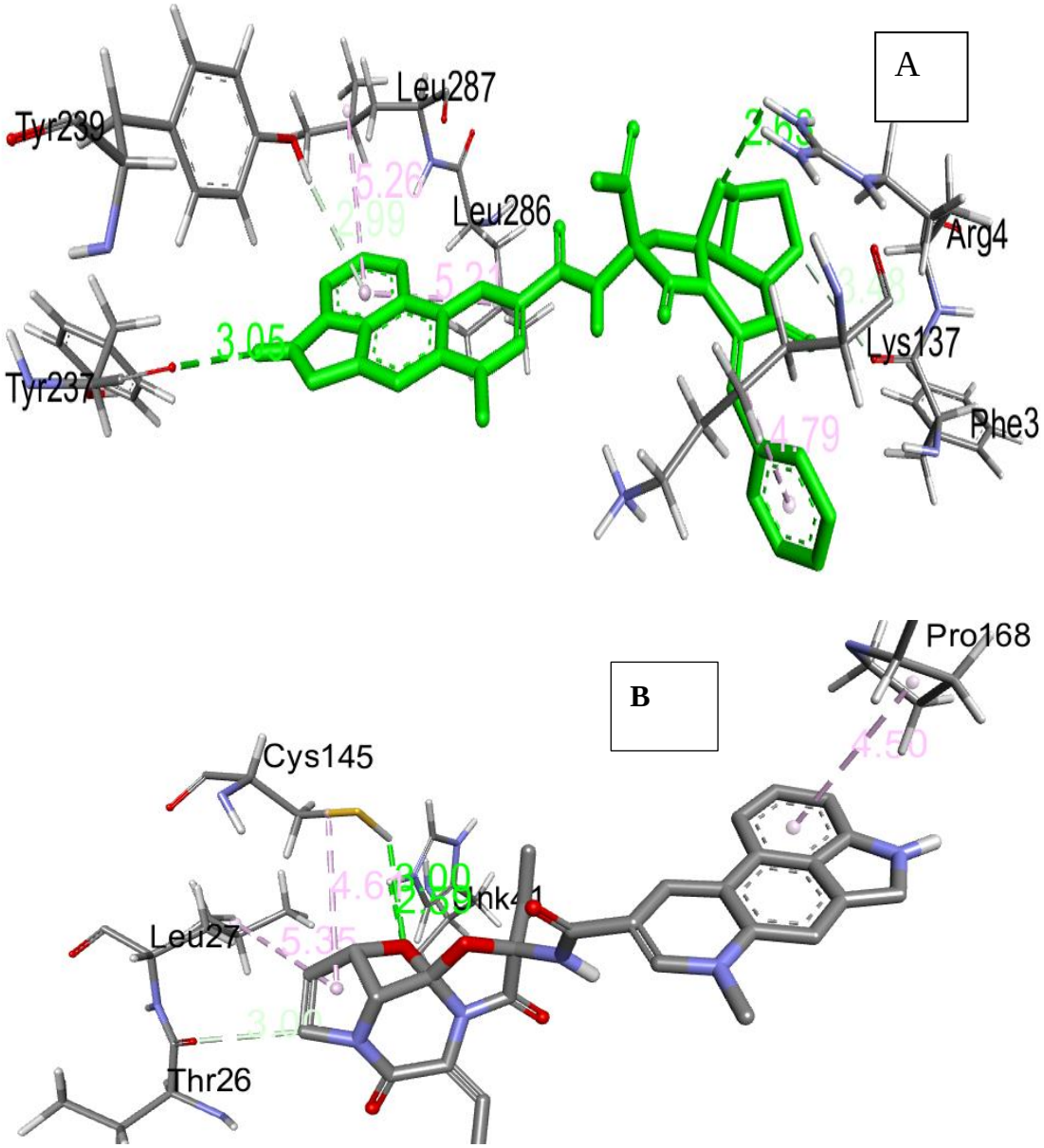


Figure 9: Visuals of few ligand binding site atoms that are involved in intermolecular bonding. Mentioned number are the distances between them in Å. Bonds in green color are the hydrogen bonds and the bonds in pink color represents hydrophobic bonds. **A**: Ligand conformation at Non-Active pocket system; **B** Ligand conformation at Active pocket system

The results showed that 7 non-Active site residues were found to interact with ligand while in the Active pocket system only 5 residues were found interacting. This inferred the good enough interaction between the two species. At the end of docking, we conclude that blind docking increased the scope of finding potential regions in the M^{pro} that maybe left out due to performing site specific docking at the known active site. The results gave us the insights to look at the known dyad.

The resulting geometrically complimentary region with significant binding energy of -11.7 kcal/mol as a vina score was the output of the docking run for the non-Active pocket system. While for Active pocket system it was found to be -9.5kcal/mol. The value of this energy was found to vary when free energy was calculated by energy landscape.

Table1: Vina scores from docking

Site	**Vina score**
non-Active	-11.7 kcal/mol
Active	-9.5 kcal/mol

4.2 MD Simulation

The section discusses system's details and analysis after completion of the simulation. The classical MD simulation of the biologically functional unit proves its stability upon analysis with RMSD, RMSF, radius of gyration. The section first discusses CHARMM-GUI outputs and then GROMACS outputs are reported.

4.2.1 CHARMM-GUI

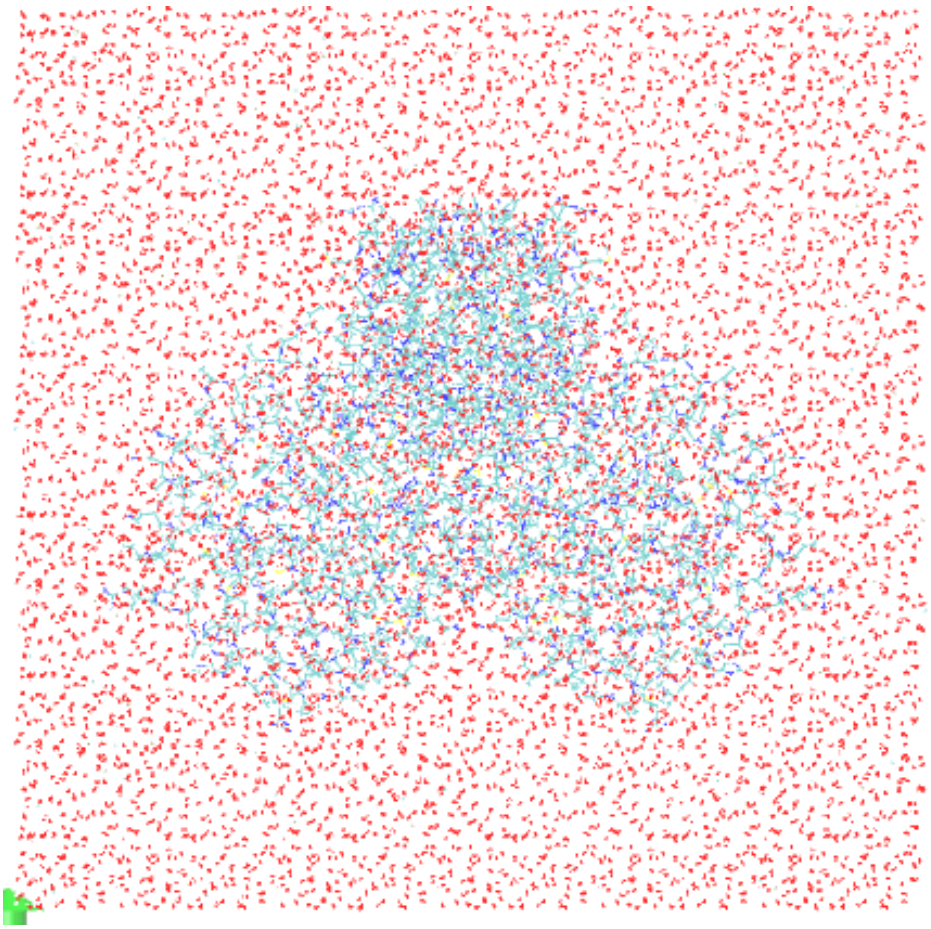
From CHARMM-GUI the required input files for GROMACS to perform MD simulation runs like energy minimization, equilibration and production run were generated with flying colors for the following three systems.

- Biologically functional unit of M^{pro} without ligand attached to it
- Biologically functional unit of M^{pro} with ligand attached to the non-active site
- Biologically functional unit of M^{pro} with ligand attached to the active site

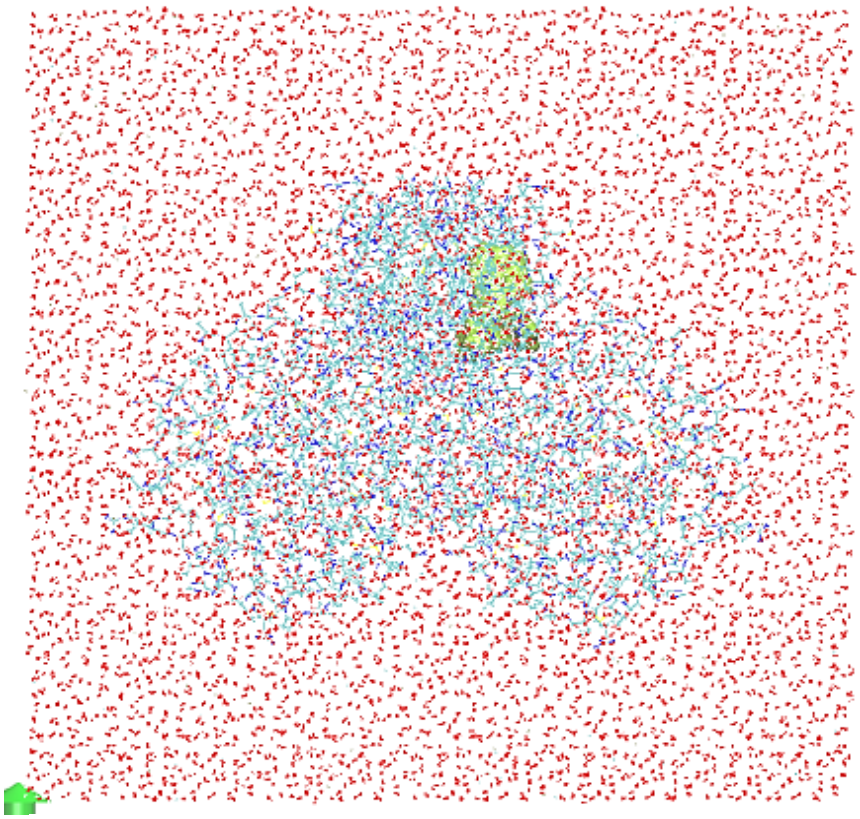
Generating topology and force fields for chemical molecules like ligands is often tedious but CHARMM-GUI modeled the biomolecular systems and let us acquire useful insights into the system and its dynamics. Fig 10 shows the three systems modelled by CHARMM-GUI. Red color dots represent water and ion molecules. The ligand molecule in yellow

color is shown in quick surf representation, the protein and DMSO molecules are in lines representation. VMD software was used to visualize the structure files generated by CHRAMM-GUI. The systems were found to be well solvated and ionized as it does not cause any instability in the later molecular dynamics and metadynamics simulation runs.

A



B



C

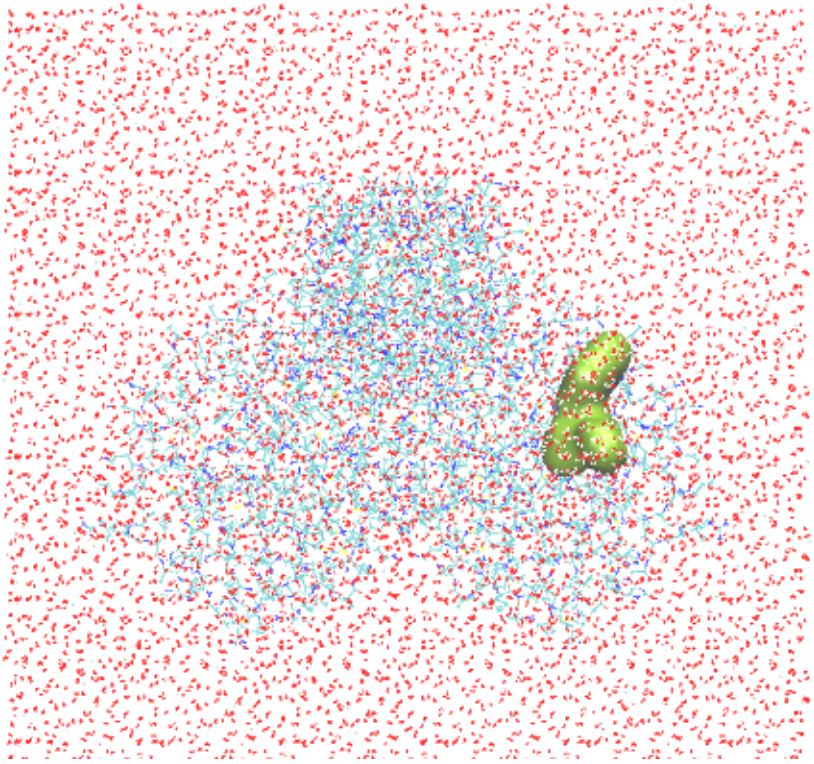


Figure 10: Snapshot by VMD of the generated biomolecular systems modelled by CHARMM-GUI
A: The biologically functional unit of M^{pro} without ligand docked at any site;
B: The biomolecular system comprising the target protein M^{pro} and ligand molecule docked at the chosen (red region) non-active site region;
C: The biomolecular system comprising the target protein M^{pro} and ligand molecule docked at the active site region.

4.2.2 GROMACS

The system energy, temperature and pressure were tracked to make sure the system has simulated successfully. The data files to plot graphs were generated using the files that were the output of MD runs. The energy, temperature, pressure, and density of the system were closely observed and were found to fluctuate around their respective values as mentioned in the .mdp file.

When the system was kept in the thermal bath it experienced sudden drop in the temperature and then a sudden rise this happened as a result of sudden but not gradual increase of the temperature. This event happened within 5 ps of time scale. Soon after the system reached its equilibrium value. This phenomenon was visible in the NVT graph (**Figure 11 B**). Simulated annealing is one solution to this phenomenon. Later the system was checked to make sure there were no artifacts caused in the system because of this jump during NVT run.

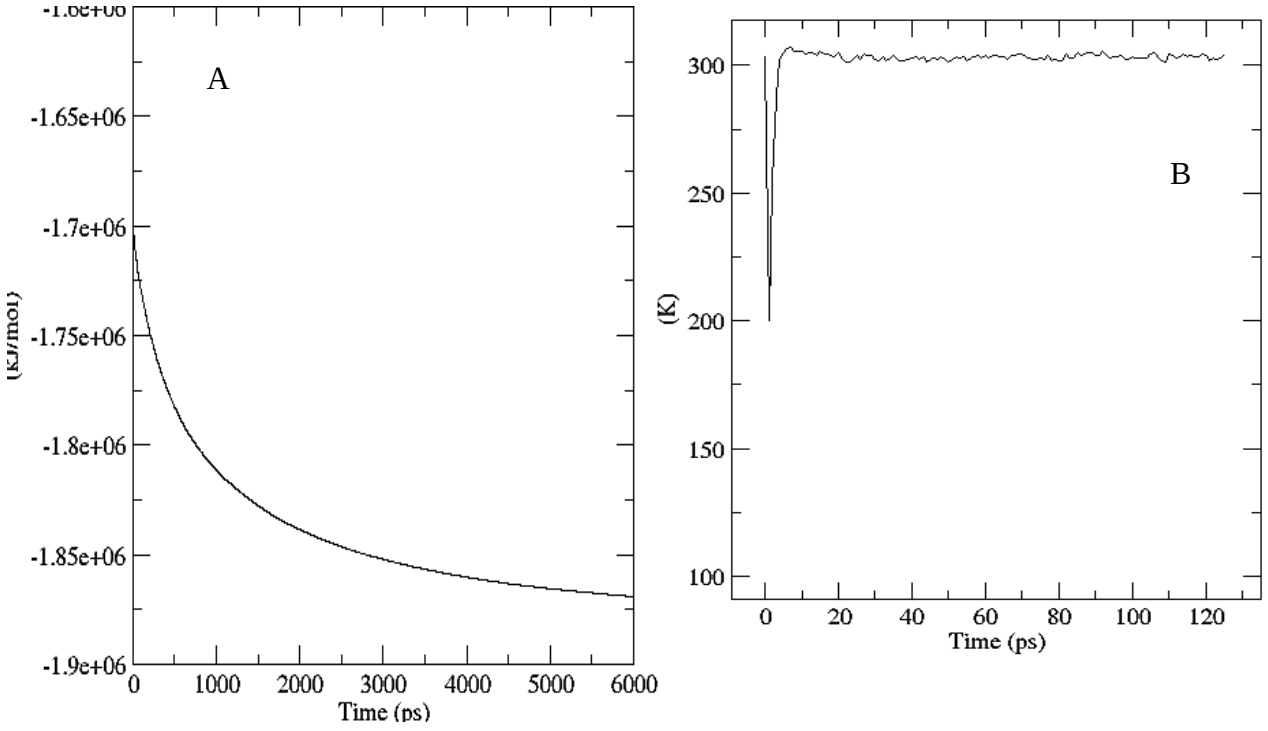


Figure 11: **A**: Energy minimization; Potential energy profile evidencing the success of energy minimization step. After 4000 ps the energy was minimized and system was found to be in its local minima.
B: NVT Equilibrium; Temperature profile of the biologically functional unit during NVT equilibration run. The temperature was found to equilibrate between the given value in the input file.

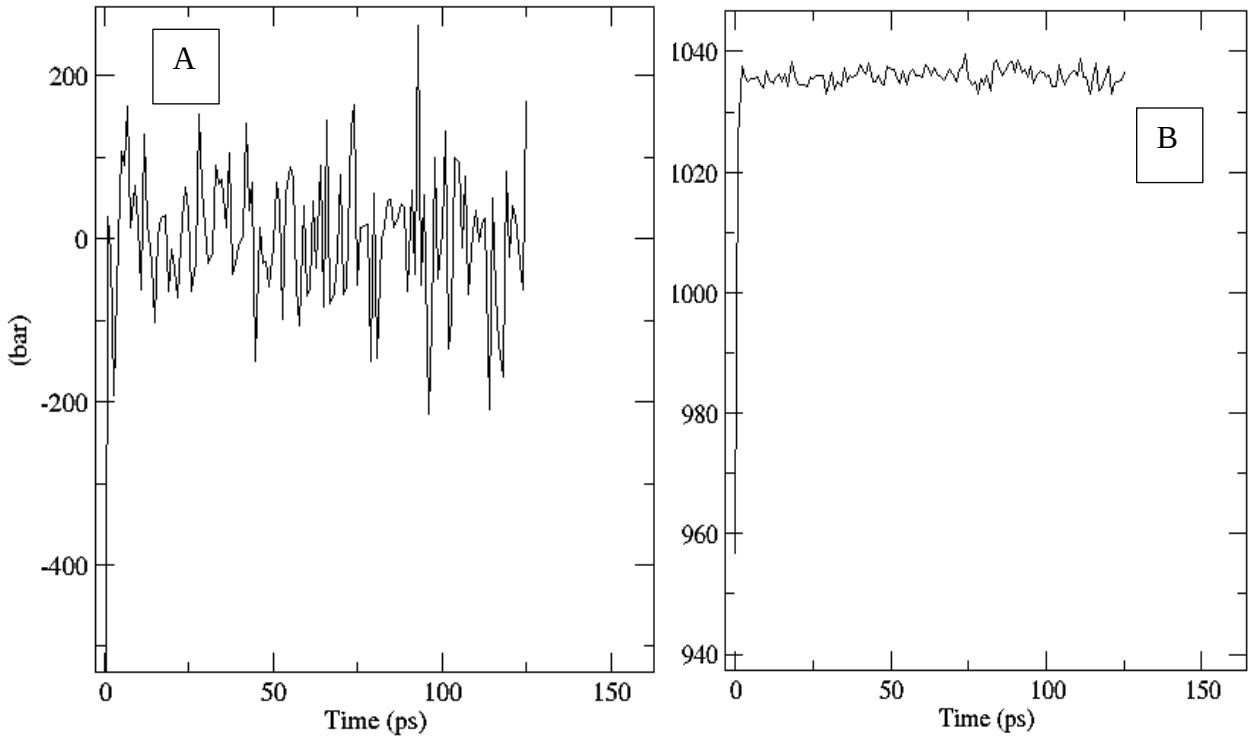


Figure 12: **A** shows the pressure profile after the NPT simulation run was over it was found to fluctuate around the given value of 1MPa.
B shows the density profile after the NPT run equilibration run has completed.

The analysis of the results of MD simulation was performed by measuring similarity between the simulated protein and its reference structure as well as structural dynamics behavior with time were estimated by quantities such as root mean square deviation (RMSD), root mean square fluctuation (RMSF), radius of gyration. RMSD deduced the average structural difference while the RMSF concluded the fluctuation of each residue in the system. This way RMSF can infer which residues fluctuates most and RMSD infers overall conformational changes.

4.2 Analysis of MD simulation run

The RMSD profile of M^{pro} 10ns simulation is shown is figure 13. Not much fluctuations were found in the graphs showing that protein is stable and well equilibrated in water. Similar profiles were found for RMSF and radius of gyration. These results

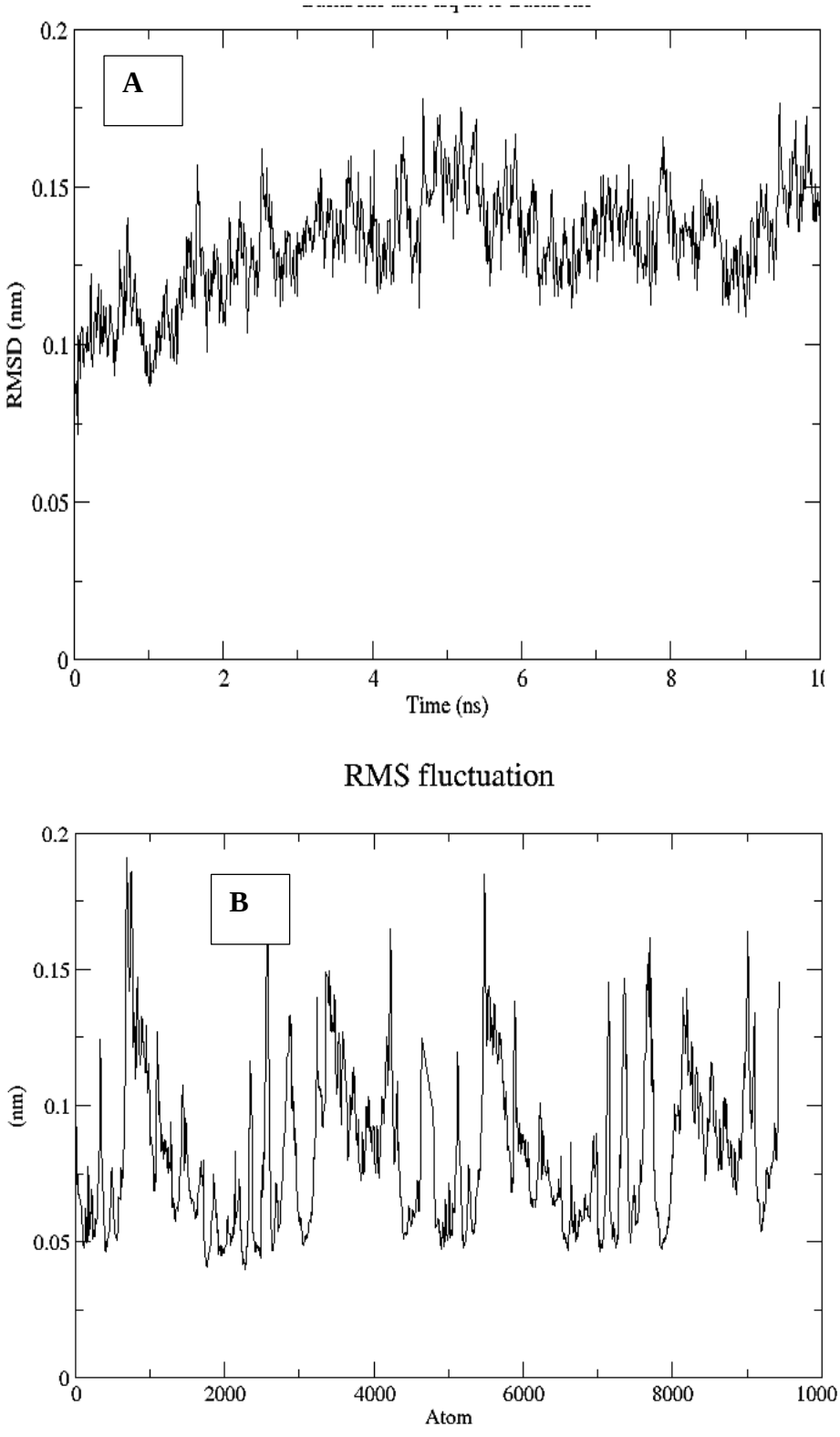


Figure 13:**A** dynamics of the system shown by RMSD and RMSF. **A**: On x-axis is simulation time scale and Y axis shows the deviation in nm
B: Fluctuation of each residue in nm on Y axis and number of reidues in the protein in x axis

The average RMSD value of the system was found to be 8.202314e-02, and the standard deviation 2.822660e-02. Firgure 14 shows the Radius of gyration plot. Radius of gyration

is generally related to the size and compactness of the system. In the later parts of results rg of protein and ligand is also reported.

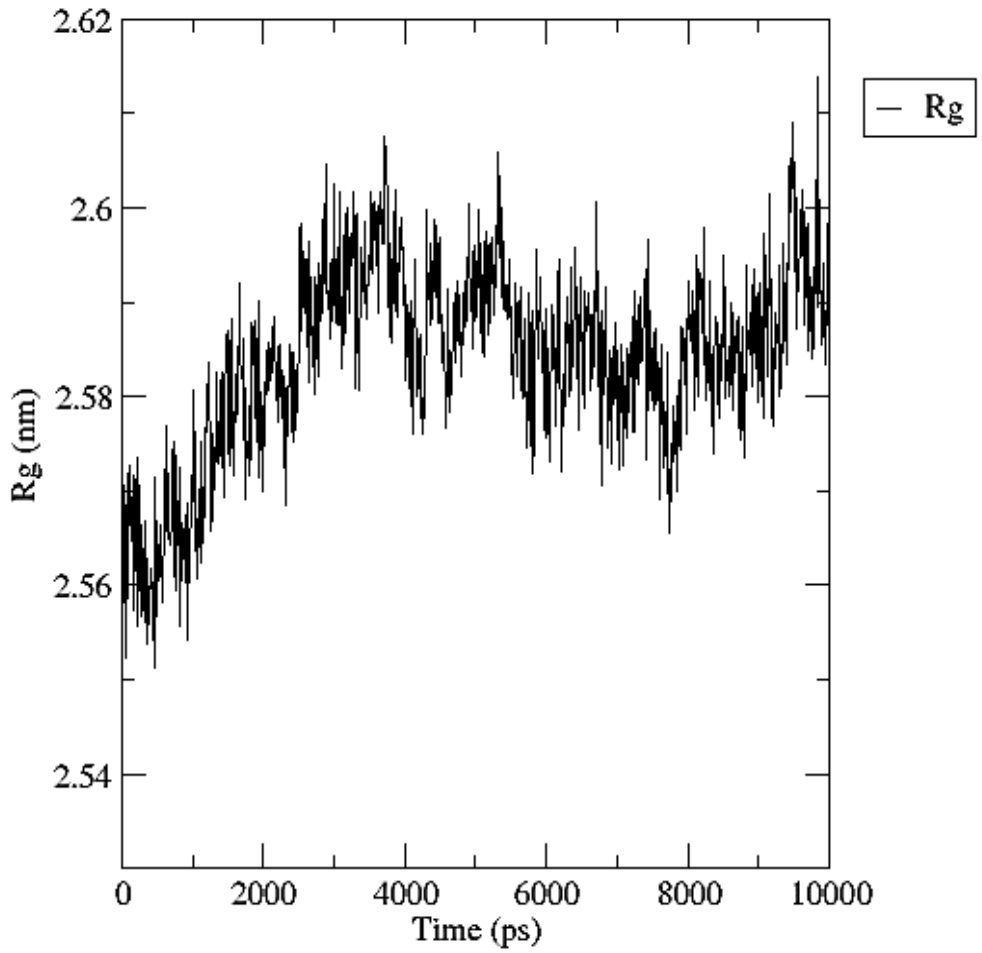


Figure 14: Radius of Gyration of the protein during 10 ns long MD run.

A trend of increased Rg was seen. A sharp increase between 0-1000 ps was found and after 3000ps it was found quite stable throughout. This dictates some expansion of the reference structure when the dynamics started.

4.3 Metadynamics

This section discusses the major results of the study on the Active and non-Active pocket systems, their details, Metadynamics outputs, Analysis of output, how unrestrained metadynamics is refined to perform another run of restrained metadynamics study, Free energy landscape construction and it's processing to report binding Free energy, and the calculation of dissociation constant.

After the systems were equilibrated a binary file, .tpr containing all the details for simulation was generated to give as an input for the metadynamics run. The same .tpr was used for restrained metadynamics run also.

4.3 non-Active pocket system

Results of unrestrained and restrained metadynamics run of the non-Active pocket system is presented in this subsection including the CV analysis, FES construction and binding free energy calculation.

Visual of the chosen three atoms on the basis of which the virtual points to command PLUMED to calculate angle and apply angle restraining is also presented.

4.3.1.1 Unrestrained Metadynamics run for the non-Active pocket system

Given the reported parameters of Gaussian kernels and CV description in the methodology section, unrestrained metadynamics simulation for less than 300 ns was conducted. Figure 15 shows distance CV a function of simulation time. Several crossings were found during the simulation, showing that the CV space has been explored by the protein-ligand system at 300 ns

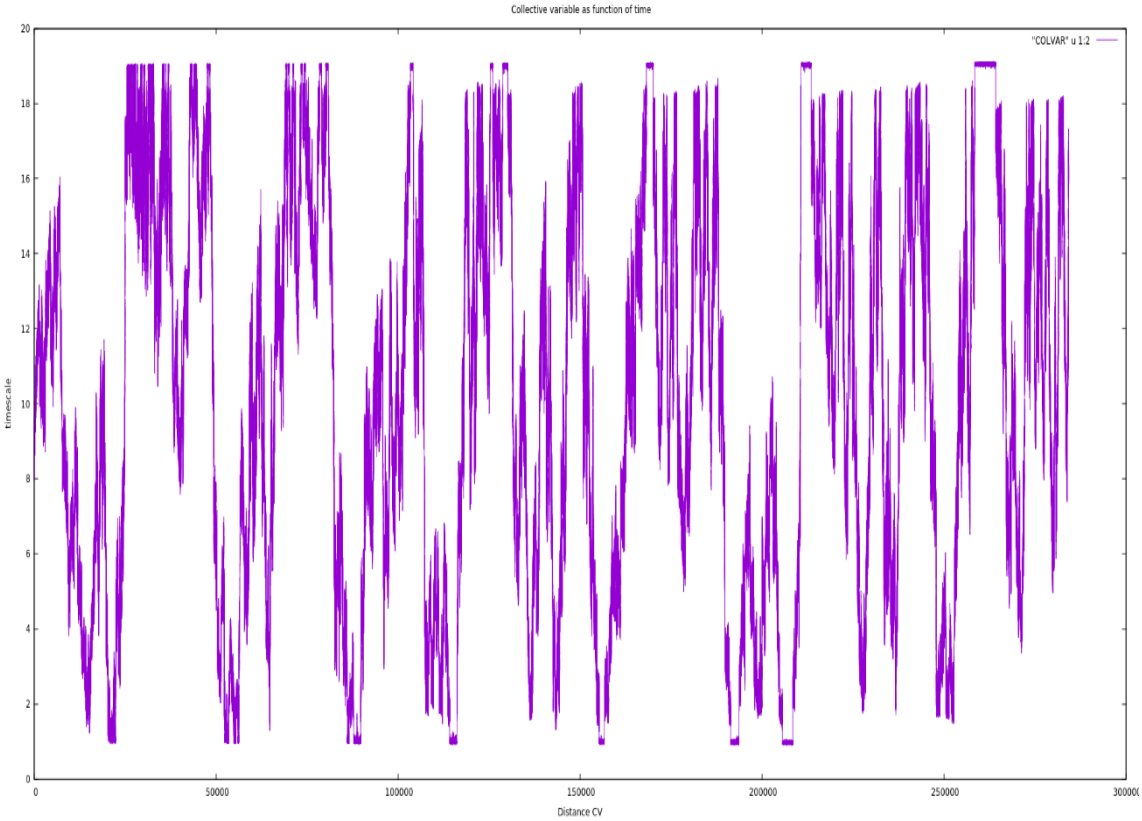


Figure 15 CV exploration shown by a plot of CV vs timescale for unrestrained metadynamics run of non-Active pocket system

X-axis shows distance CV and y-axis shows timescale of simulation in ps. Graph explains that the CV space is well explored showing the conformation corresponding to the CV values are sampled.

4.3.1.1.1 Analysing the output trajectory file

In the unrestrained simulation, we found that the ligand interacted with farther parts of the protein surface at large values of CV (results not shown). To ensure that at large values of CV, only unbound conformations of the ligand are explored, we chose to use angle restrains to limit our sampling. The unrestrained simulation trajectory was analysed to select three nearly collinear molecules details of which are mentioned in the methodology section. Here we present a VMD snapshot (figure 16) of the representation showing three collinear points and the angle between them calculated by VMD.

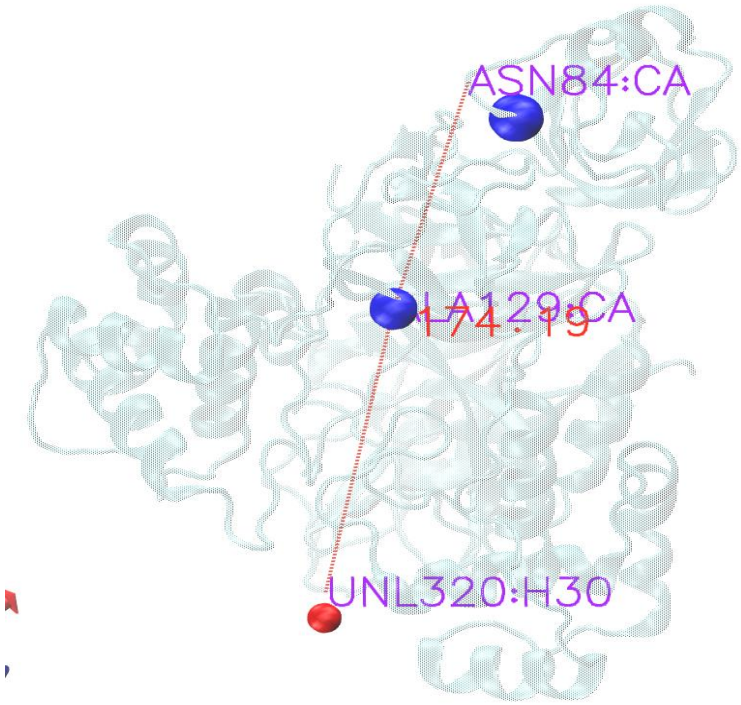


Figure 16: A visual of two residues of protein that are falling nearly collinear with the ligand molecule. Blue-colored balls represent protein residues and the red-colored ball is the ligand molecule.

Along the path of binding and unbinding the chosen 3 points were ligand and other two points on protein appeared to make angle of 174.19^{o} when visualized by VMD. The points were found to make angle around 180^{o} so were chosen to be used in the input file of restrained metadynamics. The angle restraining was imposed such that the center of masses

of these points can make angle between 160^{o} to 180^{o}. As a result of this restraining, we were able to sample the system only along the interested path.

4.3.1.2 Restrained Metadynamics run for non-Active pocket

The atomic information like residue indexes of pocket were given in the plumed input file to calculate center of mass and then angle between those centers of masses to apply angle restraining.

This can be monitored by plotting a CV graph of the data. One can refer to Figure 19,20 showing the CV values upon the application of restraining. The choice of CV greatly governs the effectiveness of metadynamics. Beautiful exploration of CV space of our system can be seen. This should be noted that the restraining through boundaries was applied on angle and the metadynamics bias potential was made to act only on the distance as CV to sample the system's conformations. We also present the result of angle restraining with the help of superimposed snapshots of the simulation trajectory visualized in VMD provided in figure 19. All the positions, the ligand has explored in complexed with the protein within all the frames are superimposed and represented in one snapshot. The area it has explored with in the protein and outside the protein was found to make a cone as a result of angle restraining. Figure 17 and 18 shows the graph representing collective variable of the system calculated by PLUMED.

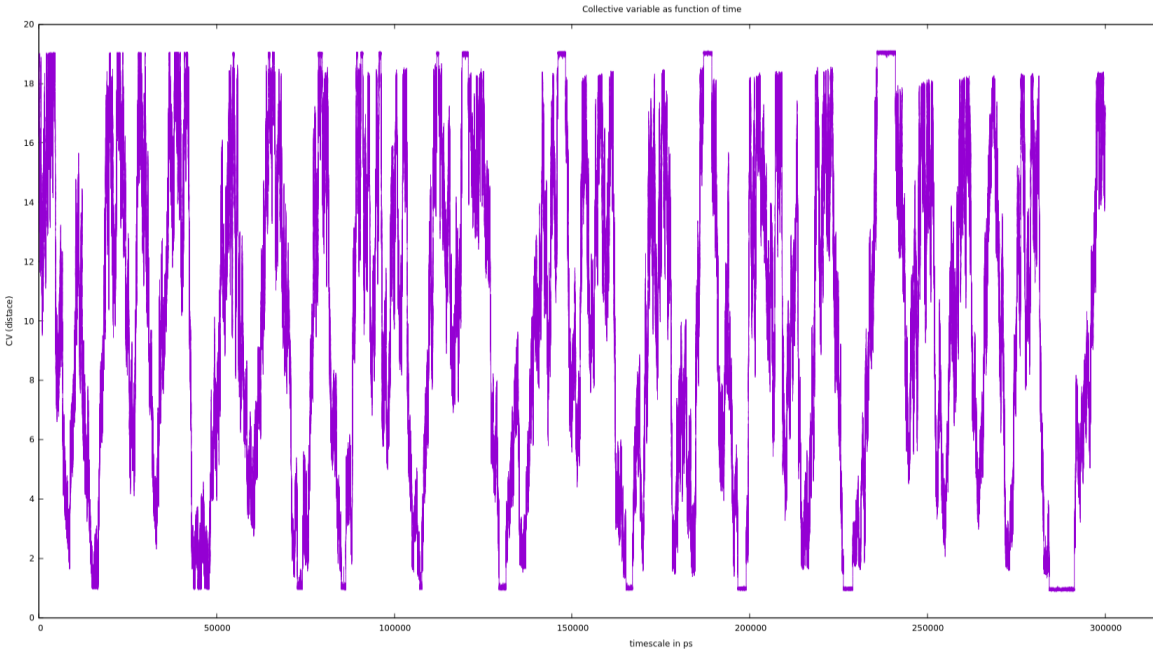


Figure17: Exploration of Distance CV space as a function of time during refined metadynamics run

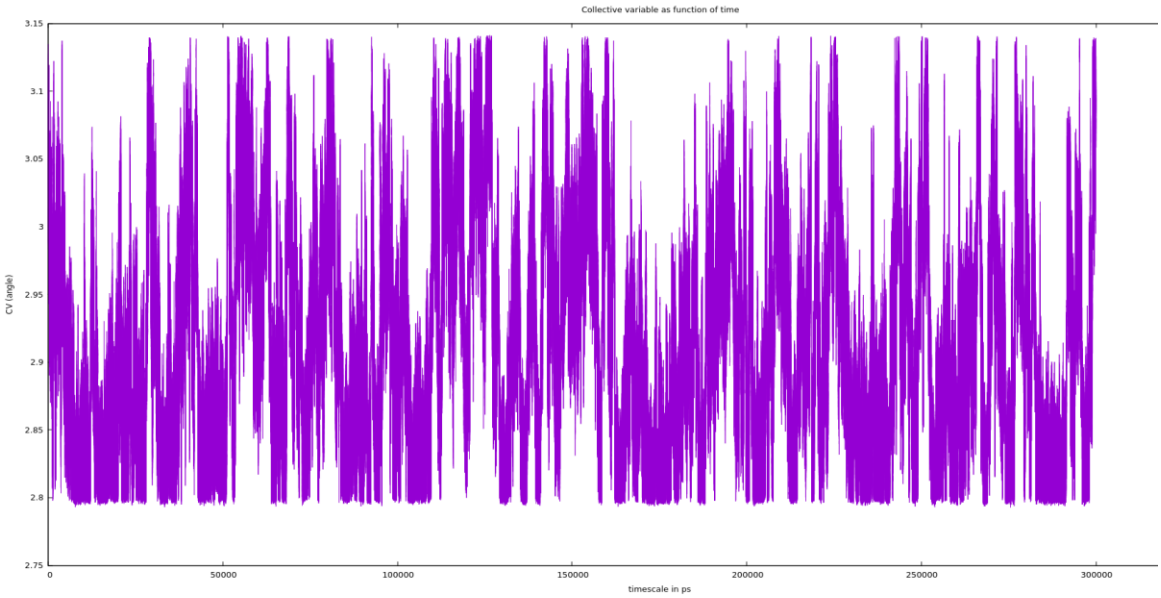


Figure18: Exploration of Angle CV space as a function of time during

The above plots dictate good exploration of sample space. The CV space is continuously explored without getting stuck at a place indicating good choice of CV as well a good exploration of phase space. The plots can predict that the choice of CV is able to access more regions of the phase space already visited.

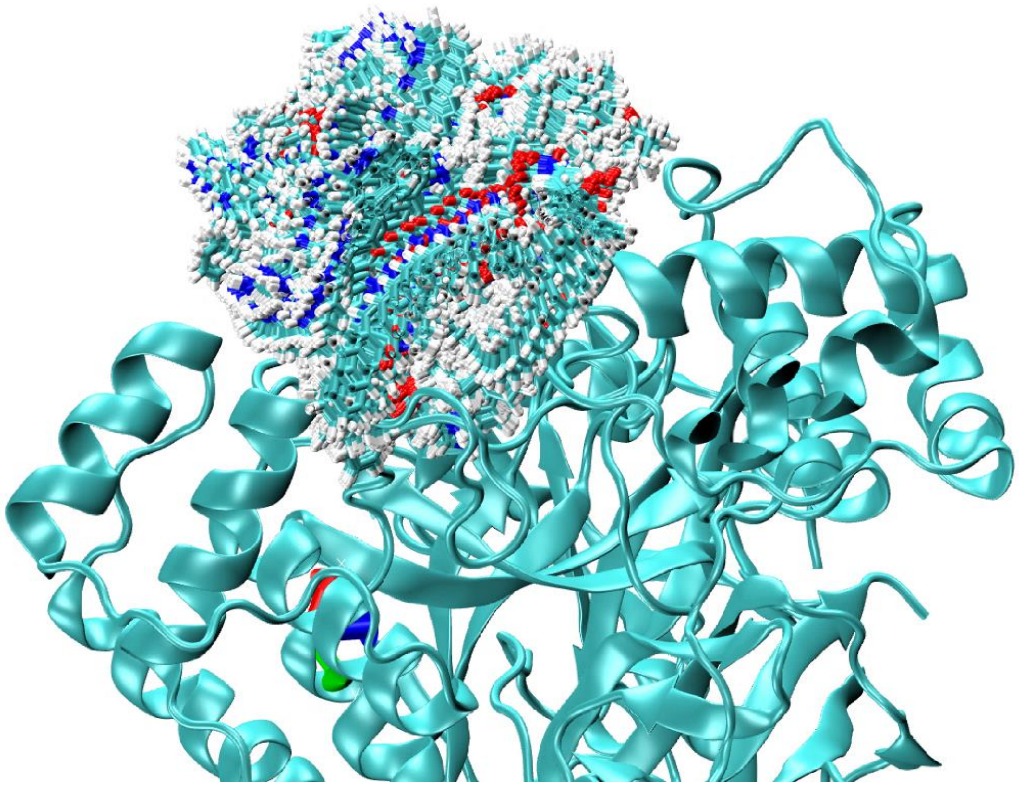


Figure 19: Top view of the superimposed ligand representing a cone as a result of angle restraining between three virtual point

Free Energy Landscape

Free energy landscape was produced after summing the hills using sum_hills command. The FES generated is shown in. figure 22 The evolution as well as convergence of the FES is shown in figure no 21.

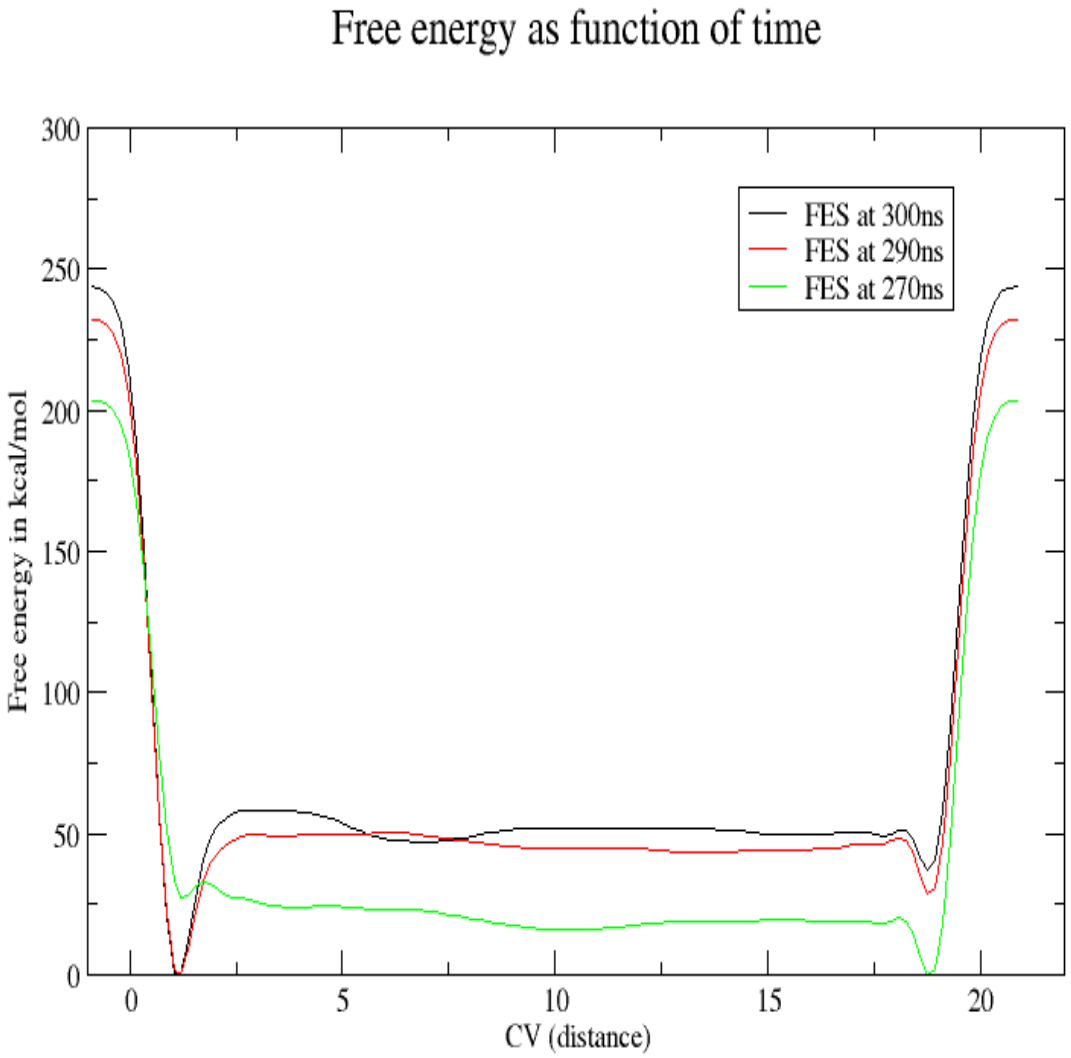


Figure 20 Free energy landscape showing evolution of FES with some unusual patterns.

However, on noticing the values of minimum free energies, the values seemed unusual. Upon plotting a histogram of the Collective variable file (figure 21), we found that due to the presence of boundary conditions near the grid min and grid max values, the CV area corresponding to those value is not sampled enough to include those CV space areas in FES calculation. Hence, those values were excluded and FES was expanded to the CV values that are sampled sufficiently to calculate FES.

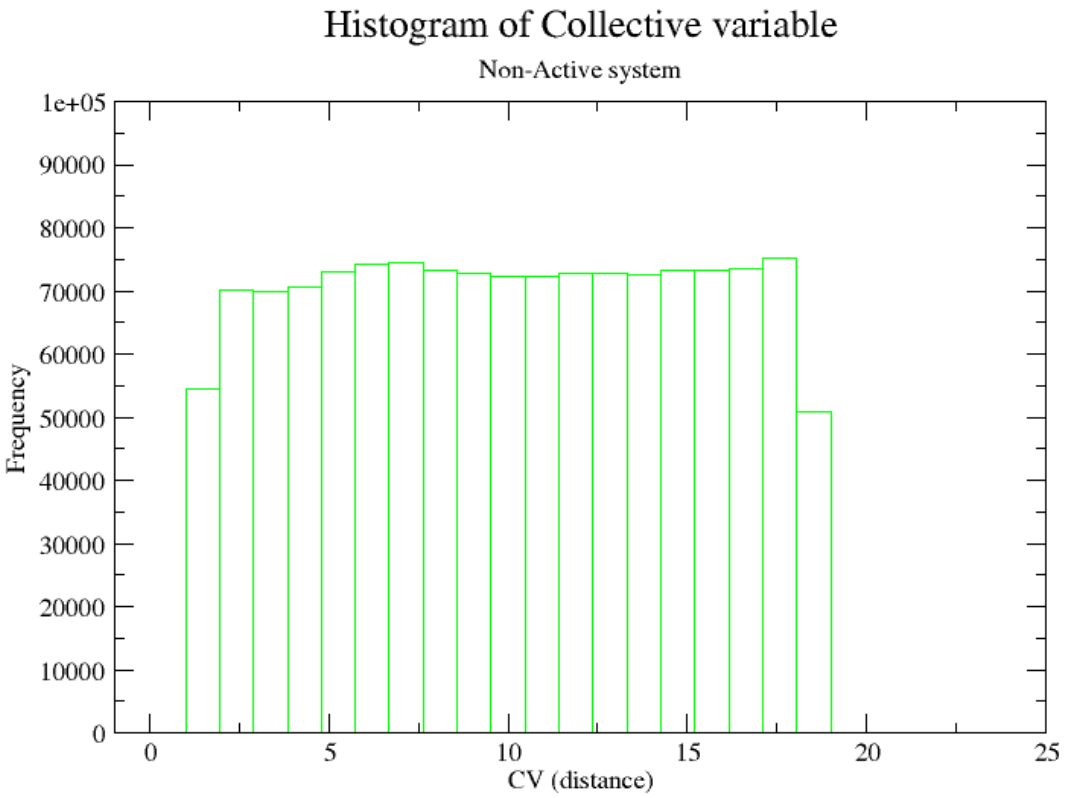


Figure 21: Histogram of the CV data with respect to frequency to investigate sampling of conformations at different CV values.

CV value 1-2 and 18-19 were found to be subjected to insufficient sampling because upper and lower bound were present near to them. The modification in FES was made to make sure these CV values are not included. After trimming this data, we constructed FES from which the subsequent difference between the two minima is calculated to report ΔΔG value. Free energy landscape after trimming the undesired values is shown in figure 23. First point represents the bound state and the second point represent unbound state. The lowest value of first valley was taken as ΔG for bound state and the average between the CV range 10-14 was taken as ΔG for unbound state.

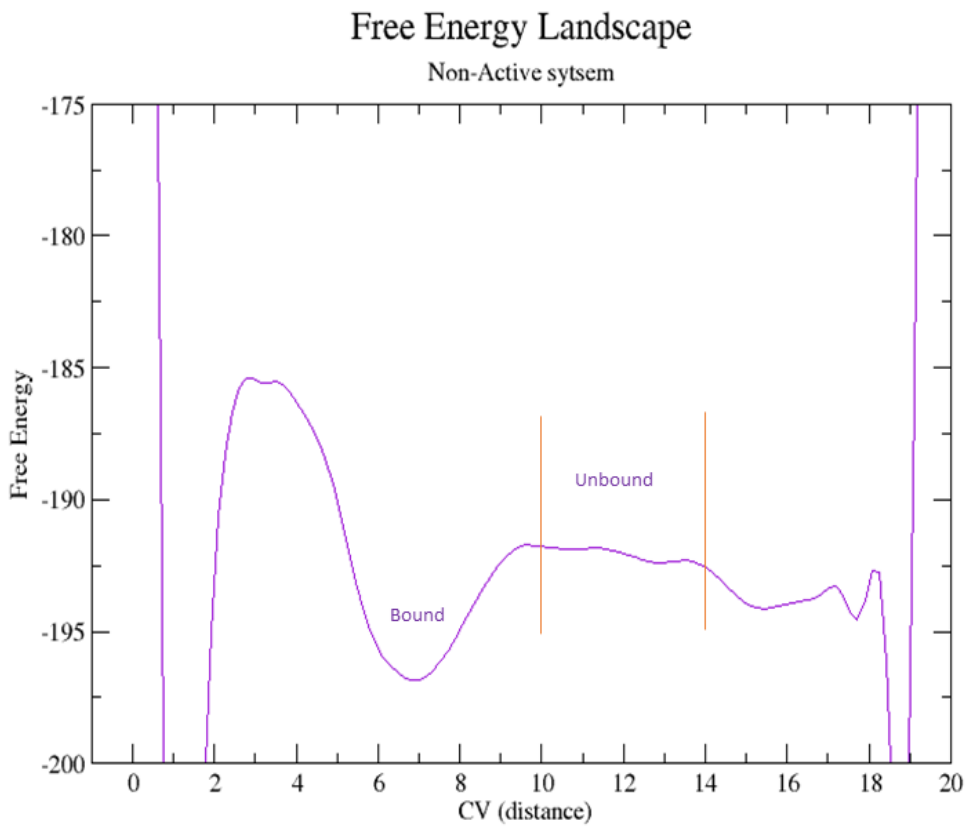


Figure 22: Free Energy Landscape Of non-Active pocket system

The system conformations at CV values near 6Å corresponds to the bound state of protein and ligand. In a broader view of this map, it was found that the structure of the map after 10Å looked nearly leveled till it reaches the upper boundary (figure 21). Corresponding to the unbound state, it exhibited similar energy when it was found unbounded with the protein. The minimum free energy taken for calculation was -196.855 at CV value 7.004 and the average value of free energy at CV value 10-14 was found to be -192.091 kcal/mol corresponding to the unbound state was taken for calculation.

$$\Delta G1 = -192.091$$

$$\Delta G2 = -192.091\, kcal/mol$$

$$\Delta G1 - \Delta G2 = \Delta\Delta G$$

$$-196.855 - (-192.091) = \Delta\Delta G$$

$$\Delta\Delta G = -4.764\, kcal/mol$$

Binding Free energy or the Free energy difference was found to be -4.764 kcal/mol

On comparing the vina score we got from docking results and the more accurate binding free energy we found the difference of -6.93 kcal/mol. Hence Vina score does not give reliable binding free energies. Although docking studies are the easy way to model a protein-ligand system if the complex does not have a crystal structure and computational examination is the objective.

Calculation of dissociation constant

Dissociation constant broadly describes the affinity between two species. Usually molecular recognition phenomenon. It states the propensity of a Protein-ligand complex to break into Protein + ligand.

$$[PL] \rightarrow [P] + [L]$$

$$Temperatureofthesystem = 303.15k$$

$$R = 1.9872\, cal/K\,.mol$$

$$RT = 596.4580\, cal/mol$$

$$\Delta\Delta G = -4.764\, kcal/mol$$

$$K_d = e^{\frac{\Delta\Delta G}{RT}}$$

$$K_d = e^{\frac{-4.764}{596.4580}}$$

$$K_d = 9920.44$$

The dissociation constant for the ligand docked at the pocket was found to be 9916.55 kcal/mol.

4.3.2 Active pocket System

The purpose of unrestrained metadynamics run was to find the nearly collinear points. Therefore, the Active pocket system was simulated for 150 ns. After discovering the colinear points in the Active pocket system angle restraining was applied for the second round of metadynamics run. Further FES calculation and analysis of trajectory were also performed and are presented by figures.

4.3.2.1Unrestrained Metadynamics run for Active pocket system

We present the exploration of CV space during the run and the result of subsequent analysis performed on the output trajectory file. The analysis of trajectory gave us the points to

apply angle restraining between them. For this section we will present the plot of COLVAR file (figure 23) and visual of three nearly collinear molecules (figure 24).

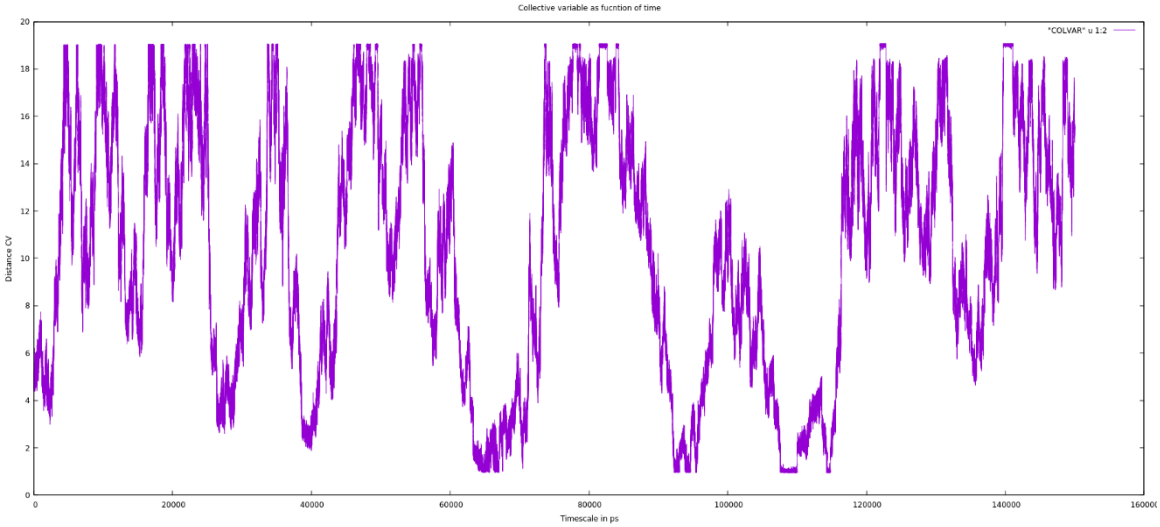


Figure 23: Exploration of distance CV space in the unrestrained metadynamics run for Active Pocket system

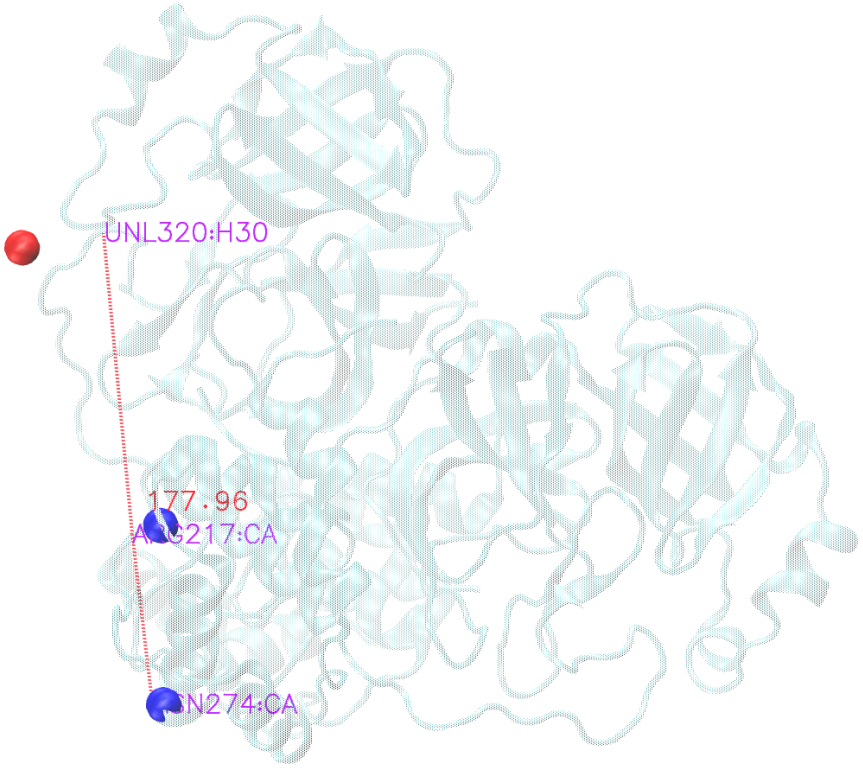


Figure 24: A visual of two residues of protein that are falling nearly collinear with the ligand molecule. Blue-colored balls represent protein residues and the red-colored ball is the ligand molecule.

The points were selected such that they make an angle closest to 180^{o} so as to provide only a few degrees of freedom when angle restraining is applied. The angle they were found to form at the time of their selection was 177.96^{o} .

Restrained metadynamics for Active Pocket system

After providing the atom indexes of residues that were found to be preset within the 7Å radius of the two protein residues that were falling collinear with the ligand (figure 25), the center of masses was calculated by the plumed and angle restraining is added along with the distance calculation. The bias potential was added as a function of distance CV. Collective variable plots represents the values of the CVs calculated by plumed are presented in figures 25 and 26.

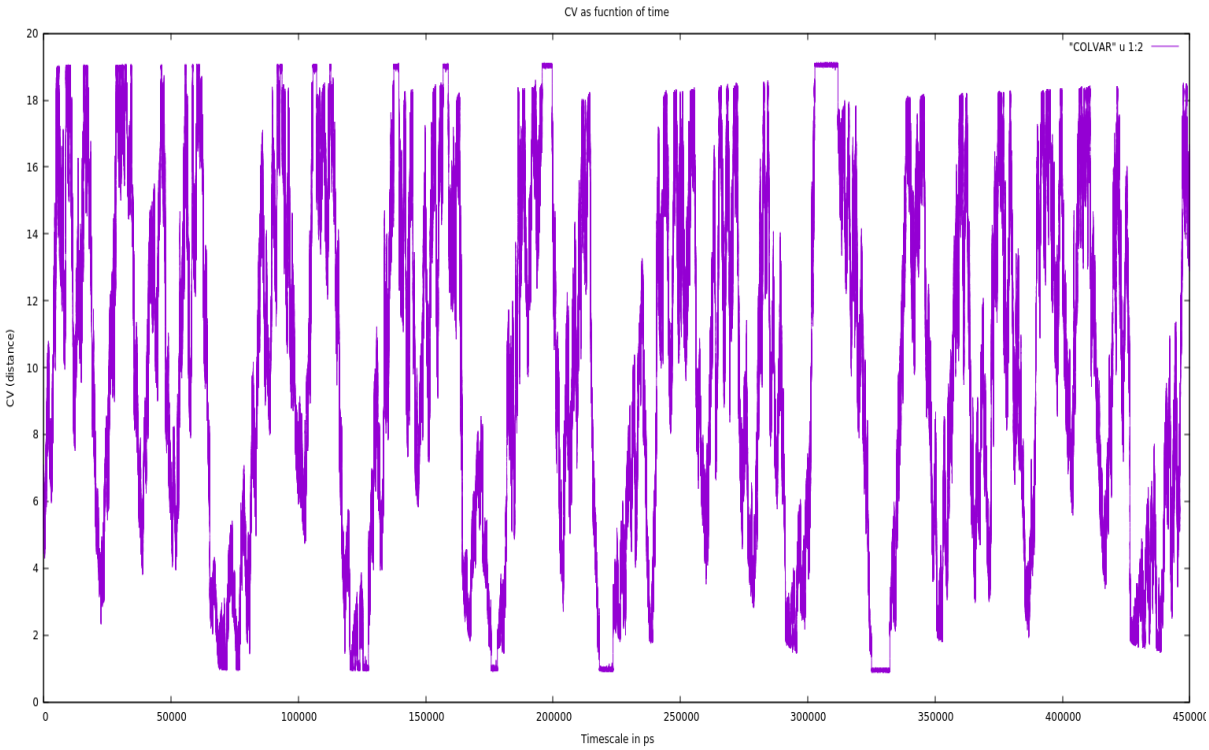


Figure 25: Exploration of distance CV space as a function of time during refined metadynamics run

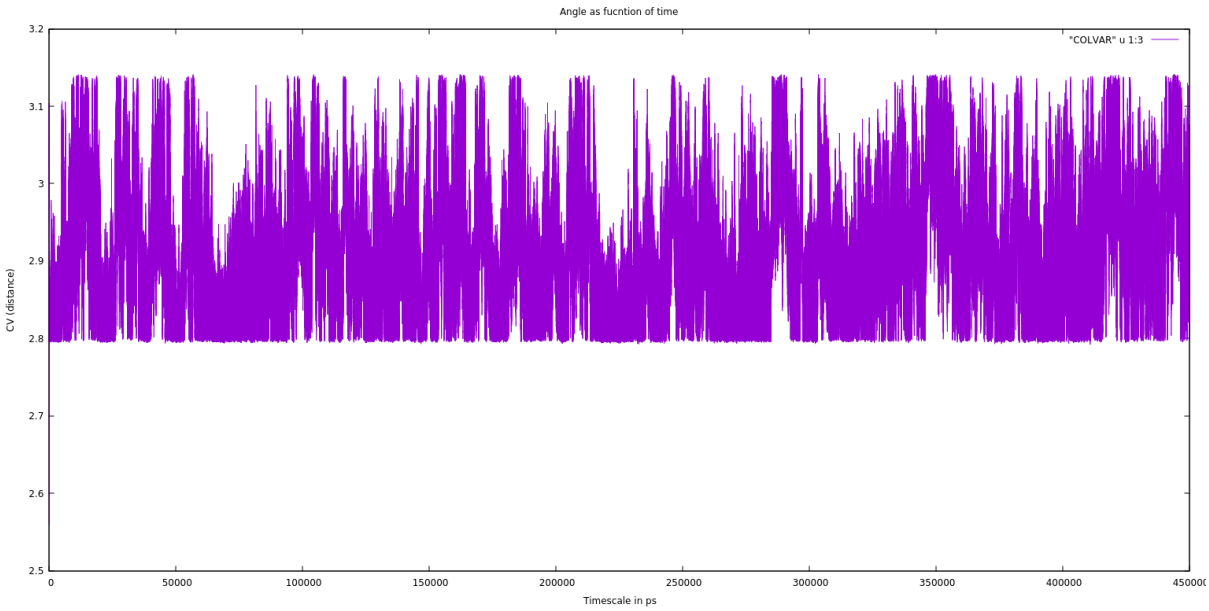


Figure 26: Angle as function of time

As a result, the ligand was restrained to travel within the range of 1-19Å making angle in the range 160^{o} -180^{o} . This event was visualized from VMD In which ligand's position within all the frames were recorded at a time (figure 27). A representation of so called was formed as a result of angle restraining.

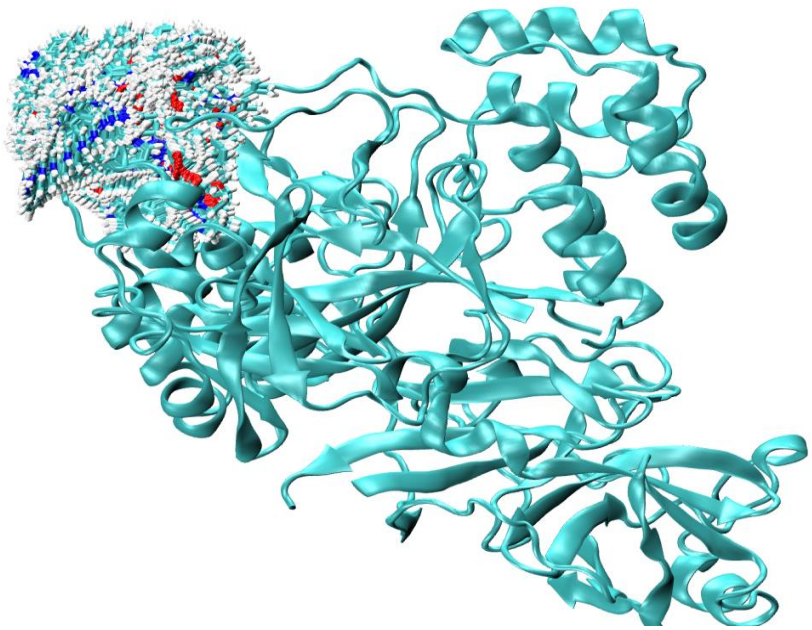


Figure 27: Superimposed snapshots of ligands forming a cone representing angle restraining

Free energy Landscape

The FES generated was checked to confirm the convergence of the metadynamics run. The last 30 ns FES were plotted and final FES was found emerging. (figure 28).

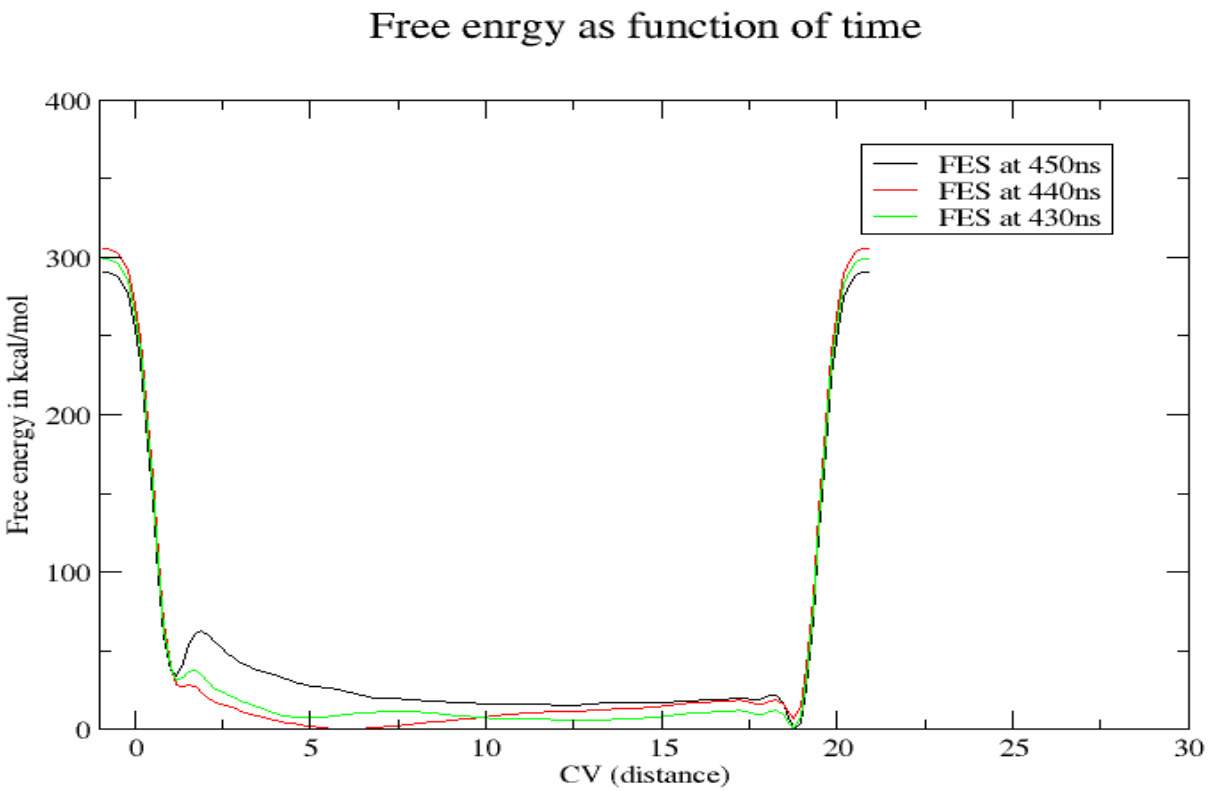


Figure 28: FES estimated at last 30 ns showing evolution of FES and convergence of metadynamics.

Noticing the unusual pattern in the FES, histogram of collective variable as function of time was formed and the same results were found which are shown in figure 29.

With further investigation of FES generated was done with a histogram of the FES data, same pattern was observed as was observed during the FES calculation of the non-Active pocket system. We found that the CV values near the boundary conditions were not sampled and hence were not fit to be included in the FES construction. So there was a need to eliminate the data during FES construction. The final FES was obtained as the expanded form of the crude FES.

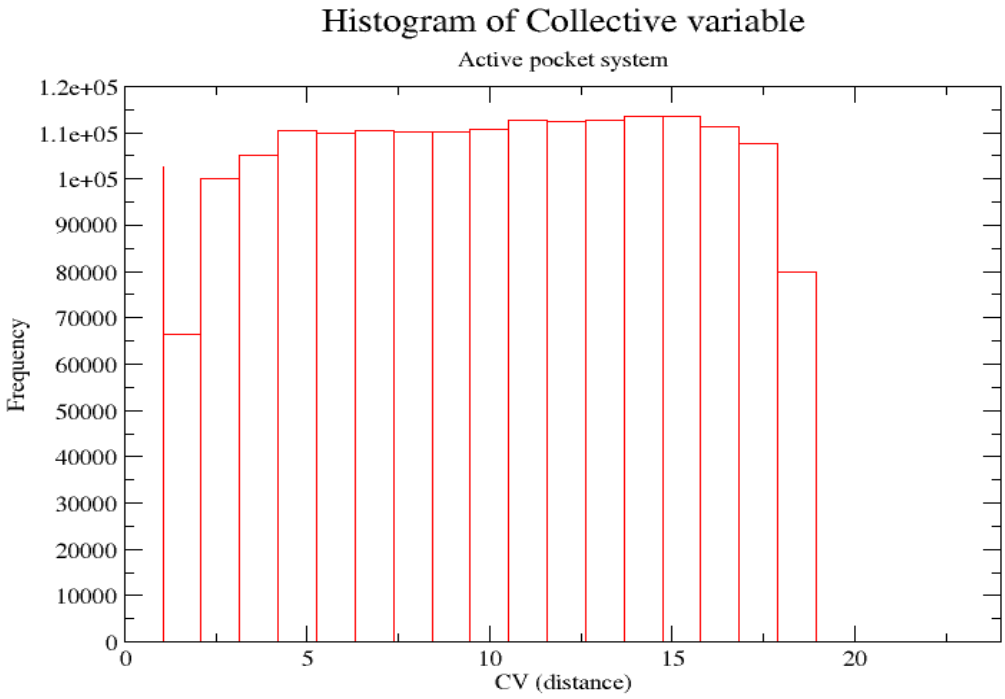


Figure 29: Histogram revealing bad sampling at CV values near restraining boundary conditions.

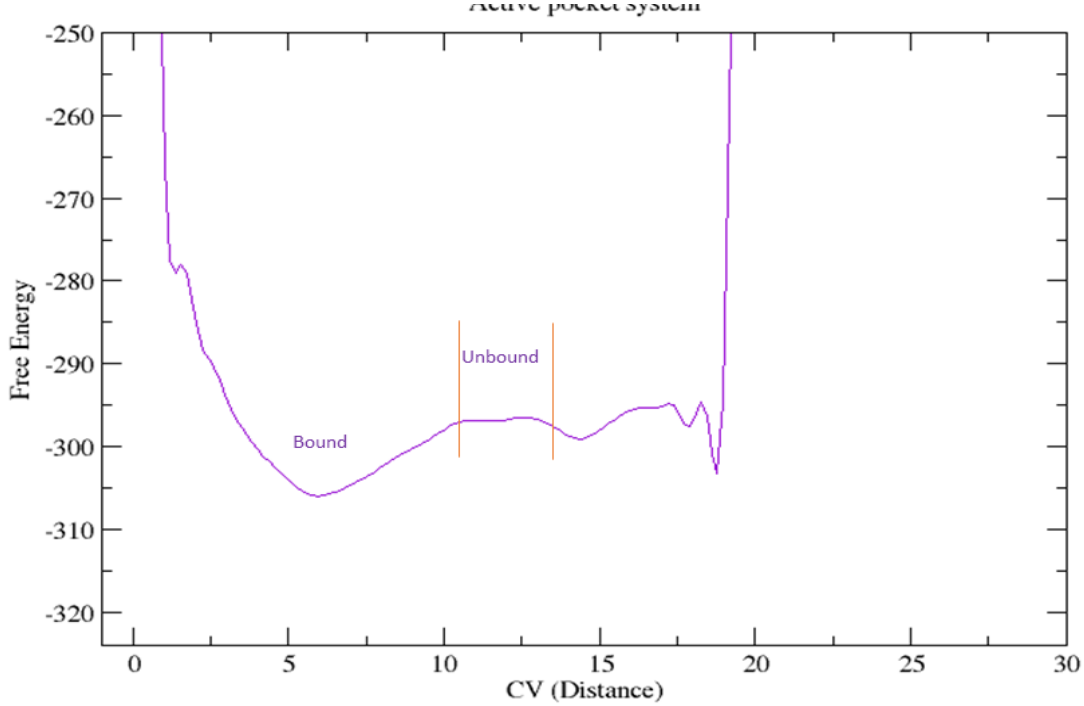


Figure 30: Free Energy Landscape of Active pocket system

At CV value 5.9Å the lowest energy, 305.95 kcal/mol was found that corresponds to the bond state while at the range of CV value 10.5-14.5 , 297.057 kcal/mol was found corresponding to the unbound state (figure 30). The diefference, i.e ΔΔG was found to be 8.893 kcal/mol.

Calculation of dissociation constant

$$R = 1.9872\, cal/K\,.mol$$

$$RT = 596.4580\, cal/mol$$

$$\Delta\Delta G = -8.893\, kcal/mol$$

$$K_d = e^{\frac{\Delta\Delta G}{RT}}$$

$$K_d = e^{\frac{-8.893}{596.4580}}$$

$$K_d = e^{-0.014909}$$

$$K_d = 9852.01$$

$$Equilibrium\ constant = \frac{k_{d_A}}{k_{d_B}}$$

$$= \frac{9852.01}{9920.44}$$

A lower value of k_d exhibiting by the ligand in novel Active pocket system suggests better affinity towards the pocket while the ligand exhibits less affinity at the Active Pocket region. On the basis of k_d values, it was indicated that Active pocket exhibits a better profile however the binding free energy values of non-Active pocket systems are not negligible. The ligand shows not better but some potency to bind to the non-Active site of the protein.

5 CONCLUSION

The search for probable binding modes with favorable interaction energies and surface complementarity by employing physics-based or geometric models was made possible by employing molecular docking techniques.

The complex of the protein and ligand system was successfully formed with docking studies. The topmost of the ligand was generated and the top most region out of three regions in the protein was also selected to prepare the system. The important substrate binding residues in nonbonded interactions were found to be Arg4, Lys 137, Tyr 237, Tyr 239, and Leu 287.

Results of docking were found to be further refined when plotting of free energy surface graph was done to depict the specificity of the ligand to the region of the pocket. Enhanced sampling through metadynamics simulation was found to be an effective way of calculating free energy surface graph and eventual binding free energy or the difference in binding energy of the two bound and unbound states. This led to the simple calculation of dissociation constant (equilibrium constant) also.

We employed enhanced sampling method, metadynamics to accurately compute the binding free energy of the ligand in two systems. The technique revealed energy difference of -4.764 kcal/mol for the non-active pocket system where ligand was present at the novel pocket that is other than the so-called active pocket (comprising the well-known his41- cys 145 dyad) while this difference for the system where the ligand was present at the active site pocket (his 41 and Cys 145) was found to be -8.893 kcal/mol. The vina score from flexible-ligand rigid-proteinn docking studies reported score of -11.7 kcal/mol for the non-Active pocket system and -9.5 kcal/mol for Active site pocket, reflecting an error of values -6.936 and 0.607 for non-Active and Active pocket systems, respectively, in docking estimates.

Implementation of such computational method can aid to speedy designing of therapeutic drugs. Employing such methods can examine new region of protein that can be compared with other pocket region. This shift from basic region could bring about scope of discovery new pocket or important residues that can be targeted in future studies.

In the metadynamics studies the selection of CV and proper restraining on angle refined the convergence and plotting of binding free energy which aid to the subsequent comparison of the two system to draw out important remarks. The study presented also

gave the insights into the molecular recognition of two species at two different sites, namely, the active pocket site and the non-active pocket site by employing advanced computational technique like metadynamics. The comparison was made between the two systems by studying their free energy landscapes, the dissociation constants. At the end we conclude that due to the broadened scope of molecular dynamics by employing enhanced sampling techniques the examination on novel protein pockets can be done efficaciously.

6 REFERENCE

A., Pedersen, L. G., & Anderson, M. W. (1993). Molecular dynamics simulation of HIV-1 protease in a crystalline environment and in solution. *Biochemistry*, *32*(6), 1443-1453.

Abdel-Maksoud, K., al-Badri, M. A., Lorenz, C., & Essex, J. W. (2020). Allosteric regulation of SARS-CoV-2 protease: towards informed structure-based drug discovery.Abraham, D. J. Structure-based Drug Design. A Historical Perspective and the Future. ComprehensiVe Medicinal Chemistry II*; Elsevier: Oxford, U.K.*, 2007.

Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., & Lindahl, E. (2015). GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, *1*, 19-25.

Adamcik, J., & Mezzenga, R. (2018). Amyloid polymorphism in the protein folding and aggregation energy landscape. *Angewandte Chemie International Edition*, *57*(28), 8370-8382.

Alberg DG, Schreiber SL (1993) Structure-based design of a cyclophilin– calcineurin bridging ligand. *Science* 262:248–250

Andersen, K. G., Rambaut, A., Lipkin, W. I., Holmes, E. C. & Garry, R. F. (2020) The proximal origin of SARS-CoV-2 *Nature Medicine* 26, 450–452

Bahar, I.; Chennubhotla, C.; Tobi, D. Intrinsic dynamics of enzymes in the unbound state and relation to allosteric regulation. *Curr. Opin. Struct. Biol*. 2007, 17, 633–640.

Bai, F., Morcos, F., Cheng, R. R., Jiang, H., & Onuchic, J. N. (2016). Elucidating the druggable interface of protein− protein interactions using fragment docking and coevolutionary analysis. *Proceedings of the National Academy of Sciences*, *113*(50), E8051-E8058.

Barducci, A., Bonomi, M., & Parrinello, M. (2011). Metadynamics. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, *1*(5), 826-843.

Barnes, J. E. (1989). Evolution of compact groups and the formation of elliptical galaxies. *Nature*, *338*(6211), 123-126.

Bekker, H., Berendsen, H. J. C., Dijkstra, E. J., Achterop, S., Vondrumen, R., VANDERSPOEL, D., ... & Renardus, M. K. R. (1993). Gromacs-a parallel computer for

molecular-dynamics simulations. In *4th International Conference on Computational Physics (PC 92)* (pp. 252-256). World Scientific Publishing.

Berendsen, H. J., van der Spoel, D., & van Drunen, R. (1995). GROMACS: A message-passing parallel molecular dynamics implementation. *Computer physics communications*, *91*(1-3), 43-56.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., & Bourne, P. E. (2000). The Protein Data Bank. Nucleic acids research, 28(1), 235–242. https://doi.org/10.1093/nar/28.1.235

Betzi, S.; Suhre, K.; Chetrit, B.; Guerlesquin, F. and Morelli, X. (2006) *J. Chem. Inf. Model*, 46(4), 1704-1712.

Bolton, W., & Perutz, M. F. (1970). Three dimensional Fourier synthesis of horse deoxyhaemoglobin at 2.8 Å resolution. *Nature*, *228*(5271), 551-552.

Bonomi M, Bussi G Et.al (2019). "Promoting transparency and reproducibility in enhanced molecular simulations." *Nature methods* 16, no. 8 670-673.

Branduardi D, Gervasio F, Parrinello M (2007). From A to B in free energy space. *J Chem Phys* , 126:054103

Brooks BR, Brooks CL 3rd, Mackerell AD Jr, Nilsson L, Petrella RJ, Roux B, Won Y, Archontis G, Bartels C, Boresch S, Caflisch A, Caves L, Cui Q, Dinner AR, Feig M, Fischer S, Gao J, Hodoscek M, Im W, Kuczera K, Lazaridis T, Ma J, Ovchinnikov V, Paci E, Pastor RW, Post CB, Pu JZ, Schaefer M, Tidor B, Venable RM, Woodcock HL, Wu X, Yang W, York DM, Karplus M (2009) CHARMM: the biomolecular simulation program. *J Comput Chem 30* (10):1545–1614

Bussi, G., Gervasio, F. L., Laio, A., & Parrinello, M. (2006). Free-energy landscape for β hairpin folding from combined parallel tempering and metadynamics. *Journal of the American Chemical Society*, *128*(41), 13435-13441.

Casalino, L., Gaieb, Z., Goldsmith, J. A., Hjorth, C. K., Dommer, A. C., Harbison, A. M., ... & Amaro, R. E. (2020). Beyond shielding: the roles of glycans in the SARS-CoV-2 spike protein. *ACS central science*, *6*(10), 1722-1734.

Chen, R., Li, L., & Weng, Z. (2003). ZDOCK: an initial-stage protein-docking algorithm. *Proteins: Structure, Function, and Bioinformatics*, *52*(1), 80-87.

Chen, R., Tong, W., Mintseris, J., Li, L., & Weng, Z. (2003). ZDOCK predictions for the CAPRI challenge. *Proteins: Structure, Function, and Bioinformatics*, *52*(1), 68-73.

Chen, Y. C. (2015). Beware of docking!. *Trends in pharmacological sciences*, *36*(2), 78-95.

Chipot, C., & Pohorille, A. (2007). *Free energy calculations* (Vol. 86). Berlin: Springer.

Chong, S. H., Im, H., & Ham, S. (2019). Explicit characterization of the free energy landscape of pkid–kix coupled folding and binding. *ACS central science*, *5*(8), 1342-1351.

Chrencik, J.E.; Staker, B.L.; Burgin, A.B.; Pourquier, P.; Pommier, Y.; Stewart, L. and Redinbo, M.R. (2004) *J. Mol. Biol.*, 339(4), 773-784.

Churakov, S. V., Iannuzzi, M., & Parrinello, M. (2004). Ab initio study of dehydroxylation− carbonation reaction on brucite surface. *The Journal of Physical Chemistry B*, *108*(31), 11567-11574.

Corman, V. M., Muth, D., Niemeyer, D., & Drosten, C. (2018). Hosts and sources of endemic human coronaviruses. *Advances in virus research*, *100*, 163-188.

Dawe, G. B., Musgaard, M., Aurousseau, M. R., Nayeem, N., Green, T., Biggin, P. C., & Bowie, D. (2016). Distinct structural pathways coordinate the activation of AMPA receptor-auxiliary subunit complexes. *Neuron*, *89*(6), 1264-1276.

de Azevedo Jr, W. F., & Dias, R. (2008). Evaluation of ligand-binding affinity using polynomial empirical scoring functions. *Bioorganic & medicinal chemistry*, *16*(20), 9378-9382.

De Clercq, E. (2002). Strategies in the design of antiviral drugs. *Nature Reviews drug discovery*, *1*(1), 13-25.

Deeks, S. G., Hecht, F. M., Swanson, M., Elbeik, T., Loftus, R., Cohen, P. T., & Grant, R. M. (1999). HIV RNA and CD4 cell count response to protease inhibitor therapy in an urban AIDS clinic: response to both initial and salvage therapy. *Aids*, *13*(6), F35-F43.

Delemotte, L., Tarek, M., Klein, M.L., Amaral, C., and Treptow, W. (2011). Intermediate states of the Kv1.2 voltage sensor from atomistic molecular dynamics simulations. *Proc. Natl. Acad. Sci. USA* 108, 6109–6114.

Dellago, Christoph, and Peter G. Bolhuis (2009). "Transition path sampling and other advanced simulation techniques for rare events." *Advanced computer simulation approaches for soft matter sciences III* 167-233.

Dias, R., de Azevedo, J., & Walter, F. (2008). Molecular docking algorithms. *Current drug targets*, *9*(12), 1040-1047.Downes, M.; Verdecia, M. A.; Roecker, A. J.; Hughes, R.; Hogenesch, J. B.; et al. 2003 A chemical, genetic, and structural analysis of the nuclear bile acid receptor FXR. Mol. Cell 11, 1079–1092.

Dror, R.O., Green, H.F., Valant, C., Borhani, D.W., Valcourt, J.R., Pan, A.C., Arlow, D.H., Canals, M., Lane, J.R., Rahmani, R., et al. (2013). Structural basis for modulation of a G-protein-coupled receptor by allosteric drugs. *Nature* 503, 295–299

Durrant, J. D., & McCammon, J. A. (2011). Molecular dynamics simulations and drug discovery. *BMC biology*, *9*(1), 1-9.

Eiben, A. E., & Smith, J. E. (2003). *Introduction to evolutionary computing* (Vol. 53, p. 18). Berlin: springer.

Ensing, B., & Klein, M. L. (2005). Perspective on the reactions between F–and CH3CH2F: The free energy landscape of the E2 and SN2 reaction channels. *Proceedings of the National Academy of Sciences*, *102*(19), 6755-6759.

Ensing, B., Laio, A., Gervasio, F. L., Parrinello, M., & Klein, M. L. (2004). A minimum free energy reaction path for the E2 reaction between fluoro ethane and a fluoride ion. *Journal of the American Chemical Society*, *126*(31), 9492-9493.

Ewing, T. J., Makino, S., Skillman, A. G., & Kuntz, I. D. (2001). DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases. *Journal of computer-aided molecular design*, *15*(5), 411-428.

Fermi, G., Perutz, M. F., Shaanan, B., & Fourme, R. (1984). The crystal structure of human deoxyhaemoglobin at 1.74 Å resolution. *Journal of molecular biology*, *175*(2), 159-174.

Fogel, D. B. (1998). *Artificial intelligence through simulated evolution* (pp. 227-296). Wiley-IEEE Press.

Forli, S. (2015). Charting a path to success in virtual screening. *Molecules*, *20*(10), 18732-18758.

Fuzo CA, Degreve L (2014) Effect of the thermostat in the molecular dynamics simulation on the folding of the model protein chignolin. *J Mol Model* 18(6):2785–2794

González, M. A. (2011). Force fields and molecular dynamics simulations. *École thématique de la Société Française de la Neutronique*, *12*, 169-200.

Goodsell, D. S., Morris, G. M., & Olson, A. J. (1996). Automated docking of flexible ligands: applications of AutoDock. *Journal of molecular recognition*, *9*(1), 1-5.

H. C. Urey and C. A. Bradley Jr. The vibrations of pentatonic tetrahedral molecules. Phys. Rev., 38:1969–1978, 1931.

Hartl, F. U., & Hayer-Hartl, M. (2009). Converging concepts of protein folding in vitro and in vivo. *Nature structural & molecular biology*, *16*(6), 574-581.

Hellal-Levy, C.; Fagart, J.; Souque, A.; Wurtz, J. M.; Moras, D.; et al. (2000). Crucial role of the H11-H12 loop in stabilizing the active conformation of the human mineralocorticoid receptor. *Mol. Endocrinol*., 14, 1210–1221.

Hellal-Levy, C.; Fagart, J.; Souque, A.; Wurtz, J. M.; Moras, D.; et al (2000). Crucial role of the H11-H12 loop in stabilizing the active conformation of the human mineralocorticoid receptor. *Mol. Endocrinol.* 14, 1210–1221.

Henzler-Wildman, K., & Kern, D. (2007). Dynamic personalities of proteins. *Nature*, *450*(7172), 964-972.

Hess, B., Kutzner, C., Van Der Spoel, D., & Lindahl, E. (2008). GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *Journal of chemical theory and computation*, *4*(3), 435-447.

Holland, J. (1975). Book: Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology. *Control, and Artificial Intelligence*.

Hollingsworth, S. A., & Dror, R. O. (2018). Molecular dynamics simulation for all. *Neuron*, *99*(6), 1129-1143.

Hoover, W. G. (1985). Canonical dynamics: Equilibrium phase-space distributions. *Physical review A*, *31*(3), 1695.

Humphrey, W., Dalke, A., & Schulten, K. (1996). VMD: visual molecular dynamics. *Journal of molecular graphics*, *14*(1), 33-38.

Iannuzzi, M., Laio, A., & Parrinello, M. (2003). Efficient exploration of reactive potential energy surfaces using Car-Parrinello molecular dynamics. *Physical Review Letters*, *90*(23), 238302.

J. D. Chodera, D. L. Mobley, M. R. Shirts, R. W. Dixon, K. Branson and V. S. Pande (2011). Alchemical Free Energy Methods for Drug Discovery: Progress and Challenges, Curr. Opin. Struct. Biol., 21(2), 150–160

Jain, A. N. (2003). Surflex: fully automatic flexible molecular docking using a molecular similarity-based search engine. *Journal of medicinal chemistry*, *46*(4), 499-511.

Jensen, M.Ø., Jogini, V., Borhani, D.W., Leffler, A.E., Dror, R.O., and Shaw, D.E. (2012). Mechanism of voltage gating in potassium channels. Science 336, 229–233.

Jerome, K. R. (2005). The road to new antiviral therapies. *Clinical and Applied Immunology Reviews*, *5*(1), 65-76.

Jin, Z., Du, X., Xu, Y., Deng, Y., Liu, M., Zhao, Y., ... & Yang, H. (2020). Structure of Mpro from SARS-CoV-2 and discovery of its inhibitors. *Nature*, *582*(7811), 289-293.

Jo, S., Kim, T., Iyer, V. G., & Im, W. (2008). CHARMM-GUI: a web-based graphical user interface for CHARMM. *Journal of computational chemistry*, *29*(11), 1859-1865.

Jorgensen WL, Tirado-Rives J (2005) Potential energy functions for atomic-level simulations of water and organic and biomolecular systems. *Proc Natl Acad Sci USA* 102(19):6665–6670

Joy, S., Nair, P. S., Hariharan, R., & Pillai, M. R. (2006). Detailed comparison of the protein-ligand docking efficiencies of GOLD, a commercial package and ArgusLab, a licensable freeware. *In silico biology*, *6*(6), 601-605.

Kendrew, J. C.; Bodo, G.; Dintzis, H. M.; Parrish, R. G.; Wyckoff, H.; et al (1958). A three-dimensional model of the myoglobin molecule obtained by X-ray analysis. *Nature* 181, 662–666.

Khan FI, Wei DQ, Gu KR, Hassan MI, Tabrez S (2016) Current updates on computer aided protein modeling and designing. *Int J Biol Macromol* 85:48–62

Khandogin, J., and Brooks, C.L., 3rd (2007). Linking folding with aggregation in Alzheimer's beta-amyloid peptides. *Proc. Natl. Acad. Sci. USA* 104, 16880– 16885

Kini RM, Evans HJ (1991) Molecular modeling of proteins: a strategy for energy minimization by molecular mechanics in the AMBER force field. *J Biomol Struct Dyn* 9(3):475–488

Koshland Jr, D. E. (2010) The key–lock theory and the induced fit theory. Angew. *Chem. Int. Ed*., 33, 2375–2378

Kramer, B., Rarey, M., & Lengauer, T. (1999). Evaluation of the FLEXX incremental construction algorithm for protein–ligand docking. *Proteins: Structure, Function, and Bioinformatics*, *37*(2), 228-241.

Krammer, A., Kirchhoff, P. D., Jiang, X., Venkatachalam, C. M., & Waldman, M. (2005). LigScore: a novel scoring function for predicting binding affinities. *Journal of Molecular Graphics and Modelling*, *23*(5), 395-407.

Krauss, G. (2006). *Biochemistry of signal transduction and regulation*. John Wiley & Sons.

Lamarre, D. et al (2003). An NS3 protease inhibitor with antiviral effects in humans infected with hepatitis C virus. Nature 426, 186–189.

lder, B.J., and Wainwright, T.E. (1957). Phase transition for a hard sphere system. *J. Chem. Phys.* 27, 1208–1209.

Levitt, M. (2001). The birth of computational structural biology. *Nature structural biology*, *8*(5), 392-393.

Li, X.; Geng, M.; Peng, Y.; Meng, L.; Lu, S. (2020). Molecular immune pathogenesis and diagnosis of COVID-19. J. *Pharm. Anal*. 19, 1–7.

Lifson, S., and Warshel, A. (1968). Consistent force field for calculations of conformations vibrational spectra and enthalpies of cycloalkane and N-alkane molecules. *J. Chem. Phys.* 49, 5116–5129

Lindahl, E., Hess, B., & Van Der Spoel, D. (2001). GROMACS 3.0: a package for molecular simulation and trajectory analysis. *Molecular modeling annual*, *7*(8), 306-317.

Liu, M., & Wang, S. (1999). MCDOCK: a Monte Carlo simulation approach to the molecular docking problem. *Journal of computer-aided molecular design*, *13*(5), 435-451

M. Bonomi, D. Branduardi, G. Bussi, C. Camilloni, D. Provasi, P. Raiteri, D. Donadio, F. Marinelli, F. Pietrucci, R.A. Broglia and M. Parrinello. (2009). PLUMED: a portable

plugin for free energy calculations with molecular dynamics, *Comp. Phys. Comm*. 180, 1961 , preprint available as arXiv:0902.0874

M. K.; Abraham, D. J. (2005). The enigma of the liganded hemoglobin end state: a novel quaternary structure of human carbonmonoxy hemoglobin. *Biochemistry*, 44, 8347–8359

Mandal, Soma, and Sanat K. Mandal (2009). "Rational drug design." *European journal of pharmacology* 625.1-3. 90-100.

Manglik, A., Lin, H., Aryal, D.K., McCorvy, J.D., Dengler, D., Corder, G., Levit, A., Kling, R.C., Bernat, V., Hubner, H., et al. (2016). Structure-based discovery € of opioid analgesics with reduced side effects. *Nature* 537, 185–190.

McCammon, J.A., Gelin, B.R., and Karplus, M. (1977). Dynamics of folded proteins. *Nature* 267, 585–59

McCorvy, J.D., Butler, K.V., Kelly, B., Rechsteiner, K., Karpiak, J., Betz, R.M., Kormos, B.L., Shoichet, B.K., Dror, R.O., Jin, J., and Roth, B.L. (2018). Structure-inspired design of b-arrestin-biased ligands for aminergic GPCRs. *Nat. Chem.* Biol. 14, 126–134.

Miao, Y., Feher, V. A. & McCammon, J. A. (2015). Gaussian Accelerated Molecular Dynamics: Unconstrained Enhanced Sampling and Free Energy Calculation. Journal of Chemical Theory and Computation 11, 3584-3595

Mobley, D. L., & Gilson, M. K. (2017). Predicting binding free energies: frontiers and benchmarks. *Annual review of biophysics*, *46*, 531.

Moré, J. J., & Wu, Z. (1999). Distance geometry optimization for protein structures. *Journal of Global Optimization*, *15*(3), 219-234.

Morris, G.M.; Goodsell, D.S.; Halliday, R.S.; Huey, R.; Hart, W.E.; Belew, R.K. and Olson, A.J. (1998) J. Comput. *Chem.*, 19(14), 1639-1662.

Morrison, Julie L., et al. (2006). "A lock-and-key model for protein–protein interactions." *Bioinformatics* 22.16. 2012-2019.

Muryshev, A. E., Tarasov, D. N., Butygin, A. V., Butygina, O. Y., Aleksandrov, A. B., & Nikitin, S. M. (2003). A novel scoring function for molecular docking. *Journal of computer-aided molecular design*, *17*(9), 597-605.

Nguyen TT, Viet MH, Li MS (2014) Effects of water models on binding affinity: evidence from all-atom simulation of binding of tamiflu to A/H5N1 neuraminidase. *Sci World J* 2014:536084

Nosé, S. (1984). A molecular dynamics method for simulations in the canonical ensemble. *Molecular physics*, *52*(2), 255-268.

Nosé, S., & Klein, M. L. (1983). Constant pressure molecular dynamics for molecular systems. *Molecular Physics*, *50*(5), 1055-1076.

Oliveira, J. S., Pereira, J. H., Canduri, F., Rodrigues, N. C., de Souza, O. N., de Azevedo Jr, W. F., ... & Santos, D. S. (2006). Crystallographic and pre-steady-state kinetics studies on binding of NADH to wild-type and isoniazid-resistant enoyl-ACP (CoA) reductase enzymes from Mycobacterium tuberculosis. *Journal of molecular biology*, *359*(3), 646-666.

Oostenbrink C, Villa A, Mark AE, van Gunsteren WF (2004) A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6. *J Comput Chem* 25(13):1656–1676

Owen, C. D., et al. "Covid-19 main protease with unliganded active site." PDB. org Code 6Y84 (2020).

Páll, S., Abraham, M. J., Kutzner, C., Hess, B., & Lindahl, E. (2014, April). Tackling exascale software challenges in molecular dynamics simulations with GROMACS. In *International conference on exascale applications and software* (pp. 3-27). Springer, Cham.

Paquet E, Viktor HL (2015) Molecular dynamics, Monte Carlo simulations, and Langevin dynamics: a computational review. *Biomed Res Int* 2015:18

Parrinello, M., & Rahman, A. (1981). Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied physics*, *52*(12), 7182-7190.

Pereira, J. H., De Oliveira, J. S., Canduri, F., Dias, M. V., Palma, M. S., Basso, L. A., ... & De Azevedo, W. F. (2004). Structure of shikimate kinase from Mycobacterium tuberculosis reveals the binding of shikimic acid. *Acta Crystallographica Section D: Biological Crystallography*, *60*(12), 2310-2319.

Perozzo, R., Kuo, M., Valiyaveettil, J. T., Bittman, R., Jacobs, W. R., Fidock, D. A., & Sacchettini, J. C. (2002). Structural Elucidation of the Specificity of the Antibacterial Agent Triclosan for Malarial Enoyl Acyl Carrier Protein Reductase∗. *Journal of Biological Chemistry*, *277*(15), 13106-13114.

Petersen, E., Koopmans, M., Go, U., Hamer, D. H., Petrosillo, N., Castelli, F., ... & Simonsen, L. (2020). Comparing SARS-CoV-2 with SARS-CoV and influenza pandemics. *The Lancet infectious diseases*, *20*(9), e238-e244.

Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L, Schulten K (2005) Scalable molecular dynamics with NAMD. *J Comput Chem 26* (16):1781–1802

Pierce, B., Tong, W., & Weng, Z. (2005). M-ZDOCK: a grid-based approach for C n symmetric multimer docking. *Bioinformatics*, *21*(8), 1472-1478.

Pietrucci F, Laio A. (2009) A collective variable for the efficient exploration of protein beta-sheet structures: application to SH3 and GB1. *J Chem Theory Comput*, 5:2197–2201.

Pillaiyar, T., Manickam, M., Namasivayam, V., Hayashi, Y. & Jung, S. H. (2016). An overview of severe acute respiratory syndrome-coronavirus (SARS-CoV) 3CL protease inhibitors: peptidomimetics and small molecule chemotherapy. *J. Med. Chem*. 59, 6595–6628

Pronk S, Pall S, Schulz R, Larsson P, Bjelkmar P, Apostolov R, Shirts MR, Smith JC, Kasson PM, van der Spoel D, Hess B, Lindahl E (2013) GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* 29(7):845–854

Pronk, S., Páll, S., Schulz, R., Larsson, P., Bjelkmar, P., Apostolov, R., ... & Lindahl, E. (2013). GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics*, *29*(7), 845-854.

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

Salomon-Ferrer R, Gotz AW, Poole D, Le Grand S, Walker RC (2013) Routine microsecond molecular dynamics simulations with AMBER on GPUs. 2. Explicit solvent particle mesh *Ewald. J Chem Theory Comput* 9(9):3878–3888

Sauton, N., Lagorce, D., Villoutreix, B. O., & Miteva, M. A. (2008). MS-DOCK: accurate multiple conformation generator and rigid docking protocol for multi-step virtual ligand screening. *BMC bioinformatics*, *9*(1), 1-12.

Schames, J. R., Henchman, R. H., Siegel, J. S., Sotriffer, C. A., Ni, H., & McCammon, J. A. (2004). Discovery of a novel binding trench in HIV integrase. *Journal of medicinal chemistry*, *47*(8), 1879-1881.

Schulze-Gahmen, U., De Bondt, H. L., & Kim, S. H. (1996). High-resolution crystal structures of human cyclin-dependent kinase 2 with and without ATP: bound waters and natural ligand as guides for inhibitor design. *Journal of medicinal chemistry*, *39*(23), 4540-4546.

Sega M, Autieri E, Pederiva F.(2009) On the calculation of puckering free energy surfaces. *J Chem Phys*, 130:225102.

Shang, J., Ye, G., Shi, K., Wan, Y., Luo, C., Aihara, H., ... & Li, F. (2020). Structural basis of receptor recognition by SARS-CoV-2. *Nature*, *581*(7807), 221-224.

Spahn, V., Del Vecchio, G., Labuz, D., Rodriguez-Gaztelumendi, A., Massaly, N., Temp, J., Durmaz, V., Sabri, P., Reidelbach, M., Machelska, H., et al. (2017). A nontoxic pain killer designed by modeling of pathological receptor conformations. *Science 355*, 966–969.

Sztain, T., Amaro, R., & McCammon, J. A. (2021). Elucidation of cryptic and allosteric pockets within the SARS-CoV-2 main protease. *Journal of chemical information and modeling*, *61*(7), 3495-3501.

Takemoto, M., Kato, H. E., Koyama, M., Ito, J., Kamiya, M., Hayashi, S., ... & Nureki, O. (2015). Molecular dynamics of channelrhodopsin at the early stages of channel opening. *PLoS One*, *10*(6), e0131094.

Tribello, G. A., Bonomi, M., Branduardi, D., Camilloni, C., & Bussi, G. (2014). PLUMED 2: New feathers for an old bird. *Computer physics communications*, *185*(2), 604-613.

Tribello, G. A., Bonomi, M., Branduardi, D., Camilloni, C., & Bussi, G. (2014). PLUMED 2: New feathers for an old bird. *Computer physics communications*, *185*(2), 604-613.

Tsai, C. J., Kumar, S., Ma, B., & Nussinov, R. (1999). Folding funnels, binding funnels, and protein function. *Protein Science*, *8*(6), 1181-1190.

Ullrich, S., & Nitsche, C. (2020). The SARS-CoV-2 main protease as drug target. *Bioorganic & medicinal chemistry letters*, *30*(17), 127377.

Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E., & Berendsen, H. J. (2005). GROMACS: fast, flexible, and free. *Journal of computational chemistry*, *26*(16), 1701-1718.

Velec, H. F., Gohlke, H., & Klebe, G. (2005). DrugScoreCSD knowledge-based scoring function derived from small molecule crystal data with superior recognition rate of near-native ligand poses and better affinity prediction. *Journal of medicinal chemistry*, *48*(20), 6296-6303.

Verdonk, M. L., Cole, J. C., Hartshorn, M. J., Murray, C. W., & Taylor, R. D. (2003). Improved protein–ligand docking using GOLD. *Proteins: Structure, Function, and Bioinformatics*, *52*(4), 609-623.

von Helmholtz, H. (1882). *Die thermodynamik chemischer Vorgänge*.

Wang, R., Lai, L., & Wang, S. (2002). Further development and validation of empirical scoring functions for structure-based binding affinity prediction. *Journal of computer-aided molecular design*, *16*(1), 11-26.

Wang, R., Liu, L., Lai, L., & Tang, Y. (1998). SCORE: A new empirical method for estimating the binding affinity of a protein-ligand complex. *Molecular modeling annual*, *4*(12), 379-394.

Wang, R., Lu, Y., & Wang, S. (2003). Comparative evaluation of 11 scoring functions for molecular docking. *Journal of medicinal chemistry*, *46*(12), 2287-2303..

Wang, R., Lu, Y., Fang, X., & Wang, S. (2004). An extensive test of 14 scoring functions using the PDBbind refined set of 800 protein− ligand complexes. *Journal of chemical information and computer sciences*, *44*(6), 2114-2125.

Watson, H. C. (1969). The stereochemistry of the protein myoglobin. *Prog. Stereochem*, *4*(299), 5.

World Health Organization. Coronavirus disease (COVID-19): situation report, 209. URL https://www.who.int/emergencies/diseases/ novel-coronavirus-2019/situation-reports. Accessed 13.07.2022

Wu, C., and Shea, J.E. (2013). Structural similarities and differences between amyloidogenic and non-amyloidogenic islet amyloid polypeptide (IAPP) sequences and implications for the dual physiological and pathological activities of these peptides. PLoS *Comput. Biol*. 9, e1003211.

Wu, F., Zhao, S., Yu, B., Chen, Y. M., Wang, W., & Song, Z. G. (2020). A novel coronavirus associated with human respiratory disease in China. *Nature*, *579*(7798), 265-269.

Y. Chong, K. Borroto-Esoda, P.A. Furman, R.F. Schinazi, C.K. Chu, (2002). Molecular mechanism of DAPD/DXG against zidovudine- and lamivudine-drug resistant mutants: A molecular modeling approach. Antivir. Chem. Chemother. 13(2), 115–128

Yang, C. Y., Wang, R., & Wang, S. (2006). M-score: a knowledge-based potential scoring function accounting for protein atom mobility. *Journal of medicinal chemistry*, *49*(20), 5903-5911.

Yao, H., Liu, J., Xu, M., Ji, J., Dai, Q., & You, Z. (2022). Discussion on molecular dynamics (MD) simulations of the asphalt materials. *Advances in Colloid and Interface Science*, *299*, 102565.York, D. M., Darden, T.

Zhang, C.-Y., Wei, J.-F. & He, S.-H (2006). Adaptive evolution of the spike gene of SARS coronavirus: changes in positively selected sites in different epidemic groups. *BMC microbiology* 6, 88

Zhang, J., Xie, B., & Hashimoto, K. (2020). Current status of potential therapeutic candidates for the COVID-19 crisis. *Brain, behavior, and immunity*, *87*, 59-73.

Zhang, L., Lin, D., Sun, X., Curth, U., Drosten, C., Sauerhering, L., Becker, S., Rox, K., & Hilgenfeld, R. (2020). Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved a-ketoamide inhibitors. *Science, 368(6489)*, 409–412. https://doi.org/10.1126/science.abb340

Zhao, X., Liu, X., Wang, Y., Chen, Z., Kang, L., Zhang, H., ... & Jiang, H. (2008). An improved PMF scoring function for universally predicting the interactions of a ligand with

protein, DNA, and RNA. *Journal of chemical information and modeling*, *48*(7), 1438-1447.

Zheng, W., Schafer, N. P., Davtyan, A., Papoian, G. A., & Wolynes, P. G. (2012). Predictive energy landscapes for protein–protein association. *Proceedings of the National Academy of Sciences*, *109*(47), 19244-19249.