Molecular Dynamics Simulation to Study Protein Conformation and Ligand Interaction

Santanu Sasidharan, Vijayakumar Gosu, Timir Tripathi, and Prakash Saudagar

Abstract The field of molecular dynamics (MD) simulations has become indispensable today to studying the conformational flexibility and dynamics of proteins as well as protein–ligand complexes. The technique helps to replicate real-time biological events like macromolecular dynamics on a computational platform and allows us to understand the fold and conformational changes in the protein–ligand complex. In addition, MD simulations enable us to estimate the thermodynamics and kinetics associated with protein–ligand binding. In this chapter, we introduce the basics of MD simulations and the theoretical aspects of the simulations. Further, we describe the sequential steps in the process of MD simulation and the background information of the steps. The chapter also discusses ligand binding and conformational changes with the help of case studies. Though the field has advanced by leaps and bounds, there is still a necessity for better force fields and methods to accurately predict the free energy of binding. In summary, research focusing on force fields supported by advancements in computational power will help researchers have better insights into protein–ligand interactions and their conformations.

S. Sasidharan
Department of Biotechnology, National Institute of Technology Warangal, Warangal, Telangana, India

Department of Physics and Astronomy, University of British Columbia, Vancouver, BC, Canada

V. Gosu
Department of Animal Biotechnology, Jeonbuk National University, Jeon Ju, Republic of Korea

T. Tripathi
Molecular and Structural Biophysics Laboratory, Department of Biochemistry, North-Eastern Hill University, Shillong, India

Regional Director's Office, Indira Gandhi National Open University (IGNOU), Regional Centre Kohima, Kohima, India

P. Saudagar (✉)
Department of Biotechnology, National Institute of Technology Warangal, Warangal, Telangana, India
e-mail: ps@nitw.ac.in

P. Saudagar, T. Tripathi (eds.), *Protein Folding Dynamics and Stability*,
https://doi.org/10.1007/978-981-99-2079-2_6

Keywords Protein · Molecular dynamic simulations · Ligand · Dynamics · Conformational flexibility

1 Introduction

Research in biomolecular dynamics has evolved over the last few decades. Among the macromolecules of interest, proteins are essential for the growth and structural integrity of any organism. Their three-dimensional (3D) structure and interactions with other macromolecules or ligands help them function properly. In addition, other molecules interact with proteins either in their active site or allosteric site, which may change the conformations of the protein. The study of these dynamics in detail helps us understand the underlying principles of protein function and interactions [1]. Moreover, advances in bioinformatics and computational power have led researchers to study the structural dynamics of proteins using various simulation algorithms.

Molecular dynamics (MD) simulation is a theoretical method that can analyze the protein structure, folding, and stability by visualizing it in a motion picture. MD simulations have been widely used for studying the complexity of protein folding and the interaction of proteins with ligands. This theoretical study has become an integral part of analyzing the interaction of the ligand with the protein and how the binding of the ligand influences the protein structure, dynamics, and conformation [2, 3]. Besides, it also helps in studying the interactions and changes in terms of energy and geometry over the evolved time period. Today, this method is a boon for protein fold analysis and drug discovery [4].

In principle, MD simulations consider the potential energy function of each of the atoms (force field) and determine the lowest energy state. This means that the state of the most stable conformation, which can be seen over the time period of the simulation run, is determined. Over the past few years, various refinements have been made in the forcefields used for MD simulations. Among the various forcefields used, AMBER, CHARMM, and GROMOS are the most widely used forcefields for studying the structural dynamics of proteins at different pH and temperature conditions [5–7]. These forcefields can be employed in various software like GROMACS, AMBER, and NAMD, and significant information can be obtained from the trajectory analysis [8–10].

To study the effect of the ligand on the protein conformation, one can utilize various techniques like principal component analysis (PCA), coarse-grained simulation, and umbrella sampling. Coarse-grained simulation helps overcome the time-scale and length-scale difference in the ligand–protein interaction by considering the atoms at a macroscopic scale. It does so by reducing the degrees of freedom of the atoms of the protein–ligand complex, providing reduced computational stress, thereby running smoothly. The umbrella sampling depends upon the biasing potential obtained from the mean force potential. It fixes or restrains the ligand toward an increasing center of mass distance via the umbrella sampling. This eventually helps in studying the ligand interaction with the atoms around it over a period of time.

Other than the specific protein–ligand interaction, any perturbation in the protein conformations may result in diseased conditions such as Alzheimer's disease and cancers. Thus, understanding how a protein folds and its dynamics change when interacting with other small molecules and macromolecules is of paramount importance.

2 Background of MD Simulation

We discussed that the MD simulation is a powerful computational method for the theoretical study of biomolecules through fluctuation and conformational changes at the atomic level. The technique uses Newton's second law of motion to calculate the time evolution of the molecular system. The results are obtained in the form of trajectories that are analyzed using different tools for the position and velocity of each atom in the system. In recent times, MD simulation is also being used to understand the thermodynamic properties of biological events like conformational transitions. The technique helps us understand that a protein is flexible and can thus undergo a variety of slow and fast structural rearrangements (also known as transitions), ligand binding, enzymatic regulations, and ion transport in biological systems.

For a protein to function, structural fluctuations and flexibility are very crucial. According to the Levinthal paradox [11, 12], the average time taken would be of the order of 10^{10} years if the process of protein folding was to occur randomly, considering all accessible configurations (around 10^{30} configurations) and a time of 10^{-12} s to search each configuration [4]. The fact that the protein folding process occurs in an immensely shorter time (between picoseconds and milliseconds) proves that the event of protein folding is not a result of a random search toward the correct functional form among the vast configurational space. To explore such configurational spaces, techniques like umbrella sampling have been developed [13].

Before performing MD simulations, it is essential to choose an initial configuration of the proposed system that does not have high potential energy. A velocity must be assigned to the system. To rule out instabilities during simulation, energy minimization is required. Further, a potential energy function (forcefield), which describes the forces that act between the atoms as a function of their positions, is assigned to the system. This gives an initial distribution of the velocities of the atoms and the values of the starting coordinates for the atoms in the system. During the course of the simulation, the trajectories are obtained at different time points and are analyzed. This equilibrium distribution of velocities throughout the system is done via the Maxwell–Boltzmann distribution.

2.1 Theory Behind MD Simulation

It is well known that the MD algorithms calculate the classical time evolution of the system using Newton's second law of motion, i.e.,

$$F_{\mathrm{i}} = m_{\mathrm{i}} a_{\mathrm{i}}$$

$$\frac{F_{\mathrm{i}}}{m_{\mathrm{i}}} = \frac{v_{\mathrm{i}}}{t}$$

$$\frac{F_{\mathrm{i}}}{m_{\mathrm{i}}} = \frac{d^2 x_{\mathrm{i}}}{dt^2}$$

where F_{i} = force exerted on molecule "i," m_{i} = mass of the particle, and a_{i} = acceleration of molecule "i."

The potential energy of the system can be explained as the sum of the individual contribution of both bonded and non-bonded interactions in the system, i.e.,

$$V(r) = V_{\mathrm{bonded}} + V_{\mathrm{nonbonded}}.$$

Bonded interactions are the sum of four simple harmonic species that describes bond stretching and angle bending. It includes all the parameters responsible for bond stretch and angular bending, including rotational torsion and improper torsion [14].

$$V_{\mathrm{bonded}} = V_{\mathrm{bond}} + V_{\mathrm{angle}} + V_{\mathrm{torsion}} + V_{\mathrm{improper}}.$$

V_{bond} represents the energy involved in stretching the bond length in an interaction and can be explained with the help of Morse potential, a robust interatomic interaction model used for the potential energy calculation of a diatomic molecule. Morse potential is computationally expensive and requires three parameters per bond evaluation. Mathematically it can be represented as

$$V_{\mathrm{morse}}(I) = D_{\mathrm{e}}\left[1 - e^{(-a(I-I_0))^2}\right]$$

$$a = \omega\sqrt{\frac{\mu}{2D_{\mathrm{e}}}}$$

$$\omega = \sqrt{\frac{k}{\mu}}$$

where k = stretching constant, D_{e} = depth of potential minimum, l = bond length, l_0 = equilibrium value of bond length, μ = reduced mass, and ω = frequency of bond vibration in small displacement from l to l_0.

However, to overcome the problem of extensive mathematical calculations, a harmonic potential (Hooke's law) was proposed with an approximation that was adequate for explaining bond stretch energy. According to this method

$$V_{\text{bond}}(I) = \frac{k}{2}(I - I_0)^2$$

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

where V_{angle} = energy due to the deviation of angles from their equilibrium values and θ = angle formed between two and three atoms.

V_{torsion} is the torsion angle term in the force field model. It represents the effective barriers for the rotation around chemical bonds. The barriers are due to the steric interactions between the atoms and a group of atoms that are separated by three covalent bonds [15].

$$V_{\text{torsion}}(\varphi) = K_{\varphi}(1 - \cos(n\varphi - \varphi_0)$$

where K_{φ} = barrier height, φ = torsional angle, φ_0 = angular position of the first minimum in the potential, and n = number of minima.

V_{improper} is the improper torsion that arises to maintain chirality.

$$V_{\text{improper}}(\omega) = K\varphi(\omega - \omega_0)^2$$

where ω = improper dihedral angle and ω_0 = improper dihedral angle at equilibrium positions.

The non-bonded interaction is composed of two components, i.e., the van der Waals interaction energy and the electrostatic interaction energy. Energy determination is considered the most time-consuming part of the simulation as they have long-range interactions of the atoms in the system to be considered.

$$V_{(\text{non}-\text{bonded})} = V_{\text{vdw}} + V_{\text{ele}}$$

V_{vdw} arises from a balance between repulsive (short-range and arises due to electron–electron interactions) and attractive forces (long-range force and arises due to electron fluctuations which generate dipole in an atom). It can be demonstrated using Lennard-Jones potential, i.e.,

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

where σ = collision diameter and ϵ = well depth.

V_{ele} act at longer ranges compared to van der Waals interactions. It can be represented as

$$V_{\text{ele}} = \sum_i \sum_j \frac{q_i q_j}{4\pi \in r_{ij}}$$

where q_i, q_j = partial atomic charge, ϵ = dielectric constant, and r_{ij} = relative distance.

Therefore, the potential energy can be represented as

$$V_{(r)} = \sum V_{\text{bond}} + \sum V_{\text{angle}} + \sum V_{\text{torsion}} + \sum V_{\text{vdw}} + \sum V_{\text{ele}}$$

Though the mathematical calculations per atom are higher for a simulation system, the molecular mechanic forcefields provide a reasonable compromise between accuracy and efficiency.

3 Steps in MD Simulation

3.1 Initialization

MD simulations for a biomolecular structure require an initial structure that can be used as a starting point. This structure can be obtained from the X-ray crystallographic or cryo-electron microscopic (cryo-EM) structures available in the protein databank (PDB). Structures from nuclear magnetic resonance (NMR) and homology models can also be used. The selection of the initial structure is critical to obtain better-quality results. Before proceeding toward simulation, energy minimization of the structure is required to eliminate structural distortions that arise due to strong van der Waals interactions and result in unstable simulation. Once the structure is obtained, the next step is to set up the periodic boundary conditions.

3.2 Periodic Boundary Conditions

Defining the periodic boundary is an important step in MD simulation. The step allows one to simulate a small part of a large system specifically. Here, all the atoms present in the computational cell or box (MD cell) are replicated to create an infinite lattice throughout the space. Each particle in the MD cell interacts not only with other particles within the computational box but also with their mirror images in the nearby boxes. Most MD simulations are done in a cubic or octahedral computational cell. The Ewald method is the most common method used for calculating the electrostatic energy of a system on the lattice with periodic boundary conditions. Total electrostatic energy from image cells can be calculated as a summation of real space (V_r), reciprocal space (V_k), correction due to excluding pairs (U_e), and a self-term (U_s) [15].

In MD simulation, constant temperature is essential that can be maintained through coupling to a Berendsen thermal bath. Velocities are scaled by a factor at every step.

$$X = \left(1 + \frac{\delta_t}{\tau}\left(\frac{T}{T_0} - 1\right)\right)$$

where T = the time constant and T_0 is the reference temperature that tells the strength of the coupling between the thermal bath and the system.

3.3 *Energy Minimization*

Minimization algorithms are employed to identify the geometry of the system that corresponds to the minima of the potential energy surface. The minima values can be very large when a biomolecular system comprises thousands of atoms, and a large number of degrees of freedom is taken. The algorithms used for minimization are important in MD because it is essential to start a simulation from a well-minimized structure that helps avoid any high-energy interactions that might hinder the system.

$$\frac{\partial f}{\partial X_i} = 0$$

$$\frac{\partial^2 f}{\partial X_i^2} > 0$$

where $I = [1, \ldots, N]$, given function f, which depends upon the variable x_1, x_2,. . . . .x_n.

The minimum of f is the point at which the first derivative of the function corresponding to each variable is zero, and the second derivatives of the function are all positive. The energy minimization method can be divided into first and second derivative techniques.

3.3.1 First-Derivative Techniques

The first-derivative techniques include the steepest descent and conjugate gradient [16, 17]. The first derivative of energy shows where the local minima lie, and the magnitude gradient indicates the steepness of the local slope. While the second derivative indicates the function's curvature and the information that can be utilized to determine where the function will change with direction. The first-order minimization algorithm is the steepest descent, and in this, the coordinates of the atoms are changed gradually until the system moves close to the minimum energy point. A line search algorithm is used iteratively to locate the minimum point. Even when the

starting initial structure is far away from the minimum, the steepest method can achieve the minimum through iterative steps. Because of this advantage, it is recommended to start with the steepest descent algorithm for energy minimization. The conjugate gradient is another first-order minimization algorithm that accumulates information about the function from one iteration to the next [15].

3.3.2 Second-Derivative Techniques

The Newton-Raphson method is a second-order derivative method used to invert the Hessian matrix for energy minimization [18]. The technique provides the curvature of the function that tells about the change in the direction of the function. For large systems, the technique requires higher computational effort and large storage requirements. In most cases, the steepest descent and the Newton-Raphson method are used in combination. However, in the steepest descent method, the structure can be brought to the minimum closely, while in the Newton-Raphson method, a few steps are required to reach the minimum.

3.4 Thermostats and Barostats

Thermostats and barostats are used for equilibrium. The objective of the equilibrium phase is to perform the simulation until the properties like structure, pressure, temperature, and energy are stable with respect to time and to bring the system to equilibrium from the initial configuration. During this phase, each atom of the system is assigned an initial velocity selected from the Maxwell–Boltzmann distribution at a low temperature. Slowly, new velocities are assigned with a gradient increase in the temperature. This process is repeated until the desired temperature is obtained. The equilibration is usually conducted using a Berendsen thermostat and a Parrinello-Rahman barostat [19, 20].

3.5 Production Stage

After the successful completion of the equilibration of the system, the desired MS simulation time length is assigned between picoseconds (ps) and milliseconds (ms). During the production run, no velocity scaling is performed, and hence the temperature becomes a calculated property. Various properties are computed during the production run and are stored for further analysis. During the production run, millions of non-bonded interactions are generated. Thus, it is necessary to evaluate the non-bonded interactions during simulation. One of the easy ways to do so is by extending the time step, which improves the simulation performance. However, we do not consider the bond vibrations during simulation because errors are generated

immediately after the production run starts in bond vibration. These errors can be excluded entirely by adding bond constraints using SHAKE algorithms [21].

3.6 Analysis of the MD Data

Simulation information generated after the production run can be analyzed in different ways. One of the most important jobs during the analysis of the ligand–protein complex is to determine whether the apoprotein is stable and close to the experimentally retrieved structure or not. The basic method to check the stability and change in conformation is by calculating the root mean square deviation (RMSD), root mean square fluctuation (RMSF), radius of gyration (Rg) and hydrogen bond (H-bond), and principal component analysis (PCA) from the simulation data.

The RMSD is used to measure the structural stability of the protein–ligand complex. It provides information about the deviation produced by the complex during the MD simulation compared to the initial reference structure by calculating the Cα values of the protein backbone. Mathematically, it can be represented as

$$\left\lfloor \left(r_{\mathrm{i}}^{\alpha} - r_{\mathrm{i}}^{\beta} \right) \right\rfloor^{1/2} = \sqrt{\frac{1}{N} \sum_{i} \left(r_{\mathrm{i}}^{\alpha} - r_{\mathrm{i}}^{\beta} \right)^2}$$

The RMSF is used to measure the local changes that are present along the chain of the protein. It is the measure of the displacement of a particular atom or a group of atoms relative to the initial structure used for the simulation and is averaged over the number of atoms in the structure. The calculation involves a rigid alignment of structure in each frame of the simulation run with respect to the reference frame. It is mathematically represented as

$$\mathrm{RMSF} = \sqrt{\frac{1}{N_{\mathrm{f}}} \sum\nolimits_{f} \left(r_{\mathrm{i}}^{f} - r_{\mathrm{i}}^{\mathrm{avg}} \right)^2}$$

The Rg determines the distribution of the atoms present in a protein around the axis of the protein. Rg is given by the length that measures the distance between the point where the atom is rotating and the point where the energy transfers with maximum effect. It is mathematically represented as

$$\mathrm{Rg} = \sqrt{\frac{1}{N_{\mathrm{i}}} \sum\nolimits_{i} (r_{\mathrm{i}} - r_{\mathrm{cm}})^2}$$

Hydrogen bonds are known to play a vital role in ligand binding. They are important for the effective ligand binding and conformational change in the protein's active site. Mathematically, it is calculated as

$$U_{\mathrm{HB}}(r) = \frac{A}{r^{12}} - \frac{C}{r^{10}}$$

The PCA is a machine learning tool that converts a set of correlated observations to a set of linearly independent components. This transformation to the new coordinate system represents the first coordinate with the highest variance, the second coordinate with the second highest variance, and so on. PCA is used to analyze the motion of flexible regions in the protein. Furthermore, it can also be used to analyze the poorly equilibrated regions in a protein. The calculation of PCA involves the following basic steps

1. Creation of coordinate covariance matrix—It is a 3×3 matrix that consists of the coordinates *x*, *y,* and *z* of the sample at different times.
2. Calculation of principal components and coordinate projections—It gives us the eigenvectors of the matrix.
3. Visualization of the principal components.

The molecular mechanics (MM) energies combined with the Poisson–Boltzmann or generalized Born and surface area continuum solvation (MM/PBSA and MM/GBSA, respectively) methods are used to estimate the ligand binding affinities in the simulation run system. They help in deciding the strength of binding of the ligand to its receptor and studying the stability of the complex. The sample is first simulated over a given period of time. Further, snapshots are taken at regular intervals in time from the simulation to calculate the free energy of the sample. For explicit solvation in water, the free energy is mathematically determined as

$$G = E_{\mathrm{int}} + E_{\mathrm{ele}} + E_{\mathrm{vdw}} + G_{\mathrm{pol}} + G_{\mathrm{np}} - TS$$

where E_{int} = molecular mechanics internal energy, E_{ele} = electrostatic internal energy, E_{vdw} = van del Walls energy, G_{pol} = polar solvation free energy, and G_{np} = nonpolar solvation free energy.

Moreover, the binding free energy between the protein and the ligand is mathematically represented as

$$\Delta G = \langle G(PL) - G(P) - G(L) \rangle_{\mathrm{PL}}$$

where *PL*, *P*, and *L* are protein–ligand complexes, protein, and ligand, respectively, whose free energies are calculated using the equation above. Brackets indicate the average over the snapshot taken. Depending on the protein being analyzed, the r^2 value obtained from the correlation coefficients ranges from 0.0 to 0.9.

4 Ligand Binding and Fold Transitions

The first-ever simulation of a protein was conducted as early as 1944 using a small protein bovine pancreatic trypsin inhibitor [22]. From the simulation, McCammon and his team revealed the fluidic characteristics of a protein interior for the first time. The simulation lasted for 9.2 ps and opened up a new realm in molecular biology and drug discovery. Today, the advancements in computational power allow one to perform even microsecond (μs) simulations at the atomistic level.

To start with any ligand binding, the primary requisite is the availability of the target protein structure. This may be a limitation as the number of experimentally determined structures is still less than the number of proteins existing in nature. The problem can partially be solved by its homologous proteins and by predicting the structure using homology modeling [23]. For proteins that do not exhibit any homology, their structures can be predicted de novo using Robetta, I-TASSER, or AlphaFold [24–26]. Once the structure is solved, the ligands can be docked into the rigid or semirigid target structures. The drawback, however, is that the docking process does not consider the flexibility of the target protein, and any critical fold change that occurs cannot be analyzed. To overcome this, all-atom MD simulations can be employed to obtain conformation ensembles of the target protein, which can be used later for ensemble docking. If there is a computational limitation, coarse grain simulation [27, 28] can be done, and representative conformations can be obtained. The atomistic models can then be converted using tools such as backward.py [29]. Once the structure is confirmed, the next step is to perform ensemble docking, where the ligand is docked against each structure of the conformational ensemble. Performing global docking against a conformational ensemble of the target protein with a large dataset of ligands requires computational power. However, the ligand-binding site can be identified using tools such as fpocket [30] and ConCavity [31]. Once the ligand-binding site is predicted based on the geometry of the ligand and the target protein, one can perform the docking more efficiently.

There are several ways to understand the protein–ligand interactions and fold changes with respect to the binding. The most accurate way is to perform all-atom simulations for the ligand–protein complex. However, there are other docking algorithms supported with CHARMM forcefield that employ multiple strategies to obtain better protein–ligand interactions, such as CDOCKER [32], EADock [33], etc. Even though the methods use forcefields, they fail to account for the entropy changes, and therefore the accuracy of the final results in the docking is compromised [34]. Long timescale MD simulations are an easy and effective way to sample the protein–ligand interactions. Long timescale simulations allow determining not only the interactions but also the fold changes that occur in the target protein due to ligand binding. These simulation results enable direct comparison to the experimental results and serve as a benchmark for ligand-binding studies. However, as discussed earlier, they are expensive, and most research groups cannot access them. However, coarse-grained model simulation can work around this problem. The method maps several heavy atoms into one site, reducing the total

number of particles, thereby, the computational power. One can refer to Souza et al. [35] for a better understanding of the concept. The coarse grain simulations fail to provide the desired accuracy in the ligand binding, even if the back mapping is performed. Therefore, to achieve high accuracy and better binding free energy results, it is recommended to run a long timescale simulation of the ligand–protein complex.

Some simulations require the sampling of the conformational space to obtain statistically reliable results. Though this is not reliant on the resolution of the model, it is critical since there are systems with two states with high-energy crossover. To overcome this issue, there are various enhanced sampling methods such as Ligand Gaussian accelerated molecular dynamics simulations [36], metadynamics [37], Markov state models [38, 39], and replica-exchange molecular dynamics [40].

5 Case Studies

In this section, we will discuss the docking and simulation of ligand–protein, how the pipeline works, and the analysis of the simulation results. For this, we will use a study by Sasidharan et al. where natural compounds were virtually screened against the tyrosine aminotransferase (TAT) from *Leishmania donovani* [41]. The initial part of the study concentrated on the virtual screening of 1,83,659 compounds from the ZINC15 database with the protein. The top 10 compounds were then docked independently against TAT using Autodock v4.2. For the docking, authors framed the grid around the active site of the TAT enzyme housing the K286 residue, and 500 LGA docks were conducted to obtain the best-docked conformation. The top 5 compounds with the highest binding affinity and interactions (Fig. 1) with the active site cavities were chosen and carried forward for simulations.

The simulations were carried out using GROMACS v5.1.4. The complexes were energy minimized by steeped descent method and were temperature and pressure equilibrated using a modified Berendsen thermostat and Parrinello–Rahman barostat, respectively. The electrostatic interactions were computed with the help of particle mesh Ewald. The trajectory analysis showed that all complexes with the protein were stable throughout the simulation period. The RMSD of the Cα backbone (Fig. 2a) showed the stability of the complexes, while the Rg (Fig. 2c), along with solvent accessible surface area (Fig. 2d), corroborated the stability of the complexes. The RMSF analysis showed higher fluctuations in the N-terminal (Fig. 2b), and the reason for the same was deciphered by the authors in another study [42]. The study then concentrated on the binding of the ligands to the TAT. An average of 1–3 hydrogen bonds formed between the compounds and the protein (Fig. 3a). The authors eliminated the compound TI 2 from further studies owing to the presence of several metastable states (Fig. 3b). The simulation data showed that the compounds TI 1, TI 3, TI 4, and TI 5 could bind to TAT with high affinity. The authors proved the inhibitory activity of the compounds by in vitro inhibition kinetics.

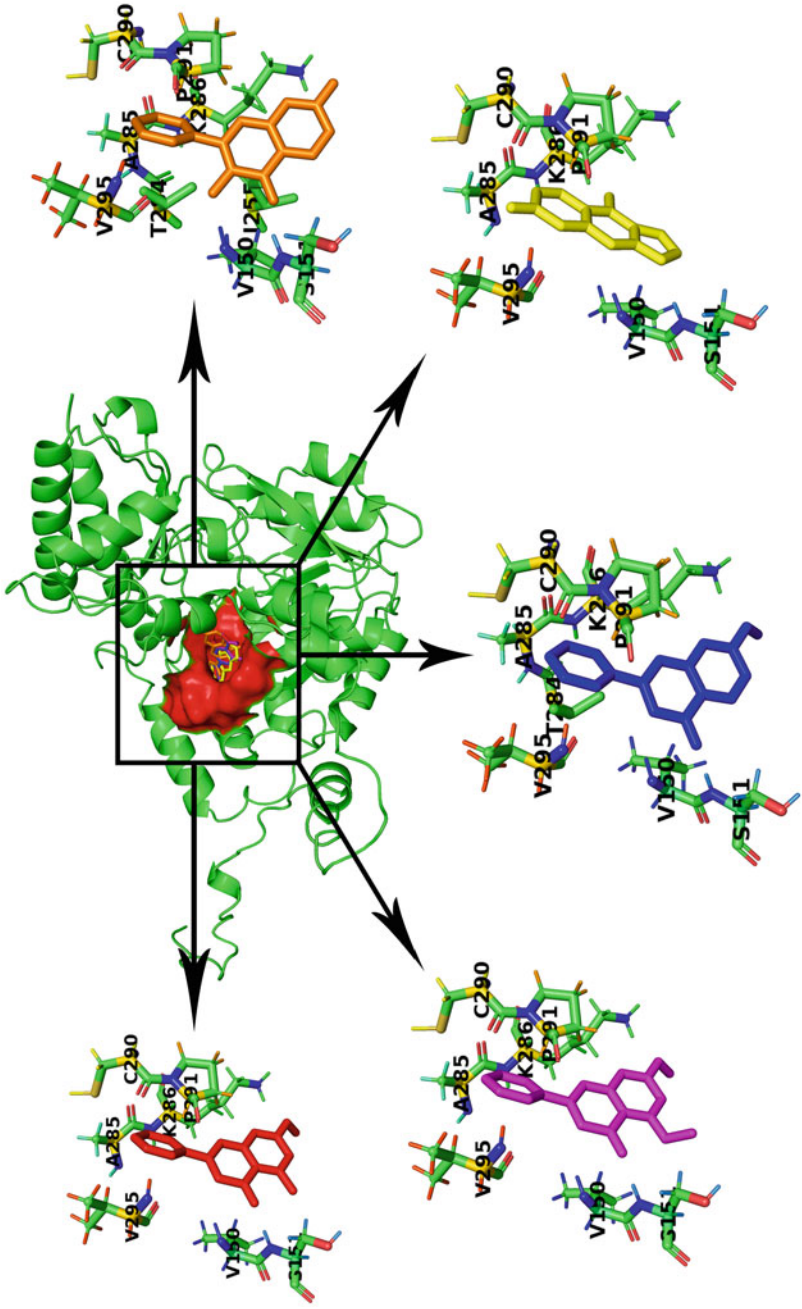

Fig. 1 Docking of the five compounds in the TAT active site. Compounds TI 1 (in red), TI 2 (in magenta), TI 3 (in blue), TI 4 (in yellow), and TI 5 (in orange) docked to the active site cavity (represented as red surface). (Figure adapted with permission from Sasidharan and Saudagar [41])

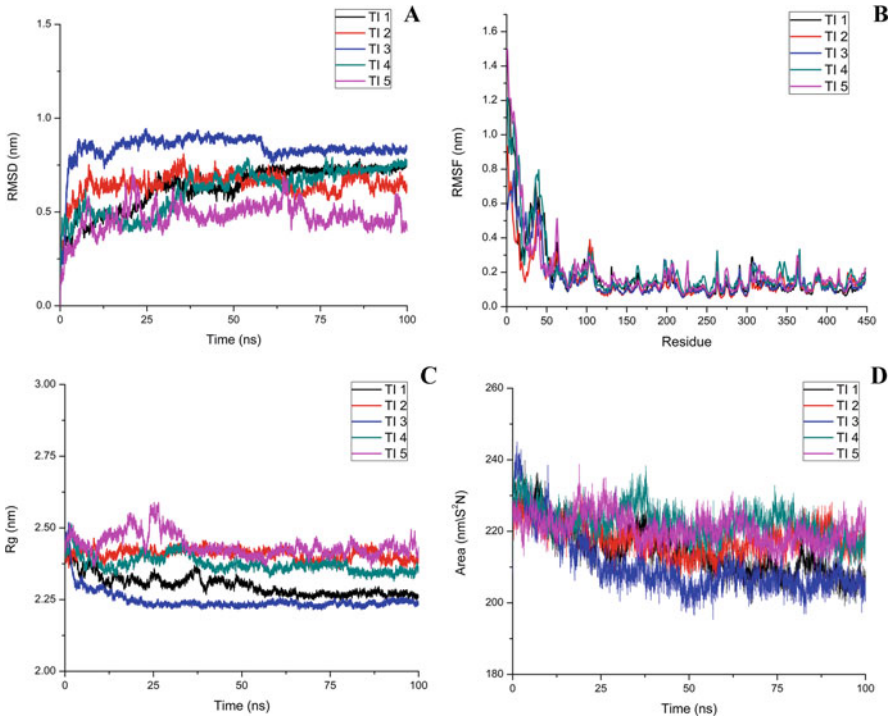

Fig. 2 Trajectory analysis of compounds TI 1–5 complexes with the TAT enzyme. (**a**) RMSD, (**b**) RMSF, (**c**) Rg, (**d**) SAS. The analysis of all four trajectories shows the stability of the TI complexes. (Figure adapted with permission from Sasidharan and Saudagar [41])

To understand the concept of PCA and MMPBSA, we use a study by Shweta et al. [43]. Here, the authors followed a similar protocol and simulated the top two protein–compound complexes. Besides RMSD, Rg, SASA, and RMSF, the authors also conducted the PCA (Fig. 4), which showed that the ligand-bound forms are more rigid than the apo-form. The changes in the large motions were limited upon binding to the ligands chrysin and genistein. Furthermore, MMPBSA calculations showed binding energies of −78 kJ/mol and −76 kJ/mol for genistein and chrysin, respectively, which were higher than the control ATP (−54 kJ/mol). The breakdown of the binding energy is given in Table 1. The binding energy contributed by each residue in the target protein using the MMPBSA tool can also be studied [44]. Several such studies can be referred to understand the protein–ligand binding analysis using MD simulations [45–50].

The studies of protein–ligand interactions and transitions are not limited to small compounds but also protein–macromolecule interactions. Gosu et al. studied the effect of mutations on the MDA5 protein responsible for Aicardi-Goutières syndrome and Singleton-Merten syndrome [51]. The effect of mutation of residues like L372F, A45T, R779H, and R822Q was studied, and the interactions of the mutated proteins with RNA were analyzed. The authors represented the PCA of the simulated

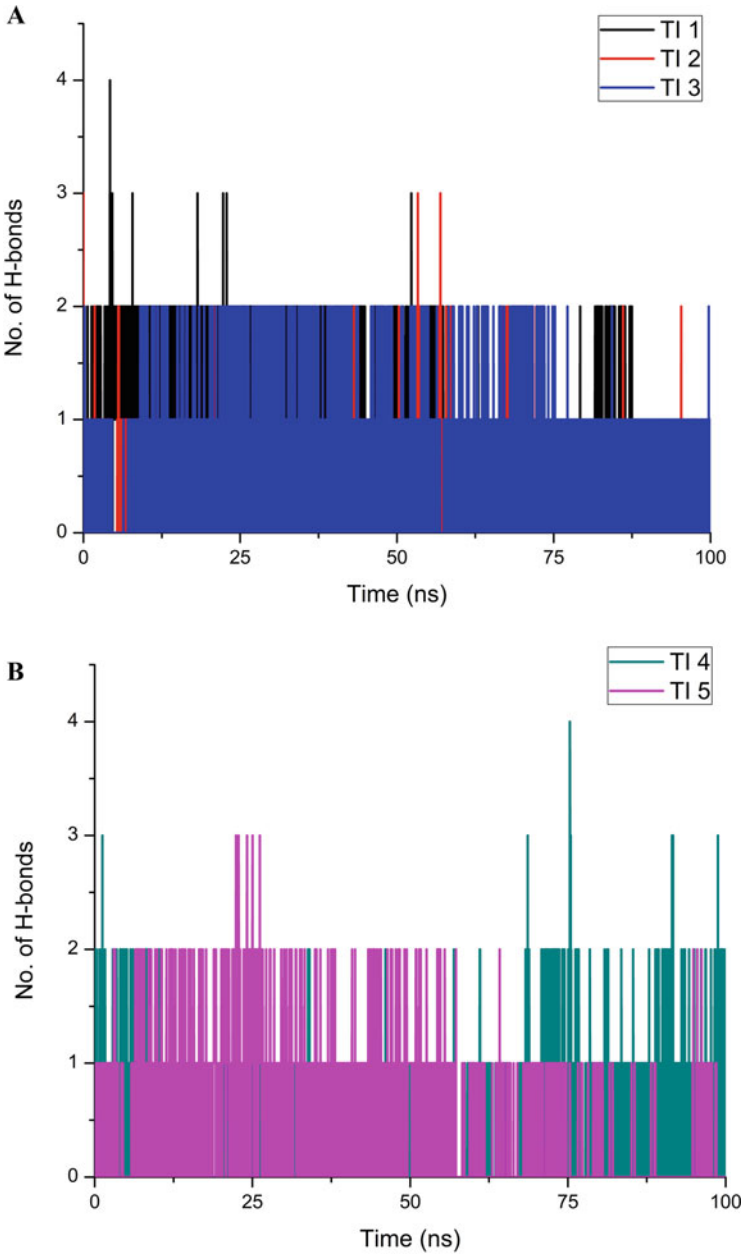

Fig. 3 Hydrogen bond analysis between TI 1–5 compounds and TAT enzyme. (**a**) H-bond analysis of TI 1–3 with TAT. (**b**) H-bond analysis of TI 4 and 5 with TAT. (Figure adapted with permission from Sasidharan and Saudagar [41])

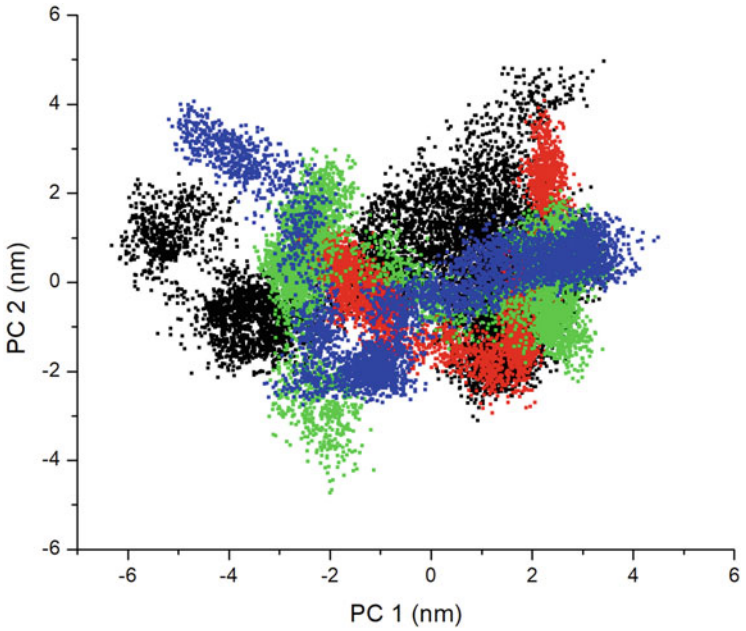

Fig. 4 Principal component analysis of large motions in the simulated structures. Apo-LdMPK4, ATP complex, GEN complex, and CHY complex with LdMPK4 are shown in black, red, green, and blue, respectively. (Figure adapted with permission from Shweta et al. [43])

Table 1 Binding free energy of MAPK4 with ligands ATP, GEN, and CHY

Component	Ligand		
	ATP	GEN	CHY
E_{vdW} (kJ/Mol)	−103.47	−104.923	−105.592
E_{elec} (kJ/Mol)	−3.588	−10.134	−11.382
G_{polar} (kJ/Mol)	60.086	46.500	49.818
$G_{non\text{-}polar}$ (kJ/Mol)	−7.973	−9.653	−9.007
ΔG_{bind} (kJ/Mol)	−54.946	−78.211	−76.164

complexes in porcupine plots (Fig. 5) that revealed the effect of mutations on the large-scale motions of the whole protein as well as the fold changes occurring over the simulation period due to mutations. Hence, MD simulations can be used to study both protein–ligand and protein–macromolecule interactions for both drug discovery and mutation effects [52–57].

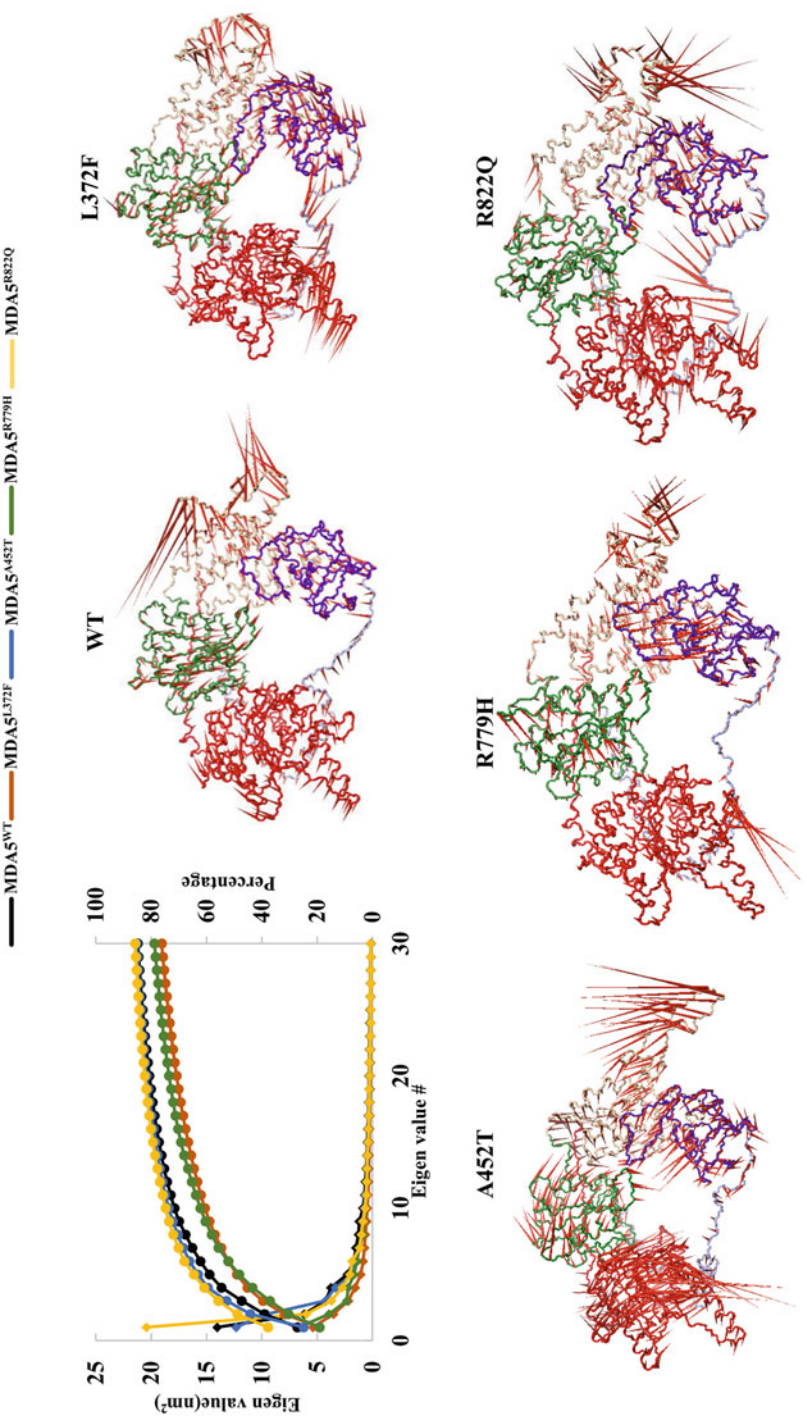

Fig. 5 Cumulative percentages and porcupine plots representing the principal component analysis. The porcupine plots were made using PyMol. (Figure adapted with permission from Gosu et al. [51])

6 Conclusions

Currently, the field of simulation is making an enormous impact in understanding the atomistic details of macromolecules. MD simulations now help researchers to drive the wet-lab experiments based on the simulation data. A detailed conformational understanding of a macromolecule explains the dynamics of the ligand binding as well as the fold transitions that accompany the ligand binding. The accuracy and free energy calculations are more accurate than the docking scores and, therefore, can be relied upon. Though the dynamics of proteins happen at msec timescale in real time, it is not possible, at least at the moment, to simulate all the proteins to that extent. Meanwhile, it is challenging to understand the dynamics using wet-lab experiments. Therefore, researchers must balance these techniques and consider the trade-off to achieve the best possible results. This research area is advancing day by day, with improvements in algorithms and forcefields. With increased computational power and refined algorithms, scientists hope to make simulations widely available at lower costs.

References

1. D.B. Singh, T. Tripathi, *Frontiers in Protein Structure, Function, and Dynamics* (Springer Nature, Singapore, 2020)
2. P. Saudagar, T. Tripathi, *Advanced Spectroscopic Methods to Study Biomolecular Structure and Dynamics*, 1st edn. (Academic Press, San Diego, 2023)
3. T. Tripathi, V.K. Dubey, *Advances in Protein Molecular and Structural Biology Methods*, 1st edn. (Academic Press, Cambridge, MA, 2022)
4. S. Sasidharan, P. Saudagar, Prediction, validation, and analysis of protein structures: a beginner's guide, in *Advances in Protein Molecular and Structural Biology Methods*, ed. by T. Tripathi, V.K. Dubey, (Academic Press, Cambridge, MA, 2022), pp. 373–385
5. J.W. Ponder, D.A. Case, Force fields for protein simulations. Adv. Protein Chem. **66**, 27–85 (2003)
6. J. Lee, M. Hitzenberger, M. Rieger, N.R. Kern, M. Zacharias, W. Im, CHARMM-GUI supports the Amber force fields. J. Chem. Phys. **153**(3), 035103 (2020)
7. W.F. van Gunsteren, X. Daura, A.E. Mark, GROMOS force field, in *Encyclopedia of Computational Chemistry*, vol. 2, (Wiley, Chichester, 2002)
8. D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A.E. Mark, H.J. Berendsen, GROMACS: fast, flexible, and free. J. Comput. Chem. **26**(16), 1701–1718 (2005)
9. D.A. Case, T.E. Cheatham III, T. Darden, H. Gohlke, R. Luo, K.M. Merz Jr., A. Onufriev, C. Simmerling, B. Wang, R.J. Woods, The Amber biomolecular simulation programs. J. Comput. Chem. **26**(16), 1668–1688 (2005)
10. J.C. Phillips, D.J. Hardy, J.D. Maia, J.E. Stone, J.V. Ribeiro, R.C. Bernardi, R. Buch, G. Fiorin, J. Hénin, W. Jiang, Scalable molecular dynamics on CPU and GPU architectures with NAMD. J. Chem. Phys. **153**(4), 044130 (2020)
11. C. Levinthal, Are there pathways for protein folding? J. Chim. Phys. **65**, 44–45 (1968)
12. C. Levinthal, How to fold graciously. Mossbauer Spectrosc. Biol. Syst. **67**, 22–24 (1969)
13. R. Shukla, T. Tripathi, Molecular dynamics simulation in drug discovery: opportunities and challenges, in *Innovations and Implementations of Computer Aided Drug Discovery Strategies in Rational Drug Design*, (Springer, Singapore, 2021), pp. 295–316

14. C.-E.A. Chang, Y.-M.M. Huang, L.J. Mueller, W. You, Investigation of structural dynamics of enzymes and protonation states of substrates using computational tools. Catalysts **6**(6), 82 (2016)
15. A. Kukol, *Molecular Modeling of Proteins* (Springer, New York, 2008)
16. P. Debye, Näherungsformeln für die Zylinderfunktionen für große Werte des Arguments und unbeschränkt veränderliche Werte des Index. Math. Ann. **67**(4), 535–558 (1909)
17. E. Stiefel, Methods of conjugate gradients for solving linear systems. J. Res. Nat. Bureau Standards **49**, 409–435 (1952)
18. J. Dedieu, Newton-Raphson method, in *Encyclopedia of Applied and Computational Mathematics*, ed. by B. Engquist, (Springer, Berlin, 2015), pp. 1023–1028
19. M. Parrinello, A. Rahman, Strain fluctuations and elastic constants. J. Chem. Phys. **76**(5), 2662–2666 (1982)
20. H.J. Berendsen, J.V. Postma, W.F. Van Gunsteren, A. DiNola, J.R. Haak, Molecular dynamics with coupling to an external bath. J. Chem. Phys. **81**(8), 3684–3690 (1984)
21. J.-P. Ryckaert, G. Ciccotti, H.J. Berendsen, Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. J. Comput. Phys. **23**(3), 327–341 (1977)
22. J.A. McCammon, B.R. Gelin, M. Karplus, Dynamics of folded proteins. Nature **267**(5612), 585–590 (1977)
23. N. Eswar, B. Webb, M.A. Marti-Renom, M. Madhusudhan, D. Eramian, M.Y. Shen, U. Pieper, A. Sali, Comparative protein structure modeling using modeller. Curr. Protoc. Bioinformatics **15**(1), 5.6.1–5.6.30 (2006)
24. D.E. Kim, D. Chivian, D. Baker, Protein structure prediction and analysis using the Robetta server. Nucleic Acids Res. **32**(Suppl_2), W526–W531 (2004)
25. J. Yang, R. Yan, A. Roy, D. Xu, J. Poisson, Y. Zhang, The I-TASSER suite: protein structure and function prediction. Nat. Methods **12**(1), 7–8 (2015)
26. J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, Highly accurate protein structure prediction with AlphaFold. Nature **596**(7873), 583–589 (2021)
27. S.J. Marrink, H.J. Risselada, S. Yefimov, D.P. Tieleman, A.H. De Vries, The MARTINI force field: coarse grained model for biomolecular simulations. J. Phys. Chem. B **111**(27), 7812–7824 (2007)
28. D.H. de Jong, G. Singh, W.D. Bennett, C. Arnarez, T.A. Wassenaar, L.V. Schafer, X. Periole, D.P. Tieleman, S.J. Marrink, Improved parameters for the martini coarse-grained protein force field. J. Chem. Theory Comput. **9**(1), 687–697 (2013)
29. T.A. Wassenaar, K. Pluhackova, R.A. Böckmann, S.J. Marrink, D.P. Tieleman, Going backward: a flexible geometric approach to reverse transformation from coarse grained to atomistic models. J. Chem. Theory Comput. **10**(2), 676–690 (2014)
30. P. Schmidtke, V. Le Guilloux, J. Maupetit, P. Tufféry, Fpocket: online tools for protein ensemble pocket detection and tracking. Nucleic Acids Res. **38**(Suppl_2), W582–W589 (2010)
31. J.A. Capra, R.A. Laskowski, J.M. Thornton, M. Singh, T.A. Funkhouser, Predicting protein ligand binding sites by combining evolutionary sequence conservation and 3D structure. PLoS Comp. Biol. **5**(12), e1000585 (2009)
32. G. Wu, D.H. Robertson, C.L. Brooks III, M. Vieth, Detailed analysis of grid-based molecular docking: a case study of CDOCKER—a CHARMm-based MD docking algorithm. J. Comput. Chem. **24**(13), 1549–1562 (2003)
33. A. Grosdidier, V. Zoete, O. Michielin, EADock: docking of small molecules into protein active sites with a multiobjective evolutionary optimization. Proteins **67**(4), 1010–1025 (2007)
34. S. Yin, L. Biedermannova, J. Vondrasek, N.V. Dokholyan, MedusaScore: an accurate force field-based scoring function for virtual drug screening. J. Chem. Inf. Model. **48**(8), 1656–1662 (2008)

35. P.C. Souza, S. Thallmair, P. Conflitti, C. Ramírez-Palacios, R. Alessandri, S. Raniolo, V. Limongelli, S.J. Marrink, Protein–ligand binding with the coarse-grained martini model. Nat. Commun. **11**(1), 1–11 (2020)
36. Y. Miao, A. Bhattarai, J. Wang, Ligand Gaussian accelerated molecular dynamics (LiGaMD): characterization of ligand binding thermodynamics and kinetics. J. Chem. Theory Comput. **16**(9), 5526–5547 (2020)
37. A. Barducci, M. Bonomi, M. Parrinello, Metadynamics. Wiley Interdiscip. Rev. Comput. Mol. Sci. **1**(5), 826–843 (2011)
38. W. Wang, S. Cao, L. Zhu, X. Huang, Constructing Markov state models to elucidate the functional conformational changes of complex biomolecules. Wiley Interdiscip. Rev. Comput. Mol. Sci. **8**(1), e1343 (2018)
39. B.E. Husic, V.S. Pande, Markov state models: from an art to a science. J. Am. Chem. Soc. **140**(7), 2386–2396 (2018)
40. L.S. Stelzl, G. Hummer, Kinetics from replica exchange molecular dynamics simulations. J. Chem. Theory Comput. **13**(8), 3927–3935 (2017)
41. S. Sasidharan, P. Saudagar, Flavones reversibly inhibit Leishmania donovani tyrosine aminotransferase by binding to the catalytic pocket: an integrated in silico-in vitro approach. Int. J. Biol. Macromol. **164**, 2987–3004 (2020)
42. S. Sasidharan, P. Saudagar, Mapping N-and C-terminals of Leishmania donovani tyrosine aminotransferase by gene truncation strategy: a functional study using in vitro and in silico approaches. Sci. Rep. **10**(1), 1–15 (2020)
43. S. Raj, S. Sasidharan, V.K. Dubey, P. Saudagar, Identification of lead molecules against potential drug target protein MAPK4 from L. Donovani: an in-silico approach using docking, molecular dynamics and binding free energy calculation. PLoS One **14**(8), e0221331 (2019)
44. R. Kumari, R. Kumar, A. Lynn, g_mmpbsa—a GROMACS tool for high-throughput MM-PBSA calculations. J. Chem. Inf. Model. **54**(7), 1951–1962 (2014)
45. R. Shukla, P.B. Chetri, A. Sonkar, M.Y. Pakharukova, V.A. Mordvinov, T. Tripathi, Identification of novel natural inhibitors of opisthorchis felineus cytochrome P450 using structure-based screening and molecular dynamic simulation. J. Biomol. Struct. Dyn. **36**(13), 3541–3556 (2018)
46. R. Shukla, H. Shukla, P. Kalita, A. Sonkar, T. Pandey, D.B. Singh, A. Kumar, T. Tripathi, Identification of potential inhibitors of Fasciola gigantica thioredoxin1: computational screening, molecular dynamics simulation, and binding free energy studies. J. Biomol. Struct. Dyn. **36**(8), 2147–2162 (2018)
47. R. Shukla, H. Shukla, P. Kalita, T. Tripathi, Structural insights into natural compounds as inhibitors of Fasciola gigantica thioredoxin glutathione reductase. J. Cell. Biochem. **119**(4), 3067–3080 (2018)
48. R. Shukla, H. Shukla, A. Sonkar, T. Pandey, T. Tripathi, Structure-based screening and molecular dynamics simulations offer novel natural compounds as potential inhibitors of mycobacterium tuberculosis isocitrate lyase. J. Biomol. Struct. Dyn. **36**(8), 2045–2057 (2018)
49. R. Shukla, H. Shukla, T. Tripathi, Structural and energetic understanding of novel natural inhibitors of mycobacterium tuberculosis malate synthase. J. Cell. Biochem. **120**(2), 2469–2482 (2019)
50. R. Shukla, H. Shukla, T. Tripathi, Structure-based discovery of phenyl-diketo acids derivatives as mycobacterium tuberculosis malate synthase inhibitors. J. Biomol. Struct. Dyn. **39**(8), 2945–2958 (2021)
51. V. Gosu, S. Sasidharan, P. Saudagar, H.-K. Lee, D. Shin, Computational insights into the structural dynamics of MDA5 variants associated with Aicardi–Goutières syndrome and Singleton–Merten syndrome. Biomol. Ther. **11**(8), 1251 (2021)
52. R. Shukla, T. Tripathi, Molecular dynamics simulation of protein and protein-ligand complexes, in *Computer-Aided Drug Design*, ed. by D.B. Singh, (Springer Nature, Singapore, 2020), pp. 133–161

53. K. Prince, S. Sasidharan, N. Nag, T. Tripathi, P. Saudagar, Integration of spectroscopic and computational data to analyze protein structure, function, folding, and dynamics, in *Advanced Spectroscopic Methods to Study Biomolecular Structure and Dynamics*, ed. by P. Saudagar, T. Tripathi, (Academic Press, San Diego, 2023), pp. 483–502
54. J. Kalita, H. Shukla, T. Tripathi, Engineering glutathione S-transferase with a point mutation at conserved F136 residue increases the xenobiotic-metabolizing activity. Int. J. Biol. Macromol. **163**, 1117–1126 (2020)
55. P. Kalita, H. Shukla, K.C. Das, T. Tripathi, Conserved Arg451 residue is critical for maintaining the stability and activity of thioredoxin glutathione reductase. Arch. Biochem. Biophys. **674**, 108098 (2019)
56. R. Shukla, H. Shukla, T. Tripathi, Activity loss by H46A mutation in mycobacterium tuberculosis isocitrate lyase is due to decrease in structural plasticity and collective motions of the active site. Tuberculosis (Edinb.) **108**, 143–150 (2018)
57. A. Sonkar, D.L. Lyngdoh, R. Shukla, H. Shukla, T. Tripathi, S. Ahmed, Point mutation A394E in the central intrinsic disordered region of Rna14 leads to chromosomal instability in fission yeast. Int. J. Biol. Macromol. **119**, 785–791 (2018)