



Molecular Dynamics Simulation of Protein and Protein–Ligand Complexes

7

Rohit Shukla and Timir Tripathi

Abstract

Biomacromolecules, including proteins and their complexes, adopt multiple conformations that are linked to their biological functions. Though some of the structural heterogeneity can be studied by methods like X-ray crystallography, NMR, or cryo-electron microscopy, these methods fail to explain the detailed conformational transitions and dynamics. The dynamic structural states in proteins are covered in magnitude between 10^{-11} and 10^{-6} m and time-scales from 10^{-12} s to 10^{-5} s. For a comprehensive analysis of the biomolecular dynamics, molecular dynamics (MD) simulation has evolved as the most powerful technique. With the advent of high-end computational power, MD simulations can be performed between μ s to the ms time-scale that can accurately describe the dynamics of any system. Various force fields like GROMOS, AMBER, and CHARMM have been developed for MD simulations. Tools like GROMACS, AMBER, CHARMM-GUI, and NAMD are the most widely used methods for MD simulation that can provide precise information on the motions and flexibility of a protein, which contributes to the interaction dynamics of protein–ligand complexes. MD simulation has several other practical applications in diverse research areas, including molecular docking and drug design, refining protein structure predictions, and studying the unfolding

R. Shukla

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

Department of Biotechnology and Bioinformatics, Jaypee University of Information Technology (JUIT), Solan, India

T. Tripathi (✉)

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

pathway of a protein. Combining MD simulation with wet-lab experiments has become an indispensable complement in the investigation of several important and intricate biological processes. Various tools like principal component analysis, cross-correlation analysis, and residues interaction network analysis are additional useful approaches for analyzing MD data. In this chapter, we will discuss MD simulation for a layman understanding and explain how it can be used for protein–ligand characterization as well as for use in diverse biomolecular applications.

Keywords

Molecular dynamics · Principal component analysis · Cross-correlation analysis · Virtual screening · Mutational analysis · Structure-function relation · Protein unfolding · Force fields

7.1 History and Background

Proteins perform a wide range of cellular functions in living organisms, such as catalysis of metabolic reactions, transport, and cell signaling; all those depend on the structure and dynamics of the protein. Several non-covalent and covalent interactions help to stabilize the native conformation of protein that dictates its function (Singh and Tripathi 2020). In practice, the degree of folded nature is generally determined by wet-lab experiments, including fluorescence and circular dichroism studies. All-atom molecular dynamics (MD) simulation has been developed as a new tool to understand the dynamics of protein motions at the atomic level. It provides information about the motion of an individual atom as a function of time, and thus describes the dynamic behavior of a molecule. The advantage of MD simulation is that it provides information about the folding/unfolding mechanism like the final folded structure, the time dependency of these events, and the inter-residue interactions. The pioneers of MD simulation were Alder and Wainwright, who introduced this technique in the late 1950s to study the interactions of a hard-sphere (Alder and Wainwright 1957, 1959). In 1964, first simulation using the realistic potential for liquid argon was carried out by Rahman (Rahman 1964). The first realistic system (liquid water) was done in the 1970s to perform simulation (Rahman and Stillinger 1971). However, the first simulation of protein was conducted in 1977 (McCammon et al. 1977). Soon in the 1980s, simulations on protein interacting with small molecules, their thermodynamics (free energy calculations), and rapid calculations of biomolecules were developed (McCammon et al. 1986). In 1998, Duan and Kollman revealed the folding mechanism of the small sub-domain of villin using a μs simulation (Duan and Kollman 1998). In 2013, the Nobel Prize in Chemistry was awarded to Martin Karplus, Michael Levitt, and Arieh Warshel for the development of multi-scale models for complex chemical systems, a technique to simulate the behavior of molecules at various scales from single molecules to proteins.

MD simulation is an emerging field. We entered the keyword “molecular dynamics simulation” to NCBI PubMed database (<https://www.ncbi.nlm.nih.gov/pubmed/>) resulted in 56,833 articles as of 31/07/2020, suggesting the growing importance of MD simulation. Development of techniques such as potential sampling methods, force field advancement, and high-end computational power is allowing us to perform simulations in a range of μs to ms time-scale. MD simulation can thus be highly useful in the study of biomolecular dynamics. However, the use of MD simulation requires optimal models that can mimic the cellular environment, physical applications that can provide the motions to the model, and large-scale computations. However, recently MD simulations have been extended to cellular scales, and simulations of an entire cell have been performed (Heidari et al. 2016). The development of a more robust algorithm and theory for modeling, docking, scoring, energy-calculations will make the MD simulation more effective. In this chapter, we will discuss the principle, methods, tools, and important applications of MD simulation.

7.2 Introduction

To date, it is not possible to accurately predict detailed biomolecular conformational dynamics *in vitro*. Techniques like X-ray crystallography, nuclear magnetic resonance (NMR), and recent cryo-electron microscopy (cryo-EM) methods have provided breakthroughs in structural biology. Still, a vast gap exists between the numbers of the available protein sequences and protein structures. There are 177,754,527 protein sequences in the latest release of UniProtKB as of 12/04/2020, while the protein data bank (PDB) has only 1,62,259 protein structures on 12/04/2020, suggesting only a small fraction of the total sequences have known structures. The PDB statistics, as on 12/04/2020, are shown in Table 7.1. Thus, the prediction of protein structures is essential to fill this significant gap.

Most wet-lab experimental methods provide structural information of proteins in static form, while practically proteins are highly dynamic (Dror et al. 2012). Molecular docking only provides a static pose, and it cannot illustrate the dynamics of the protein–ligand complex (Kalita et al. 2018b; Mamgain et al. 2015). In addition to the prediction of the dynamic behavior of biological systems, MD simulations can also help to explore the kinetic behavior and assemblies of molecules at the atomic level

Table 7.1 The PDB statistics as on dated 12/04/2020

Experimental method	Proteins	Nucleic acids	Protein/NA complex	Others	Total
X-Ray	135,436	2044	6562	460	144,502
NMR	11,344	1284	264	49	12,941
Electron microscopy	3638	35	1029	114	4816
Other	32	1	0	4	37
Hybrid methods	155	5	3	1	164
Total	150,605	3369	7858	628	162,460

(Alder and Wainwright 1959; Rajendran et al. 2018). The conformational dynamics in proteins cover large ranges in both magnitude and time-scale (Vogeli et al. 2012), and due to the conformational changes proteins can function in a variety of ways including acting as transporters, signaling molecules, sensors, and mechanical effectors and also interact with the substrate, drugs, and hormones through conformational changes (Sonkar et al. 2017; Pandey et al. 2017). Structural dynamics in protein conformations are fast, covering a magnitude from 10^{-11} to 10^{-6} m, within of time-scale between 10^{-12} s and 10^5 s (Boehr et al. 2006; Wolf and Kirschner 2013; Krishnamoorthy 2012). The local motions: loop and side-chain motions take 10^{-15} s to 10^{-1} s to complete, while the helix, domain (hinge bending), and subunit motions take 10^{-9} s to 1 s. Processes, such as helix-coil transitions, association/dissociation, and folding/unfolding, come under large-scale motions and may take 10^{-7} s to 10^4 s to complete. Practically, it is challenging to study such changes as they take place in a very short time, however, but by resembling the in vivo conditions computationally, these processes can be examined, visualized, and analyzed (Shukla et al. 2018a, d).

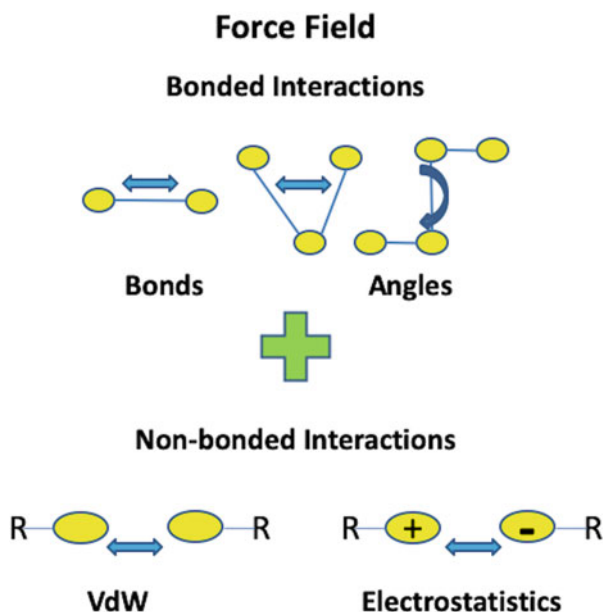
In MD simulation, a molecule can be understood as a series of charged points (atoms) linked by springs (bonds). Now, to describe the time evolution of bond lengths, bond angles, torsions, and also the non-bonding interactions between atoms, the force field is used. In MD simulation, an in-vivo like environment is created using protein and water molecules, and the atoms of protein and water move with a short time step (in the fs time duration), where forces of every atom are computed, and written in a file using a force field. This force field is a collection of equations and associated constants and includes the potential energy functions with bonded and non-bonded potential terms. It reproduces molecular geometry and selected properties of test structures. Cammon et al. performed the first MD simulation of biological macromolecules in 1977 at 9.2 ps for bovine pancreatic trypsin inhibitors. Before advances in MD simulations, experimental observations like hydrogen bond exchange were already studied (Berger and Linderstrom-Lang 1957). The role of thermal factor (B) in internal motions of proteins was investigated (Jeremy Smith et al. 1986; Brunger et al. 1985; Brooks and Karplus 1983) by that time. From the MD simulation data, we can also deduce and calculate principal components (PCs) and perform its analysis (Wolf and Kirschner 2013; David and Jacobs 2014). The calculated PCs are arranged according to their contribution to the total fluctuation along with the ensemble of conformations. The global and correlated motions can be predicted using the computed data. MD simulation also provides information on the conformational flexibility of macromolecules as well as in understanding the experimental results, such as the analysis of fluorescence depolarization (Frauenfelder et al. 1987), dynamics of NMR parameters (Brunger et al. 1987), and effect of solvent and temperature on protein stability (Nilsson et al. 1986; Colonna-Cesari et al. 1986). The simulated annealing is a widely used method for the refinement of X-ray structures (Harvey et al. 1984) and the determination of NMR structure (Case and Karplus 1979).

7.3 Principle of MD Simulation

The underlying principle of MD simulation is based on Newton's law for molecular mechanics (Adcock and McCammon 2006). For the computational study of biomolecular dynamics, MD simulation is the most important established technique (Adcock and McCammon 2006; Levitt and Warshel 1975; Karplus and Kuriyan 2005). In MD simulation, the interaction between the atoms and the molecules is examined for a time period by approximations of known physical attributes (Levitt and Warshel 1975; Karplus and Kuriyan 2005). Presently, significant progress has been made in the simulation of biomolecules. By now, we can examine the movement of atoms, the side-chain conformation of residues, and predict secondary structure and domains in a protein, as well as the binding pattern of nucleic acids and lipid membranes (Perilla et al. 2015). The binding free energy and conformational changes of systems can also be predicted by MD simulation as they are based on statistical mechanics. Due to this reason, the MD simulation has been used in the field of drug designing also (Paquet and Viktor 2015). Using the force field, the displacement of the particles, and energy values in each time step is calculated to define the new position of the atom (Adcock and McCammon 2006). The bonded (angles and atom bonds) and non-bonded contributions are included in the forcefield as an energy function in the classical MD simulation (Fig. 7.1).

The later contributions are made mainly by the van der Waals interaction, which is built by the Lennard-Jones 6-potential (Jones 1924). Coulomb's law is employed for the calculation of the electrostatic interaction (Cornell et al. 1995). Several algorithms involving Monte Carlo (MC) simulations, and Langevin dynamics, and

Fig. 7.1 The constituents of a force field, which represents bonded and non-bonded interactions



MD with their corresponding particularities and advantages have been reported (Adcock and McCammon 2006). The designed force field parameters and defined equations are well fitted and can reproduce the data from higher-level calculations or/and experiments. Most biomolecules, including proteins, nucleic acids, lipids, and sugars, are well parameterized in the force field for general use (Paquet and Viktor 2015). The parameters of force field for new ligands can be calculated using quantum chemistry treatment, along with many web servers, like PRODRG (van Aalten et al. 1996; Schuttelkopf and van Aalten 2004), ATP topology builder (Malde et al. 2011), and SwissParam (Zoete et al. 2010) that can generate the topology of the ligand. The bond lengths, bond dihedral angles, bond valence angles, and non-bonded interactions like van der Waals and electrostatic interactions contribute to the total energy of the systems (Hernandez-Rodriguez et al. 2016). Several force fields, like AMBER, GROMOS, and CHARMM, have also been developed (MacKerell et al. 1998). Every force field has a unique property, and a user defines the force field according to his choice based on the objective of the work (Ponder and Case 2003; Salsbury 2010). Once the force field and solvation of the proteins in an MD simulation are fixed, several parameters are also set by the users, which are defined below in brief.

7.3.1 Periodic Boundary Conditions

The periodic boundary condition (PBC) is an approach by which one can define a set of rules for the boundary of a simulation box so that atoms cannot move beyond the defined boundary during MD simulations. If the user does not define the PBC, the simulation box is repeated infinitely in every path and result in forming a lattice. For better computational efficiency, most MD simulations use this potential. Each particle interacts with adjacent images of the other particle in all these cut-off schemes (Holden et al. 2013). The long-range interactions are calculated in the case of molecular modeling and simulation by the isotropic periodic sum (IPS) method (Wu and Brooks 2009). Four significant advantages of using the IPS methods are as (1) it can eliminate the unnecessary symmetry artifacts, which originates in the PBC condition, (2) in the case of any functional form of the potential, it can be applied, (3) it can be used easily in the parallelized multi-processor computer, which indicates that it is computationally more efficient, and (4) it can predict the estimation of self-diffusion coefficient at the cut-off radius greater than 2.2 nm (Takahashi et al. 2010). Here, the long-range interactions are calculated based on the homogeneity of the simulation systems in the IPS method. Long-range interactions are represented by interactions with IPS images of a defined local region and can be reduced to short-range IPS potentials (Wu and Brooks 2009).

7.3.2 Ewald Summation Techniques

The calculation of long-range Coulombic interactions is time-dependent and labor-intensive in most of the MD simulation methods. Here, the Ewald summation method developed in 1921 for the prediction of long-range interactions is mostly used (Ewald 1921). Long-range interactions are estimated as sums that converge very slowly. Figure 7.2 shows that the conversion of the summation of two series of potential energy is the principal to obtain the Ewald sum in MD simulation.

7.3.3 Particle Mesh Ewald Method

In the particle mesh Ewald (PME) method, the potential energy is divided into two sums- Ewald's standard direct sum and the reciprocal sums. The classical Gaussian charge distributions are used in the PME method (Norberto de Souza and Ornstein 1999; Sagui and Darden 1999). The direct sum is computed directly utilizing cut-offs. In contrast, the reciprocal sum is determined by Fast Fourier Transform (FFT) with convolutions on a grid where charges interpolate in the grid points (Fig. 7.3) (Darden et al. 1993; Dessailly et al. 2007). Additionally, it does not interpolate while the forces are calculated by analytically differentiating energies. This significantly reduces the memory requirements for computation (Norberto de Souza and Ornstein 1999).

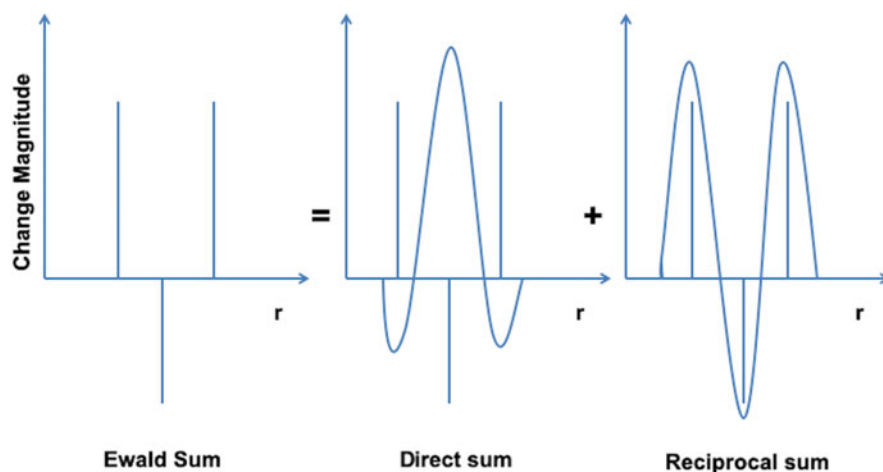


Fig. 7.2 Charge splitting into discrete and smeared distributions in the direct and reciprocal space

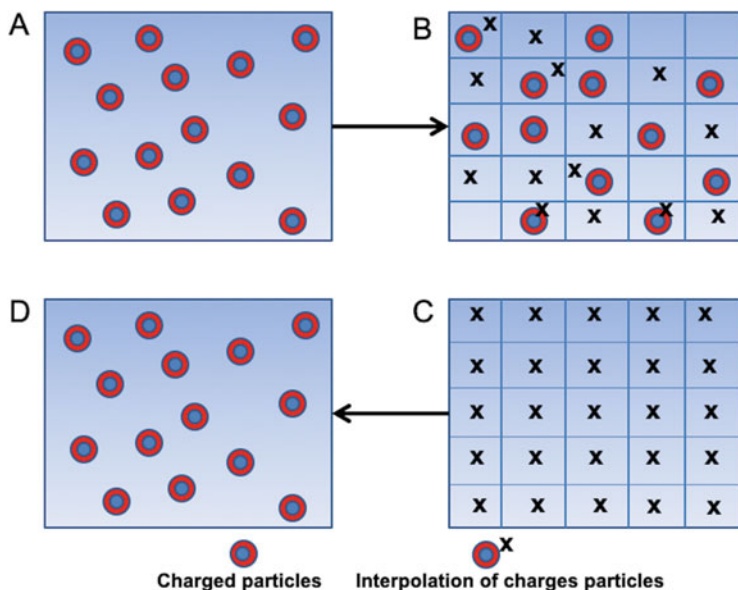


Fig. 7.3 The particle-mesh Ewald technique. A 2D representation, which is used by the majority of the Fourier-based methods. (a) The charged particle system, (b) Interpolation of charge in a 2D grid, (c) The forces and potential are computed at grid points by using FFT, (d) The coordinate updates and interpolate forces back to particles

7.3.4 Thermostats in MD

There are several thermostat methods for adding and removing the energy from the boundaries in the MD simulation system in a comparatively rational manner, for the approximation of classical ensemble (Fuzo and Degreve 2014). The number of particles (N), fixed volume (V), and the defined temperature (T) are conserved in the canonical ensemble. In a thermostat method, the energy exchange occurs between the endothermic and exothermic processes (Fuzo and Degreve 2014).

7.3.5 Solvent Models

Biomolecular MD simulation is performed in a realistic type water environment where the explicit solvent model is used (Nguyen et al. 2014). Several solvent models (mostly water models) are used in MD simulation as available: SPC/E, TIP3P, TIP4P, and TIP5P (Jorgensen and Tirado-Rives 2005). These water models are well optimized with one or many physical properties of water, such as density anomaly, diffusivity, and radial distribution function. To mimic the real cell-like environment, the MD simulation system contain explicit water molecules.

7.3.6 Energy-Minimization Methods in MD Simulations

There are different energy minimization approaches for MD structural data. By using a grid search method, the low energy regions are identified by the use of energy function in the zero-order method. In the case of gradient as an energy function, the conjugant gradient or steepest descent method is the first derivative technique. The Newton–Raphson algorithm uses the Hessian function for locating the energy minima as it is a second derived method (Kini and Evans 1991). Two main methods for energy minimization are (1) steepest descent method and (2) conjugant gradient algorithm method.

To remove the bad contacts and correction of bad geometries, the steepest descent method is widely used (Kini and Evans 1991). This method is particularly useful when the molecular system is farther from the minimum energy state, and it drifts down to the steepest slope on the potential energy surface by inducing minor structural modifications. In this method, the gradient is computed from its initial location and moves in the opposite direction to reach the minimum state. When the atoms are moving in a small increment pattern from one direction to another in coordinated systems, then for the initial geometry, the energy is computed. This process repeats for all the atoms, that eventually move to a new position downhill on the energy surface where every new step is at right angles to the one before it. This process occurs in the smaller steps to proceed down along a narrow valley and halts when the condition of a predetermined threshold value is achieved. The steepest descent method is used as the first rough and introductory run, followed by the subsequent minimization.

The conjugant gradient algorithm method is another method for energy minimization and is a primary order of minimization. This method performs minimization by using a mutually current gradient and preceding search command (Kini and Evans 1991). As compared to the steepest descent method, this method is congregated faster because it computes the search direction by using the history of minimization. It is the first derivative rate of change of the total energy in relation to the atomic positions with units of the gradient ($\text{kcal.mol}^{-1} \text{ \AA}^{-1}$). An array of directions is produced by which it succeeds over the oscillatory actions in the narrow valleys for the steepest descents method. In each minimization step, the gradient calculation is done for vector computations to predict the new direction for the minimization procedure as additional information (Feyfant et al. 2007). For the prediction of minimum energy, the direction is defined by each consecutive step. This method is preferred for larger systems (with a high number of atoms), and more storage space and calculation efforts are required. The expense of total computational and the long time period per iteration is compensated by efficient convergence to the minimum (Kini and Evans 1991). For illustrating convergence, there are various types of minimization procedures for molecular structures. For non-gradient minimizers, the augmentations in the energy and the coordinates can be measured to find the real geometry of the particular molecular system. All the gradient minimizers use atomic gradients.

7.4 Current Tools for MD Simulation

Several tools are available to investigate the atomic-level changes in the biomolecules using the MD simulation method (Khan et al. 2016). Some provide the graphical user interface like Desmond, while some run in command lines like GROMACS and AMBER. Some famous and widely used tools for MD simulation are GROMACS (Pronk et al. 2013; Oostenbrink et al. 2004), (AMBER) (Case et al. 2005; Salomon-Ferrer et al. 2013), Nanoscale MD (NAMD) (Phillips et al. 2005), and (CHARMM-GUI) (Brooks et al. 2009). For running such MD simulations, increased hardware power and software are essential components.

7.4.1 Recent Advances in Hardware to Run MD Simulation

Rapid development in computer hardware is a crucial part of MD simulation. Two reasons have an impact on trajectory analysis. The first one is the run of the long simulation result in GBs to TBs data storage, and the other is to develop the new rendering engines for the visualization effects using the latest video chipsets. Due to the advancement in the computer hardware, simulations can be performed from ns to μ s with the help of GPUs (graphics processing units) that is configured with the molecular simulation suite (Hernandez-Rodriguez et al. 2016; Gotz et al. 2012; Salomon-Ferrer et al. 2013). The GPU cards are replacing the CPU (central processing unit) and becoming commodity software and play a crucial role in decreasing the time for MD simulation. The CUDA (Compute Unified Device Architecture) is a newly invented parallel computing platform, and its use in GPU increases the number of cores to run a long simulation within time (Zhou et al. 2012; Krieger and Vriend 2015; Ge et al. 2013). Due to the emergence of GPU-CUDA technology, vigorous and massively parallel clusters are developed, such as special-purpose supercomputer Anton and Blue waters (David et al. 2007). They are precise for running the MD simulation of biomolecules from μ s to ms time-scale. But such resources are limited for limited researchers. To remove the time-scale gap, there is an urgent need to develop newer algorithms that allow enhanced sampling in the defined areas of conformational space and access long time-scale actions using necessary hardware. The purpose of this algorithm is to collect sufficient sampling that could result in the Boltzmann distribution of the diverse conformational states for the accurate calculation of the thermodynamic and kinetic properties of the system (Doshi and Hamelberg 2015). By the modification of the Hamiltonian method is to add a bias potential, several approaches have been developed like hyper dynamics (Voter 1997), local elevation (Huber et al. 1994; Voter 1997), and accelerated MD (Rodriguez-Bussey et al. 2016). In the case of hyper dynamics simulation, the identification of transition state required, but it is not necessary for classical MD simulation. Several tools are available to perform the MD simulation study with CPU or GPU. A few of the widely used tools are described in brief below.

7.4.2 GROMACS

GROMACS is the most widely used software for MD simulation. 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). In the GROMACS simulation kit, MD simulation can be performed at various temperatures and pH values. In this simulation tool, several commands are available to perform a distinct function and calculate specific structural parameters. GROMACS, which is one of the MD simulation software, can read only the 20 natural amino acids, i.e., the non-standard amino acids are not read by GROMACS algorithms. Sometimes, there are force field limitations, for instance, Gromos and Amber cannot read the nicked DNA, but the same force field can read the same non-nicked DNA. The brief methodology for MD simulation using GROMACS is shown in Fig. 7.4.

To start, the user creates a box and fills the solvent (water). The solvent model depends on the force field. After placing the protein in the defined box in the solvent, the charge of the system is neutralized either by the addition of Na^+ or Cl^- ions; this is followed by the minimization of the system using the steepest descent method. Then, NVT (the constant Number of particles, Volume, and Temperature) simulation is run to maintain the volume and temperature of the defined system. The

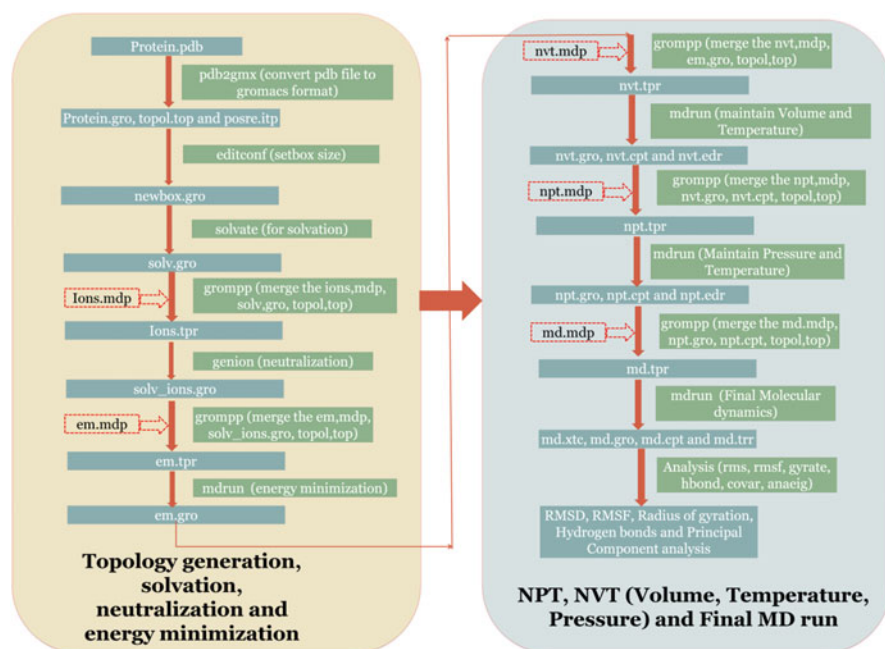
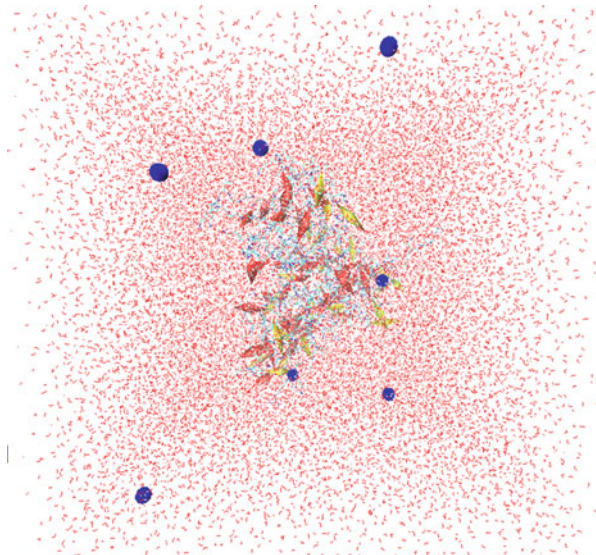


Fig. 7.4 Schematic representation showing the methodology of the MD simulation steps for GROMACS

Fig. 7.5 A protein placed in a cubic box in the water system. The red color shows water molecules, while the blue color represents the ions



temperature of the system arises from 0 and attains the desired temperature that is set by the user. After that, NPT (the constant Number of particles, Pressure, and Temperature) simulation is run to maintain the pressure of the defined systems. Several parameters are set by the addition of the .mdp file. Finally, MD simulation is performed that provides the coordinates of each step in the form of a trajectory. The trajectory can be analyzed by using various tools that are embedded in GROMACS, like `gmx-rms`, `gmx-rmsf`, `gmx-gyration`, and `gmx-hbond`. These data can be plotted in an interactive form by using GRACE (Graphing Advanced Computation and Exploration of data), a Linux based software. For example, a water embedded protein molecule, placed in a box visualized by VMD (Visual Molecular Dynamics) shown in Fig. 7.5.

7.4.3 AMBER

AMBER simulation suite is a collection of programs that are used to carry out and analyze the MD simulations for proteins, carbohydrates, and nucleic acids. Three main components of the AMBER tool are preparation, simulation, and analysis. The Antechamber and LEaP are the main program for the preparation of macromolecules. The Antechamber tool prepares the files into the force field descriptor files, which is read by the LEaP program for molecular modeling. The LEaP program then creates the topology files and Amber coordinates, which is then used in the MD simulation. The Sander program performs the MD simulation by fixing the temperature, pressure, and pH of the defined system. Lastly, the analysis part is performed by the ptraj module, which calculates the RMSD, RMSF, radius of gyration (R_g), H-bonds, and cross-correlation functions.

7.4.4 CHARMM-GUI

CHARMM-GUI is a simulation tool for the analysis of macromolecular dynamics and associated mechanical attributes. It performs standard MD simulations by using state-of-the-art algorithms for time stepping, long-range force calculation, and periodic images. Various analyses, such as energy minimization, crystal optimization, and normal mode analysis, can be performed using CHARMM.

7.4.5 NAMD

The simulations of much large biomolecular systems are performed using NAMD. The NAMD is available free of charge. The source code documentation including a set of compiled binary files configured with various parallel source software for calculations are freely available to the user. It supports massively parallel CUDA technology. NAMD can be used with graphical user interface software VMD. The simulation can be set and analyzed using the VMD as an interface. It is also compatible with AMBER and CHARMM (Khan et al. 2016).

7.4.6 Quantum-Mechanics/Molecular-Mechanics (QM/MM)

The QM/MM methods are a widely used approaches for biomolecular systems modeling (Groenhof 2013; Warshel and Levitt 1976). The various processes and charge transfer can be described using the QM method, but the QM methods are restricted to only a few hundred atom systems. Though simulations of a large system and a long-time period are performed by highly efficient force field-based MM methods. For modeling of the large biomolecules, the hybrid QM/MM method is very efficient (Senn and Thiel 2009). The QM/MM method is widely used with various applications, such as MD simulation, free energy calculation, geometry optimization, and computational spectroscopy (Groenhof 2013).

7.4.7 HyperChem

The HyperChem tool is also a tool for molecular modeling (Froimowitz 1993). It is an attractive commercial programming product manufactured by Hypercube Company and also given for free 30 days trial. It has a set of tools for molecular mechanics, quantum chemistry, and MD of the biological systems. The attractiveness of this software is attributed to the availability of complete documentation supported by examples, making this package ideal for studying the principle and practical approaches to molecular modeling (Gutowska et al. 2005). However, this program is comparatively slow as it cannot use multiprocessor support. An efficient strategy to use this tool is to employ it as an interactive molecular designer.

7.5 Other Advance Methods for MD Simulation

The MD simulation is a very progressive field. Several advancements have happened, and several other new methods are also introduced regularly to reduce the time complexity.

7.5.1 Metadynamics

It is an advantageous and powerful method of MD simulation to enhance sampling. The collective variables (CVs) define the free energy landscape reconstruction as a function of a few selected degrees of freedom. In this method, the sampling is accelerated by the history-dependent bias potential (Barducci et al. 2011). The space of collective variables is used for the adaptive construction of bias potential. In recent times, considerable improvements have been made to the actual algorithm, leading to a well-organized, flexible, and precise method that has found many successful applications in several domains. There are various examples of metadynamics study, and the most common umbrella sampling method is discussed below (Barducci et al. 2011).

The main challenge in computational biology is to predict the accurate binding free energy difference between two or more systems. For this problem, a new method umbrella sampling introduced, which is a biased MD simulation method, and it calculates free energy using the reaction coordinates. In this method, the system is driven from one thermodynamic state to other thermodynamics states. For example, reactant and product by using the bias potential reaction coordinate along with one or more directional (Kastner 2011). The intermediate steps are covered by a series of windows, at every stage of which an MD simulation is performed. Any functional form can represent bias potential. The harmonic potential is used in this method. Using the reaction coordinates, with the sampled distribution of a system, the free energy change in each window can be calculated. By using the umbrella integration or weighted histogram analysis method, the obtained windows are combined. The bias directly gives the free energy change between the two systems. The replica-exchange method can be used to improve the sampling of each window, either by replacing between successive windows or by running additional simulations at higher temperatures (Kastner 2011).

7.6 Analysis of MD Trajectories Through GUI-Based Software

The resulting output trajectories of any MD simulation can be visualized using GUI-based software. In the section below, we will discuss a few most popular software.

7.6.1 Visual Molecular Dynamics

VMD is developed at the University of Illinois (Hsin et al. 2008; Falsafi-Zadeh et al. 2012; Humphrey et al. 1996; Knapp et al. 2010). It is a potent tool for analysis and visualization of various biological systems such as proteins, nucleic acids, carbohydrates, and lipids. It is compatible with a large number of file formats, such as PDB and GROMOS. It can process a large amount of data for the visualization of trajectory movements (Falsafi-Zadeh et al. 2012). The molecules can be viewed in the animated form, and a movie can also be created from the input trajectory. It can be operated from a remote computer and also compatible with any operating systems with basic computer configuration. It is included with NAMD also. The additional functions of this tool are given below (Likhachev et al. 2016).

1. The visualization and analysis of macromolecules.
2. The atoms and amino acids can be selected.
3. Two structures can be aligned.
4. Support of user's action recording in the scripts.
5. The Raster3D format support (This format can give a high-quality image).
6. Ramachandran plots can be generated.
7. Support various types of molecular images.
8. Stereoscopic output.
9. Command-line support.
10. Working with arrays and vectors.
11. Support of JavaScript.

7.6.2 PyMOL

The PyMOL is among the most widely used software in the field of structural biology. It can accept various formats like PDB, Mol2, SDF, and several other file formats. The user can import the trajectory and analyze the simulation result. The user can generate a surface view model. Several additional plug-ins are also available to analyze the result of MD simulation. The user can create high-quality figures and animated movies from this tool.

7.6.3 Chimera

UCSF Chimera is an advanced software (Pettersen et al. 2004). It is widely used and freely available for academics. It was developed by Resource for biocomputing, visualization, and informatics (RBVI) and funded by the National Institutes of Health. The UCSF ChimeraX is also available, which is more advanced than Chimera. The Chimera 1.13.1 is released on 14-08-2018 and accepts the GROMACS and AMBER trajectory formats. After importing these trajectories, the user can make an animated movie with the time frame and generate high-

quality pictures. The user can align two or more structures. The user can also generate the surface for cavity analysis during trajectory run. It has several features and supports the command line option also.

7.7 Structural Parameters for Analysis of MD Simulation

The MD simulation produces the result in the form of a trajectory. The trajectory contains all the parameters that were generated during each step movement of atoms. Various structural parameters that can be used to analyze the results are the following:

7.7.1 RMSD

The root mean square deviation (RMSD) is the most important and first parameter to analyze any MD trajectory (Kuzmanic and Zagrovic 2010). RMSD is used to measure the difference between the backbones of a protein from its initial structural conformation to its final position. The stability of the protein relative to its conformation can be determined by the deviations produced during the MD simulation process. RMSD is calculated with respect to the reference native conformation r_{ref} using the following formula:

$$\text{RMSD}(t) = \left[\frac{1}{M} \sum_{i=1}^N m_i |r_i(t) - r_i^{\text{ref}}|^2 \right]^{1/2}$$

where $M = \sum_i m_i$ and $r_i(t)$ represents the atom, i position at the time t after least square fitting the structure to a reference structure.

The RMSD for all residues, backbone, side-chain, and C α atoms can be calculated. RMSD is calculated with respect to the simulation time. Smaller deviations indicate a more stable protein structure and vice versa. In general, an RMSD value for a macromolecule should be less than 2 Å to extract any meaningful data. A comparative study of RMSD for native isocitrate lyase from *Mycobacterium tuberculosis* (MtbICL) and its mutant has been performed (Fig. 7.6a). A major difference in the RMSD of native (black) and mutant MtbICL (red) has been observed, and it indicates that a single mutation L148A in MtbICL protein (MtbICL_{L148A}) causes structural perturbations in the enzyme and reduces its stability (Shukla et al. 2018c).

7.7.2 RMSF

The root mean square fluctuation (RMSF) is the best way to study the residue-wise fluctuation of the protein from the MD trajectory. It describes the fluctuation of each residue or domain in a protein. The RMSF can be plotted as RMSF (nm) vs. residue

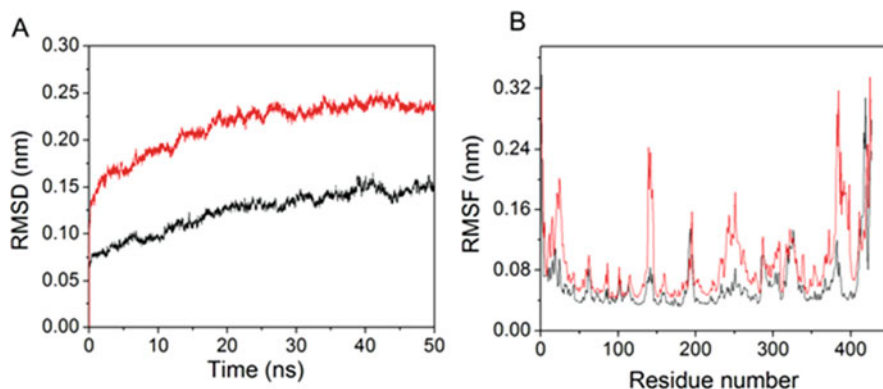


Fig. 7.6 (a) RMSD of native and mutant MtbICL. (b) RMSF of the C α atoms calculated from the last 30 ns of the MD trajectory. Black and red lines represent MtbICL and MtbICL_{L148A}, respectively

number (Kuzmanic and Zagrovic 2010). The RMSF of a protein is measured by the deviation between the position of particle i with its reference position, and T represents the time, and r_{i}^{ref} is the reference position of particle i .

$$\text{RMSF}_i = \left[\frac{1}{T} \sum_{t_j=1}^T |r_i(t_j) - r_i^{ref}|^2 \right]^{1/2}$$

Well organized and rigid structures, like helix and sheets, show low RMSF, while loosely structure like bends and coils showed higher RMSF value because atom can have more fluctuation in the bends and coils as compared to helix and sheet. RMSF of the C α atoms for native and mutant MtbICL was calculated from the MD trajectory to compare the residue-wise fluctuations (Fig. 7.6b). A high degree of residue-wise fluctuations observed in mutant MtbICL as compared to the native structure, and this happens due to a single mutation L148A, which causes more fluctuations and instability in the mutant. This mutation can also deform the shape of the binding site and which in turn can prevent the function of the enzyme (Shukla et al. 2018c).

7.7.3 Radius of Gyration

The radius of gyration (R_g) of a protein describes the compactness of the folded protein. For the same size proteins, ideal R_g value should be less for the globular folded state, while in expanded form or protein form with more number of loops and turns, the R_g value should be relatively high. The R_g value of a structure is calculated from the following equation:

$$R_g = \left(\frac{\sum_i |r_i|^2 m_i}{\sum_i m_i} \right)^2$$

In this equation, the m_i represents the mass of atom i and r_i is the position of atom i with respect to the center of mass of the molecule. The system total mean energy is calculated by the following equation:

$$\langle E \rangle = \frac{1}{N} \sum_{i=1}^N E_i$$

In the above equation, E_i represents the energy of atom i .

7.7.4 Protein–Ligand Contacts

Non-covalent bonds play a significant role in determining the stability of the protein and ligand. These bonds play a pivotal role in protein structure folding and protein–ligand stability. During the folding, the hydrogen bonds provide a path for proper folding of a protein. Other interactions, such as hydrophobic bonds, make the inner patches and the catalytic triad of a biomacromolecule. The hydrogen bonds between residues can be calculated with respect to time during the simulation. In MD simulation analysis of apo-protein and protein–ligand, the hydrogen bonds, as well as other interactions, can be calculated during simulation time to find the potential residues necessary for ligand stabilization. The protein–ligand contact map is generated to find the residues that are in contact (interaction) with ligand during most of the simulation time.

7.7.5 SASA

The solvent accessible surface area (SASA) defines the area of the protein that interacts with the solvents in a simulation box (Mazola et al. 2015). For the same size proteins, the folded globular state shows lesser SASA value, while the expanded form of the protein shows higher SASA value. It is well-known that an increase in the temperature of the system will lead to protein unfolding, and the hydrophobic core of protein gets exposed toward the solvent. As a result, the SASA value increases upon unfolding.

7.7.6 Principal Component Analysis or Essential Dynamics

Essential dynamics (ED) reflect the overall increase of the motions in a protein during the time-scale of simulations (Maisuradze et al. 2009). The principal component analysis (PCA) predicts the collective motions of the protein and reveals the

atomic fluctuations in the structure (David and Jacobs 2014). Every atom is related to each other in the MD simulation. The correlation motion analysis is very important for predicting the behavior of the molecules. The covariance matrix of atomic fluctuations is diagonalized for predicting the eigen values. The first eigenvectors play an essential role in the overall motions of the protein. The PCA analysis is used to compare the correlated motions of a protein under various conditions.

7.7.7 Secondary Structure Analysis

The secondary structure is a crucial parameter in the analysis of the MD results, and it describes the contents of secondary structure with respect to time. It clearly describes the structural contents to understand the stability of each domain. It is instrumental in mutational and protein unfolding studies, as it can explain the unfolding of a domain and its stability during a simulation.

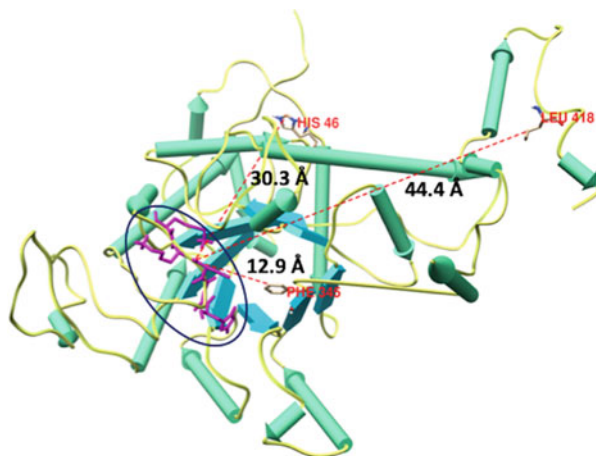
7.8 Application of MD Simulation

MD simulation is a widely used technique for studying biological systems. It is applicable in several fields like mutational analysis of protein, protein–ligand complex stability prediction, protein unfolding studies, conformational protein stability prediction, etc. It is a potent tool that requires high computational power to reveal biological mysteries. In the following sub-sections, we will briefly discuss the applications of MD simulation in all these above-stated fields using case studies.

7.8.1 Mutational Analysis

MD simulation can be used to confirm the results of in vitro mutational studies, as it can predict the conformational movement of the atoms after and before mutation. It can also predict the alternate binding site. A case study of mutational analysis from our previous works is presented here. In these studies, the role of structurally distant amino acids ($>10 \text{ \AA}$) that are away from the active site signature motif ($^{189}\text{KKCGH}^{193}$) was studied (Fig. 7.7). We showed that point mutations brought the vast extent of conformation changes in the active site conformation, which results in loss of activity of the enzyme. Three mutations (F345A, L418A, and H46A) were introduced in the structure by the in silico approach, and then MD simulation was performed to see the effect of the mutation on the structure and function of the enzyme. Here, we take an example of several MD simulation analyses of an enzyme isocitrate lyase from *Mycobacterium tuberculosis* (MtbICL) and its three point mutants and extract structural and dynamic information from the MD simulation trajectories. Using MD simulation studies, we showed that distant point mutations at H46, F345, and L418, which are structurally distant ($>10 \text{ \AA}$) to the active site sequence ($^{189}\text{KKCGH}^{193}$), completely abrogates the enzyme activity.

Fig. 7.7 Barrel structure of MtbICL. The distance of the mutated amino acids from the active site catalytic residues is shown in red lines



MtbICL is an important anti-tubercular drug target that catalyzes the first step of the glyoxylate shunt and is required for survival of the pathogen under dormancy condition. To understand the role of structurally distant amino acids in regulating the structural dynamics and conformational flexibility of MtbICL, several point mutants were created by site-directed mutagenesis, and their structure–activity relationship was studied. Three mutations, F345A (Shukla et al. 2018b), L418A (Shukla et al. 2018c), and H46A (Shukla et al. 2018f), were made. The experimental data revealed that these mutants were causing complete loss of enzyme activity, though these residues were not present within or near the active-site. Structurally, these residues were present more than 10 Å from the active site (Fig. 7.6).

To study the structural changes upon mutation and its effect on the dynamics of the enzyme, MD simulations were performed. A comparative study of RMSD for native and the three point mutants of MtbICL was done. The RMSD for the native, H46A, F345A, and L418A MtbICL mutant was found to be 0.12 nm, 0.13 nm, 0.14 nm, and 0.21 nm, respectively. The RMSD of F345A and H46A mutation shows an insignificant difference in the average RMSD values, thereby indicating that they do not significantly influence the overall stability of the protein. The RMSD value of L418A mutation suggests that it negatively affects protein stability, inducing destabilization in the mutant protein structure.

For understanding the effects of the mutations on the structural fluctuation of the entire protein and also on individual residues, the RMSF of C α atoms was analyzed for all the proteins. RMSF of the native and the three mutants was calculated from the MD trajectories to compare the residues wise fluctuations. In H46A and F345A mutant, minor changes were observed in the RMSF (Shukla et al. 2018c; Shukla et al. 2018f), while the L418A mutant showed significant fluctuations (Shukla et al. 2018f). The data suggest that H46A and F345A mutation does not induce changes in the overall protein structure. In contrast, a high degree of residue wise fluctuations observed in the L148A mutant, which causes more fluctuations and instability in the structure (Shukla et al. 2018c).

H46A mutation did not show any global structural alternations; it caused changes in the catalytic site (Shukla et al. 2018f). The PCA and cross-correlation analysis showed that H46A mutation caused a change in conformational stability and collective motions of the protein, particularly in the active site region. The residue interaction network (RIN) analysis indicates that the active site geometry was disturbed in the H46A mutant. These results suggest that due to the mutation, the dynamic perturbation of the active site leads to enzyme transition from its active form to inactive form (Shukla et al. 2018f).

The mutation in F345 induced structural flexibility and conformational rearrangements near the active site. This mutation increased the collective motions and residual mobility of the enzyme, resulting in a decrease in the mutant enzyme stability. The result was confirmed by the lower free energy in the mutant enzyme indicating the destabilized structure (Shukla et al. 2018b).

The L418A mutation also nullifies the activity of the protein. The correlated motions, residual mobility, and flexibility in the enzyme increased upon mutation. Upon L418A mutation, the global conformational dynamics and the RIN of the protein changed. The RIN depicts that several hydrogen bonds, hydrophobic bonds were distorted due to the mutation. This alteration in RIN brings conformational changes in the active site leading to the loss of enzyme activity (Shukla et al. 2018f).

Ultimately, molecular docking data indicated that all three mutations affected the substrate interactions with the active site residues of MtbICL. These results reveal the internal dynamics of the enzyme structure and feature the importance of residue-level interactions in the enzyme.

7.8.2 Application in the Drug Designing

The protein–ligand docking is a significant phase in the field of drug designing. Nowadays, several software are available for structure-based virtual screening. The PDB entries are increasing due to newer structure determination methods. Most of the docking software considers the protein as a rigid body, while ligand is always considered as flexible. Some docking software recognizes the protein and ligand both as a flexible molecule, and they can produce a better pose with a binding affinity. The question arises now is: will this docking pose exist in the cell because every protein is dynamic? To solve this problem, MD simulation is an excellent approach used to predict the stability and dynamics of the protein–ligand complexes.

7.8.2.1 Inhibitor Designing Against MtbICL

Several MtbICL structures are available in the PDB in complex with inhibitor and substrates. An inhibitor-bound structure was retrieved from the PDB (PDB ID: 1F8I), and 167,674 compounds were screened in various runs. Following rounds of screening and docking refinement, 340 compounds were selected. For validation of the docking results and to study the dynamics of the system, MD simulation was performed as docking does not provide insight into the dynamics of the system. The MD simulation data of three compounds were compared and only one compound

was found to have high potential to inhibit MtbICL. Thus, the MD simulation can be used to remove the false positive binders and provides information on the detailed mechanism of ligand-induced inhibition of enzyme function (Shukla et al. 2018e).

7.8.2.2 Inhibitor Designing Against *Fasciola gigantica* Thioredoxin Glutathione Reductase

Fasciola gigantica thioredoxin glutathione reductase (FgTGR) is a key drug target against fascioliasis caused by the helminth parasite *Fasciola gigantica*. We reported some novel inhibitors of FgTGR using the structure-based virtual screening approach, and the selected hits were validated by MD simulation. The compounds were screened against FgTGR in several runs. The selected compounds were evaluated through ADMET, and some compounds were chosen for further docking. Ultimately, three compounds were used for the MD simulation that resulted in one highly potent compound with a high affinity towards FgTGR. Thus, the MD simulation can play an important role in the screening of potential inhibitor against a target considering the physiological conditions of the interaction environment (Shukla et al. 2018b).

7.8.3 Unfolding Studies

MD simulation is a compelling technique to study the protein unfolding mechanism at an atomic level (Prakash et al. 2018; Sonkar et al. 2018). We can track the mechanism of unfolding as a result of chemical-, pH-, or temperature-induced denaturation using MD simulation, which provides a clear view of structural alternations taking place at a particular time-scale. Protein unfolding by pH, urea, GdnHCl, and temperature has already been performed using MD simulation.

7.8.3.1 Urea Induced Unfolding of FgGST1

Urea is a widely used denaturant to study the mechanism of protein unfolding. It has been proposed to disrupt the hydrophobic interactions, as a result of which the hydrophobic patches of proteins open and come into contact with water (Kalita et al. 2018a). We took an 8 M urea environment for analysis and performed 100 ns simulation at 300 K and 400 K temperature to understand the unfolding mechanism of a protein. The dynamics of protein were recorded at an interval of 40 ns; the data indicate that 8 M urea induces unfolding, and finally leads to the disruption of the complete protein 3D structure (Fig. 7.8). The RMSD, RMSF, Rg, and PCA analyses suggest that urea induces the unfolding of the protein. Different secondary structure alternations, such as loss of alpha-helices, loss of bends, and beta-sheets, were observed in the protein during MD simulation (300 K), which indicates the loss of native structure. In MD simulation at 400 K, a greater extent of unfolding in protein structure was observed.

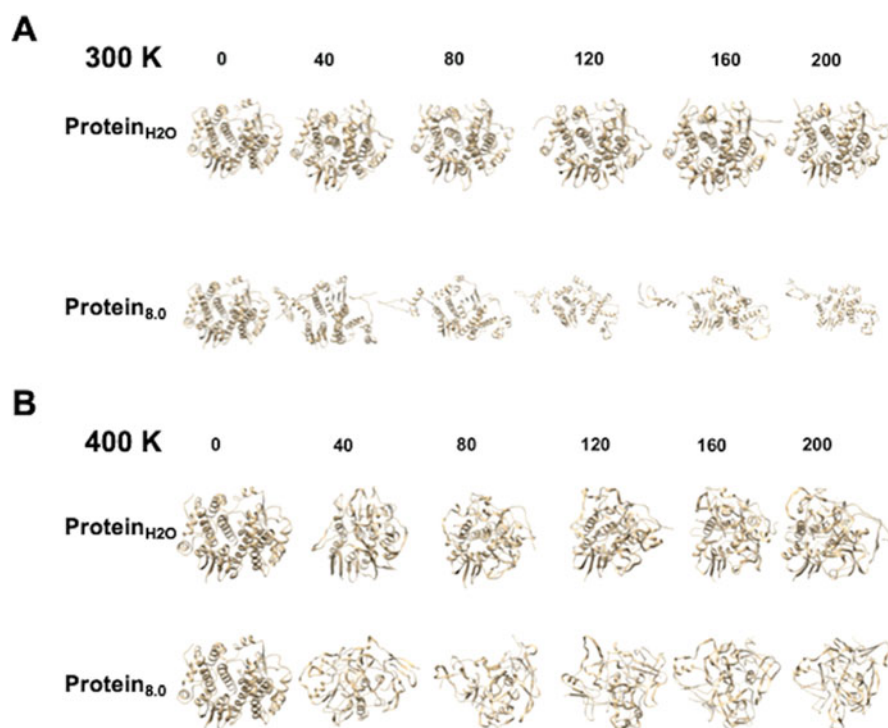


Fig. 7.8 Different time frame snapshots of unfolding at 40 ns time interval for a protein with water and urea at (a) 300 K (b) at 400 K temperature. Protein_{H₂O} and Protein_{8.0} represent the protein in water and 8 M urea, respectively (Kalita et al. 2018a). The MD simulation was run for 200 ns

7.8.3.2 GdnHCl-Induced Unfolding Analysis

Syed et al. performed the denaturation study of a protein using GdnHCl at 300 K (Syed et al. 2018). 100 ns MD simulation was run to investigate the atomic-level changes. Various structural parameters such as RMSD, RMSF, Rg, SASA, the number of hydrogen bonds, and PCA were calculated and revealed the unfolding mechanism of the protein (Syed et al. 2018).

7.8.3.3 pH-Induced Effects on the Structure and Stability of the Protein

MD simulation can also be performed in different pH values to model the structural rearrangement pattern in different pH conditions. Syed et al. also used pH in MD simulation (2, 4, 6, 8, 10, and 12) to study protein unfolding. Their analyses on a particular protein suggest that it can maintain the secondary and tertiary structure in the alkaline pH. In contrast, in the acidic condition (pH 2.0–5.5), significant structural changes occur (Syed et al. 2015).

7.9 Conclusions

Proteins are dynamic entities, and the dynamic nature defines its function. The availability of an accurate 3D structure is essential to understand the protein dynamics and function. The 3D structure may be solved using X-ray crystallography, NMR, or computational methods. Although these methods provide detailed information on protein structure, they still fail to provide sufficient information on protein dynamics and motion. MD simulation has a history of more than 43 years. It is a widely used technique for predicting the dynamic picture of any biological system. Development of GPUs based high computational capability system is a milestone for the MD simulation to reduce the time for predicting the dynamics of biological molecules such as nucleic acids, proteins, carbohydrates, and many more and their molecular interactions with each other or with small molecules inhibitors. MD simulation is a potent tool to solve difficult biological problems, which happen in second to millisecond time-scales like molecular interaction of protein–protein, protein–ligand interaction, protein folding, and unfolding analysis. It considers the biological pH, the surrounding molecules for creating the cell-like environment like water and lipids, as well as co-enzymes, ions, and nucleic acid. MD simulations provide atomic-level details of atom interaction that are associated with the function of the molecules. The binding free energy, various energy constituents, and residue-wise binding contribution with a ligand can be predicted using the MM-PBSA tool. This information is beneficial for further improvement in the binding affinity. The QM and MM method implementation in the MD simulation has drastically changed towards the enhancement of accuracy of the binding free energy. By using these methods, we can easily find out the role of polarization and electronic effects in protein–protein and protein–ligand interactions. MD simulation can be used to reveal the chemistry and dynamics of a biological system by providing an appropriate model and physical conditions.

Competing Interest The authors declare that there are no competing interests.

References

- Adcock SA, McCammon JA (2006) Molecular dynamics: survey of methods for simulating the activity of proteins. *Chem Rev* 106(5):1589–1615
- Alder BJ, Wainwright TE (1957) Phase transition for a hard sphere system. *J Chem Phys* 27(5):1208–1209
- Alder BJ, Wainwright TE (1959) Studies in molecular dynamics. I. general method. *J Chem Phys* 31(2):459–466
- Barducci A, Bonomi M, Parrinello M (2011) Metadynamics. *Wiley Interdiscip Rev Comput Mol Sci* 1(5):826–843
- Berger A, Linderstrom-Lang K (1957) Deuterium exchange of poly-DL-alanine in aqueous solution. *Arch Biochem Biophys* 69:106–118
- Boehr DD, Dyson HJ, Wright PE (2006) An NMR perspective on enzyme dynamics. *Chem Rev* 106(8):3055–3079

- Brooks B, Karplus M (1983) Harmonic dynamics of proteins: normal modes and fluctuations in bovine pancreatic trypsin inhibitor. *Proc Natl Acad Sci U S A* 80(21):6571–6575
- Brooks BR, Brooks CL 3rd, Mackerell AD Jr, Nilsson L, Petrella RJ, Roux B, Won Y, Archontis G, Bartels C, Boresch S, Caffisch 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
- Brunger AT, Brooks CL 3rd, Karplus M (1985) Active site dynamics of ribonuclease. *Proc Natl Acad Sci U S A* 82(24):8458–8462
- Brunger AT, Kuriyan J, Karplus M (1987) Crystallographic R factor refinement by molecular dynamics. *Science* 235(4787):458–460
- Case DA, Karplus M (1979) Dynamics of ligand binding to heme proteins. *J Mol Biol* 132(3):343–368
- Case DA, Cheatham TE 3rd, Darden T, Gohlke H, Luo R, Merz KM Jr, Onufriev A, Simmerling C, Wang B, Woods RJ (2005) The Amber biomolecular simulation programs. *J Comput Chem* 26(16):1668–1688
- Colonna-Cesari F, Perahia D, Karplus M, Eklund H, Braden CI, Tapia O (1986) Interdomain motion in liver alcohol dehydrogenase. Structural and energetic analysis of the hinge bending mode. *J Biol Chem* 261(32):15273–15280
- Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA (1995) A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J Am Chem Soc* 117(19):5179–5197
- Darden T, York D, Pedersen L (1993) Particle mesh Ewald: an N-log(N) method for Ewald sums in large systems. *J Chem Phys* 98(12):10089–10092
- David CC, Jacobs DJ (2014) Principal component analysis: a method for determining the essential dynamics of proteins. *Methods Mol Biol* 1084:193–226
- David ES, Martin MD, Ron OD, Jeffrey SK, Richard HL, John KS, Cliff Y, Brannon B, Kevin JB, Jack CC, Michael PE, Joseph G, Grossman JP, Ho CR, Douglas JI, John LK, Timothy L, Christine M, Mark AM, Rolf M, Edward CP, Yibing S, Jochen S, Michael T, Brian T, Stanley CW (2007) Anton, a special-purpose machine for molecular dynamics simulation. Paper presented at the proceedings of the 34th annual international symposium on computer architecture, San Diego, California, USA
- Dessailly BH, Lensink MF, Wodak SJ (2007) Relating destabilizing regions to known functional sites in proteins. *BMC Bioinf* 8:141
- Doshi U, Hamelberg D (2015) Towards fast, rigorous and efficient conformational sampling of biomolecules: advances in accelerated molecular dynamics. *Biochim Biophys Acta* 1850(5):878–888
- Dror RO, Dirks RM, Grossman JP, Xu H, Shaw DE (2012) Biomolecular simulation: a computational microscope for molecular biology. *Annu Rev Biophys* 41:429–452
- Duan Y, Kollman PA (1998) Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution. *Science* 282(5389):740–744
- Ewald PP (1921) Die Berechnung optischer und elektrostatischer Gitterpotentiale. *Ann Phys* 369(3):253–287
- Falsafi-Zadeh S, Karimi Z, Galehdari H (2012) VMD DisRg: new user-friendly implement for calculation distance and radius of gyration in VMD program. *Bioinformatics* 8(7):341–343
- Feyfant E, Sali A, Fiser A (2007) Modeling mutations in protein structures. *Protein Sci* 16(9):2030–2041
- Frauenfelder H, Hartmann H, Karplus M, Kuntz ID, Kuriyan J, Parak F, Petsko GA, Ringe D, Tilton RF (1987) Thermal expansion of a protein. *Biochemistry* 26(1):254–261
- Froimowitz M (1993) HyperChem: a software package for computational chemistry and molecular modeling. *BioTechniques* 14(6):1010–1013

- 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
- Ge H, Wang Y, Li C, Chen N, Xie Y, Xu M, He Y, Gu X, Wu R, Gu Q, Zeng L, Xu J (2013) Molecular dynamics-based virtual screening: accelerating the drug discovery process by high-performance computing. *J Chem Inf Model* 53(10):2757–2764
- Gotz AW, Williamson MJ, Xu D, Poole D, Le Grand S, Walker RC (2012) Routine microsecond molecular dynamics simulations with AMBER on GPUs. 1. Generalized born. *J Chem Theory Comput* 8(5):1542–1555
- Groenhof G (2013) Introduction to QM/MM simulations. *Methods Mol Biol* 924:43–66
- Gutowska I, Machoy Z, Machalinski B (2005) The role of bivalent metals in hydroxyapatite structures as revealed by molecular modeling with the HyperChem software. *J Biomed Mater Res A* 75(4):788–793
- Harvey SC, Prabhakaran M, Mao B, McCammon JA (1984) Phenylalanine transfer RNA: molecular dynamics simulation. *Science* 223(4641):1189–1191
- Heidari Z, Roe DR, Galindo-Murillo R, Ghasemi JB, Cheatham TE (2016) Using wavelet analysis to assist in identification of significant events in molecular dynamics simulations. *J Chem Inf Model* 56(7):1282–1291
- Hernandez-Rodriguez M, Rosales-Hernandez MC, Mendieta-Wejebe JE, Martinez-Archundia M, Basurto JC (2016) Current tools and methods in molecular dynamics (MD) simulations for drug design. *Curr Med Chem* 23(34):3909–3924
- Holden ZC, Richard RM, Herbert JM (2013) Periodic boundary conditions for QM/MM calculations: Ewald summation for extended Gaussian basis sets. *J Chem Phys* 139(24):244108
- Hsin J, Arkhipov A, Yin Y, Stone JE, Schulten K (2008) Using VMD: an introductory tutorial. *Curr Protoc Bioinformatics* 24:5–7
- Huber T, Torda AE, van Gunsteren WF (1994) Local elevation: a method for improving the searching properties of molecular dynamics simulation. *J Comput Aided Mol Des* 8(6):695–708
- Humphrey W, Dalke A, Schulten K (1996) VMD: visual molecular dynamics. *J Mol Graph* 14(1):33–38, 27–38
- Jeremy Smith SC, Pezzeca U, Brooks B, Karplus M (1986) Inelastic neutron scattering analysis of low frequency motion in proteins: a normal mode study of the bovine pancreatic trypsin inhibitor. *J Chem Phys* 85(6):3636–3654
- Jones JE (1924) On the determination of molecular fields. -II from the equation of state of a gas. *Proc Roy Soc A* 106(738):463
- Jorgensen WL, Tirado-Rives J (2005) Potential energy functions for atomic-level simulations of water and organic and biomolecular systems. *Proc Natl Acad Sci U S A* 102(19):6665–6670
- Kalita J, Shukla R, Tripathi T (2018a) Structural basis of urea-induced unfolding of *Fasciola gigantica* glutathione S-transferase. *J Cell Physiol* 234(4):4491–4503
- Kalita P, Shukla H, Shukla R, Tripathi T (2018b) Biochemical and thermodynamic comparison of the selenocysteine containing and non-containing thioredoxin glutathione reductase of *Fasciola gigantica*. *Biochim Biophys Acta Gen Subj* 1862:1306–1316
- Karplus M, Kuriyan J (2005) Molecular dynamics and protein function. *Proc Natl Acad Sci U S A* 102(19):6679–6685
- Kastner J (2011) Umbrella sampling. *Wiley Interdiscip Rev Comput Mol Sci* 1(6):932–942
- 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
- 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
- Knapp B, Lederer N, Omasits U, Schreiner W (2010) vmdICE: a plug-in for rapid evaluation of molecular dynamics simulations using VMD. *J Comput Chem* 31(16):2868–2873
- Krieger E, Vriend G (2015) New ways to boost molecular dynamics simulations. *J Comput Chem* 36(13):996–1007
- Krishnamoorthy G (2012) Motional dynamics in proteins and nucleic acids control their function: revelation by time-domain fluorescence. *Curr Sci* 102(2):266–276

- Kuzmanic A, Zagrovic B (2010) Determination of ensemble-average pairwise root mean-square deviation from experimental B-factors. *Biophys J* 98(5):861–871
- Levitt M, Warshel A (1975) Computer simulation of protein folding. *Nature* 253(5494):694–698
- Likhachev IV, Balabaev NK, Galzitskaya OV (2016) Available instruments for analyzing molecular dynamics trajectories. *Open Biochem J* 10:1–11
- MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FT, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiorkiewicz-Kuczera J, Yin D, Karplus M (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J Phys Chem B* 102(18):3586–3616
- Maisuradze GG, Liwo A, Scheraga HA (2009) Principal component analysis for protein folding dynamics. *J Mol Biol* 385(1):312–329
- Malde AK, Zuo L, Breeze M, Stroet M, Poger D, Nair PC, Oostenbrink C, Mark AE (2011) An automated force field topology builder (ATB) and repository: version 1.0. *J Chem Theory Comput* 7(12):4026–4037
- Mamgain S, Sharma P, Pathak RK, Baunthiyal M (2015) Computer aided screening of natural compounds targeting the E6 protein of HPV using molecular docking. *Bioinformation* 5:236–242
- Mazola Y, Guirola O, Palomares S, Chinae G, Menendez C, Hernandez L, Musacchio A (2015) A comparative molecular dynamics study of thermophilic and mesophilic beta-fructosidase enzymes. *J Mol Model* 21(9):228
- McCammon JA, Gelin BR, Karplus M (1977) Dynamics of folded proteins. *Nature* 267(5612):585–590
- McCammon JA, Karim OA, Lybrand TP, Wong CF (1986) Ionic association in water: from atoms to enzymes. *Ann N Y Acad Sci* 482:210–221
- 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
- Nilsson L, Clore GM, Gronenborn AM, Brunger AT, Karplus M (1986) Structure refinement of oligonucleotides by molecular dynamics with nuclear overhauser effect interproton distance restraints: application to 5' d(C-G-T-A-C-G)2. *J Mol Biol* 188(3):455–475
- Norberto de Souza O, Ornstein RL (1999) Molecular dynamics simulations of a protein-protein dimer: particle-mesh Ewald electrostatic model yields far superior results to standard cutoff model. *J Biomol Struct Dyn* 16(6):1205–1218
- 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
- Pandey T, Shukla R, Shukla H, Sonkar A, Tripathi T, Singh AK (2017) A combined biochemical and computational studies of the rho-class glutathione s-transferase sll1545 of *Synechocystis* PCC 6803. *Int J Biol Macromol* 94:378–385
- Paquet E, Viktor HL (2015) Molecular dynamics, Monte Carlo simulations, and Langevin dynamics: a computational review. *Biomed Res Int* 2015:18
- Perilla JR, Goh BC, Cassidy CK, Liu B, Bernardi RC, Rudack T, Yu H, Wu Z, Schulten K (2015) Molecular dynamics simulations of large macromolecular complexes. *Curr Opin Struct Biol* 31:64–74
- Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem* 25(13):1605–1612
- 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
- Ponder JW, Case DA (2003) Force fields for protein simulations. *Adv Protein Chem* 66:27–85

- Prakash A, Kumar V, Meena NK, Hassan MI, Lynn AM (2018) Comparative analysis of thermal unfolding simulations of RNA recognition motifs (RRMs) of TAR DNA-binding protein 43 (TDP-43). *J Biomol Struct Dyn* 37:178–194
- 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
- Rahman A (1964) Correlations in the motion of atoms in liquid argon. *Phys Rev* 136(2A):A405–A411
- Rahman A, Stillinger FH (1971) Molecular dynamics study of liquid water. *J Chem Phys* 55(7):3336–3359
- Rajendran V, Shukla R, Shukla H, Tripathi T (2018) Structure-function studies of the asparaginyl-tRNA synthetase from *Fasciola gigantica*: understanding the role of catalytic and non-catalytic domains. *Biochem J* 475(21):3377–3391
- Rodriguez-Bussey IG, Doshi U, Hamelberg D (2016) Enhanced molecular dynamics sampling of drug target conformations. *Biopolymers* 105(1):35–42
- Sagui C, Darden TA (1999) Molecular dynamics simulations of biomolecules: long-range electrostatic effects. *Annu Rev Biophys Biomol Struct* 28:155–179
- 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
- Salsbury FR Jr (2010) Molecular dynamics simulations of protein dynamics and their relevance to drug discovery. *Curr Opin Pharmacol* 10(6):738–744
- Schuttelkopf AW, van Aalten DM (2004) PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallogr D Biol Crystallogr* 60(Pt 8):1355–1363
- Senn HM, Thiel W (2009) QM/MM methods for biomolecular systems. *Angew Chem Int Ed Engl* 48(7):1198–1229
- Shukla H, Khan SR, Shukla R, Krishnan MY, Akhtar MS, Tripathi T (2018a) Alternate pathway to ascorbate induced inhibition of *Mycobacterium tuberculosis*. *Tuberculosis (Edinb)* 111:161–169
- Shukla H, Shukla R, Sonkar A, Pandey T, Tripathi T (2018b) Distant Phe345 mutation compromises the stability and activity of *Mycobacterium tuberculosis* isocitrate lyase by modulating its structural flexibility. *Sci Rep* 7(1):1058
- Shukla H, Shukla R, Sonkar A, Tripathi T (2018c) Alterations in conformational topology and interaction dynamics caused by L418A mutation leads to activity loss of *Mycobacterium tuberculosis* isocitrate lyase. *Biochem Biophys Res Commun* 490(2):276–282
- Shukla R, Shukla H, Kalita P, Sonkar A, Pandey T, Singh DB, Kumar A, Tripathi T (2018d) 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
- Shukla R, Shukla H, Sonkar A, Pandey T, Tripathi T (2018e) 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
- Shukla R, Shukla H, Tripathi T (2018f) 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
- Singh DB, Tripathi T (2020) *Frontiers in protein structure, function, and dynamics*. Springer, Singapore. <https://doi.org/10.1007/978-981-15-5530-5>. ISBN 978-981-15-5529-9
- Sonkar A, Shukla H, Shukla R, Kalita J, Pandey T, Tripathi T (2017) UDP-N-Acetylglucosamine enolpyruvyl transferase (MurA) of *Acinetobacter baumannii* (AbMurA): structural and functional properties. *Int J Biol Macromol* 97:106–114
- Sonkar A, Shukla H, Shukla R, Kalita J, Tripathi T (2018) Unfolding of *Acinetobacter baumannii* MurA proceeds through a metastable intermediate: a combined spectroscopic and computational investigation. *Int J Biol Macromol* 126:941–951

- Syed SB, Shahbaaz M, Khan SH, Srivastava S, Islam A, Ahmad F, Hassan MI (2015) Estimation of pH effect on the structure and stability of kinase domain of human integrin-linked kinase. *J Biomol Struct Dyn* 37:156–165
- Syed SB, Khan FI, Khan SH, Srivastava S, Hasan GM, Lobb KA, Islam A, Hassan MI, Ahmad F (2018) Unravelling the unfolding mechanism of human integrin linked kinase by GdmCl-induced denaturation. *Int J Biol Macromol* 117:1252–1263
- Takahashi K, Narumi T, Yasuoka K (2010) Cutoff radius effect of the isotropic periodic sum method in homogeneous system. II. Water. *J Chem Phys* 133(1):014109
- van Aalten DM, Bywater R, Findlay JB, Hendlich M, Hooft RW, Vriend G (1996) PRODRG, a program for generating molecular topologies and unique molecular descriptors from coordinates of small molecules. *J Comput Aided Mol Des* 10(3):255–262
- Vogeli B, Kazemi S, Guntert P, Riek R (2012) Spatial elucidation of motion in proteins by ensemble-based structure calculation using exact NOEs. *Nat Struct Mol Biol* 19(10):1053–1057
- Voter AF (1997) Hyperdynamics: accelerated molecular dynamics of infrequent events. *Phys Rev Lett* 78(20):3908–3911
- Warshel A, Levitt M (1976) Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J Mol Biol* 103(2):227–249
- Wolf A, Kirschner KN (2013) Principal component and clustering analysis on molecular dynamics data of the ribosomal L11.23S subdomain. *J Mol Model* 19(2):539–549
- Wu X, Brooks BR (2009) Isotropic periodic sum of electrostatic interactions for polar systems. *J Chem Phys* 131(2):024107
- Zhou Y, Liepe J, Sheng X, Stumpf MP, Barnes C (2012) GPU accelerated biochemical network simulation. *Bioinformatics* 27(6):874–876
- Zoete V, Cuendet MA, Grosdidier A, Michielin O (2010) SwissParam: a fast force field generation tool for small organic molecules. *J Comput Chem* 32(11):2359–2368