Molecular Dynamics Simulations: Concept, Methods, and Applications

7

Mohammad Sufian Badar, Shazmeen Shamsi, Jawed Ahmed, and Md. Afshar Alam

"Our future world will have to find equilibrium in the technology pendulum swing."

Stephane Nappo.

M. S. Badar (✉)
Department of Bioengineering, University of California, Riverside, CA, USA
e-mail: msbadar@engr.ucr.edu; director_academic@salfiainstitute.edu.in

Integrated Science Association (ISA), Universal Scientific Education and Research Network (USERN), Riverside, CA, USA

Salfia Paramedical Institute, Darbhanga, India

S. Shamsi
Department of Computer Science, Jamia Millia Islamia, Jamia Nagar, New Delhi 10025, India
e-mail: shazmeen184002@st.jmi.ac.in

PGD (Big Data Analytics), School of Computational and Integrative Sciences, Jawaharlal Nehru University, New Delhi, India

J. Ahmed
Department of Computer Science, Jamia Hamdard University, Delhi, India
e-mail: jahmed2047@jamiahamdard.ac.in

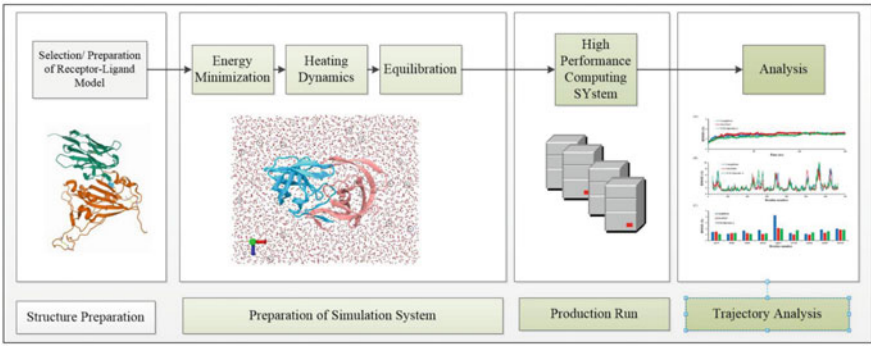
Md.A. Alam
Jamia Hamdard, New Delhi, India
e-mail: aalam@jamiahamdard.ac.in

School of Engineering Science and Technology, Department of Computer Science and Engineering, Jamia Hamdard, New Delhi, India

N. Rezaei (ed.), *Transdisciplinarity*, Integrated Science 5,
https://doi.org/10.1007/978-3-030-94651-7_7

Summary

Molecular dynamics (MD) is a computer simulation that deals with biological molecules, such as proteins and nucleic acid, and visualizes their movement in atoms and molecules. Computer simulation is executed with these atoms and molecules that are capable of interacting with each other over time and thereby can define the dynamic evolution of the system. MD simulation mimics the changes in biological molecules' structures over a given time, giving us atomic insights into the change in structure. This data helps us understand biological functions. These simulations give us comprehensive information about the fluctuations and flexibility of the proteins and nucleic acids under study. These approaches are applied to thoroughly study the organization and dynamics of biological molecules, their complexes, and conformational changes in proteins and nucleic acids. Many mysteries, on the femtoseconds scale, have been revealed through the study of these conformational changes. These methods are applied in chemical physics, materials science, and biophysics. MD simulations are often used in computational biology to generate a comprehensive understanding of interactions between proteins and their ligands and address how much these interactions are flexible and shape conformational changes in molecules when a particular mutation is introduced. Currently, it is being used to determine the tertiary structure of proteins from x-ray crystallography and NMR (or Nuclear Magnetic Resonance, a technique used in analytical chemistry for determining the structural properties and purity of samples) experiments.



The molecular dynamics simulation process.

The code of this chapter is 01001101 01101111 01101100 01100101 01100011 01110101 01101100 01100001 01110010 00100000 01100100 01111001 01101110 01100001 01101101 01101001 01100011 01110011 00100000.

1 Introduction

Paraphrasing Stephane Nappo, the Global Head of Information Security for Société Générale International Banking pole, this chapter attempts to address how the world of research in biological science is heading toward an incredibly revolutionary technology by amalgamating the focal aspects of Biological Science with Physical, Chemical, and Computer Science which aids it in manifesting the core characteristic of modern interdisciplinary research thus giving life to the technical outlook of Nappo's statement.

The ultimate goal of man is the sheer modernization of lifestyle for his comfort. Science aims at developing tools to fulfill this objective. Everything in the universe is evolving- be it nature or some human-made technology. The only difference is that nature evolves gradually while human-made technology is evolving at a much faster pace. Today the cure, antidotes, or therapies to almost all kinds of diseases are possible. Experimentation is important to gain insights into what the outcome of a particular trial could be. Determination of the structure of biomolecules under different conditions is a requisite to develop a proper understanding of the interactions between them to develop drugs for different diseases as per the need of the situation. It is experimentally impracticable to accurately determine the time-dependent behavior of biological molecules under real-time laboratory conditions, but developing a proper understanding of the dynamic behavior of complex biological processes such as protein folding and stability, conformational changes, ion transport, the central dogma of life, enzymatic reactions, etc. dynamic. is important for the development of drugs, therapies, and techniques that help cure diseases. Simulations help us to steer clear of this problem by using computational techniques. It also helps overcome time and cost issues in the long term. Experimental biological samples are pretty costly, and if a trial goes unsuccessful, the researchers have to bear a heavy loss in terms of time and money. Also, sometimes an experiment needs to be done within a limited period, take, for example, the development of drugs against coronavirus disease (COVID-19). Researchers need to do this task as soon as possible. However, considering the newness of this virus and the fact that not much is known about its reaction with different molecules in the human body, it is frustrating for the researchers to find the exact point of action of the virus experimentally because, one, it will take much time and keeping in mind the number of new cases that are emerging each day, the scientists cannot afford to waste time. Two, it will be very costly and risky to get real-life biological samples, and clinical trials take years to approve the validation of newly designed drugs. This is where simulation techniques come to the rescue. The interaction of the different biological molecules with the designed drug molecules can be seen using computational simulation tools, and the lead compounds can be separated. Then laboratory examination of only the screened compounds could be done further. This method saves time, money, and energy for the researchers, and the fact that further experimentation is done on the screened molecules only, there is a much higher chance of getting positive/desired results.

Simulation of biological molecules was unknown until as late as the 1950s, but within the next ten years from then, it was one of the hottest topics in the research world. The literal meaning of 'simulation', as we all know, is the imitation of an anticipated event. *Molecular Dynamic Simulation* is executed with computer techniques to apprehend the dynamicity of biomolecules. These methods visualize atoms and molecules when interacting with each other for a secure duration of time and analyze their physical movement and chemical interactions. Therefore, they help us realize the structure, fluctuations, flexibility, conformational changes, dynamics, and thermodynamics of simple biological molecules as well as their complexes. Understanding these complex biomolecular motions is doubtlessly pertinent to drug discovery [1]. The initial 'lock-and-key' mechanism of ligand binding proposed by Emil Fischer in 1890, in which a stationary, fixed receptor was assumed to house a small molecule without going through any conformational rearrangements, has now been forsaken to accept new binding models that consider not only the conformational changes but also the random motions of ligands and receptors [2–6], thus proving Richard Feynman's statement true. He was a Nobel Prize recipient (1965) in Physics and said, "All things are made of atoms, and that everything that living things do can be understood in terms of the jiggling and wiggling of atoms" [7]. Today, biophysics is a field devoted to comprehending the true essence of this jiggling and wiggling of biological molecules.

1.1 Aim

The goal of molecular dynamic simulation is to predict the behavior of atoms in a biological system and how they move as a time-dependent function, thereby providing the ultimate details concerning the atoms based on algorithms of physics that govern the interatomic interactions [8]. Through this, we hope to discern the properties of molecules concerning their structure and their conduct under different conditions. It serves as an important suffix to the lab experiments, saving time, cost, and labor of the scientists and bridges the gap between the latest technological advancements in the modern scientific community and the conventional experimental scientists. It aims at lowering the amount of guesswork and fittings traditional scientists make and helps them get an idea about the simulations that are difficult or unfeasible in the laboratory. We should always keep in mind that it is possible that one might not necessarily have a flawlessly realistic molecular model. However, the model should be able to portray the essential properties of physics and chemistry and also follow the concerned laws of mathematics along with possessing the correct biological attributes, and that should be enough.

1.2 Brief History

Alder and Wainwright first introduced Molecular Dynamic Simulation in 1957–1959 to understand hard spheres' interactions through thorough study. Even though

the first proper usage of simulation dates back to 1964 when Rahman et al. initiated developing real-world liquid argon. The numerical methods used for this process were developed much before, preceding the use of computers. In 1969, Barker and Watts first performed the Monte Carlo simulation of water, while McCammon et al. in 1977 performed the first MD protein simulation. The protein of interest was the bovine pancreatic trypsin inhibitor (BPTI). Duan and Kollman, in the 1990s, made an amazing revelation by discovering the folding mechanism of villin protein by applying techniques of molecular dynamics simulation, and this achievement is considered a landmark event of this field [9].

Now you must be wondering what Monte Carlo Simulation is? For that, we need to understand that there are two main classes of simulation techniques: the molecular dynamic (MD) simulation and Monte Carlo (MC) simulation. Moreover, there are other composite techniques that integrate the features of both these MD and MC depending upon the need of the research [10]. For a simulation of low-density systems like gas, where the molecules possibly get trapped in low-energy conformations, Monte Carlo simulations are preferable, while MD simulation is the choice technique for the simulation of liquids [11]. Further discussion on MC is beyond the scope of this chapter.

2 Concepts

Computer simulation for studying the dynamic behavior of molecules to comprehend the enigma behind the complexity of the biological world is a demanding task. It necessitates the need for optimally developed models capable of mimicking the cellular environment. These physical forces can simulate the laws of physics and thermodynamics and provide dynamicity to the model and heavy computations keeping in view the temporal aspect of the technique. Today, tools have been developed for molecular modeling, energy calculations, algorithms to simulate the real systems' chemical aspect, docking-scoring techniques, etc., thereby making the whole technique robust. To make the simulation naturalistic, the structure is placed in a "bath" of thousands of water molecules. Let us generate a fundamental idea about this incredibly amazing technology-enhanced technique.

2.1 Molecular Modeling

Molecular modeling is one of the fastest spreading techniques in computational biology, which encompasses all the tasks from visualization, derivation, manipulation, and representation of the structures of molecules keeping in view the physical and chemical properties that depend on these structures. As per recent studies, the modeled molecules should simulate their behavior, taking into account the equations of classical and quantum physics [12]. At present, the total number of entries in the UniProtKB/TrEMBL database is 184,998,855, while in PDB, it is

Table 1 The existence of proteins identified at different levels of information

a	1
b	2
c	3
d	4

(Prepared with data from UniprotKB/TrEMBL database as of July 25, 2020)

166891, which can be seen in the table given below, which has been taken from the UniProtKB/TrEMBL database.

Table 1 represents the number of entries of proteins in UniProtKB/TrEMBL database at different levels of its existence as of July 25, 2020.

MD simulation considers molecules as a ball-on-spring model. This model is apt to simulate the dynamic behavior of the molecules. Molecular modeling helps generate the structures of biomolecules by supplying the geometrical coordinates of biomolecules available as NMR or X-Ray crystallographic structures. However, if the ready-made structures are unavailable, one can easily deduce them by using computational algorithms and then assigning the x-, y- and z- coordinates to the molecules from the knowledge of their geometry. Three prime methods used for modeling are the ab-initio method, threading, and homology modeling.

2.2 Molecular Interaction and Force Field

MD simulation requires equations of motion for classical mechanics, which can in the simplest form be written as

$$m_i r_i = f_i$$

where $f_i = -\overline{\delta}\,\overset{\delta}{u}_{r_i}$

For this calculation, we should numerically know f_i as the force that acts on the atoms, which, in turn, results from a potential energy $U(r^N)$, where $r^N = (r_1, r_2, \ldots r_N)$ stands for the entire set of the geometrical 3 N coordinates of each atom.

For this potential energy calculation, we first need to develop a clear concept of its functional form, the force field. Force field can be understood as an empirical set of energy functions that helps us get an understanding of the energy associated with the interaction between atoms [13]. Typically, a force field is the summation of bonded and non-bonded terms or covalent and non- covalent interactions among the atoms and molecules as,

$$E_{\text{Total}} = E_{\text{Stretch}} + E_{\text{Bend}} + E_{\text{Torsion}} + E_{\text{Electrostatic}} + E_{\text{van der Walls}} + E_{\text{Hydrogen Bond}}$$

Now let us get a brief idea as to what these terms are:

- Bond stretching (E_{Stretch}) describes the energy of deformation of the bond length *w.r.t.* their equilibrium value. The energy near-equilibrium can be approximated by using harmonic potential, which does not allow the breaking of bonds [14]. The determination of the stretching force constant can be done using vibration spectroscopy.
- Angle bending (E_{Bend}) describes the deformation energy of the bond angles w.r.t. their equilibrium value. The energy near-equilibrium can be approximated by using harmonic potential. This force constant can be determined by vibration spectroscopic studies.

Torsional Term (E_{Torsion}) originates through space and accounts for the rotation of covalent bonds. This approximation of this term can be made with the help of a series of geometric functions.

- Fig. 1 is a representation of the non-bonded interactions that we just studied.
- Electrostatic term ($E_{\text{Electrostatic}}$) is evaluated using Coulomb's Law with the inclusion of partial charges, which are calculated by Quantum Mechanics. For better calculations, static partial charges and polarizable charges can also be considered as per one's needs.
- Van der Waals Term ($E_{\text{Van der Waals}}$)-It describes the interactive and repulsive interactions between atoms, in simpler terms, the interatomic forces. This term can be approximated by using Lennard Jones 12–6 potential, which can be thought of as a function of the distance between the centers of the two interacting atoms/molecules.
- Hydrogen Bond Term ($E_{\text{Hydrogen Bond}}$)—It describes the energy between atoms that have the potential to form hydrogen bonds. It is approximated by using 12–6 potential, which is similar to the Lennard Jones Potential, but the attractive interaction between atoms disappears faster in this case.

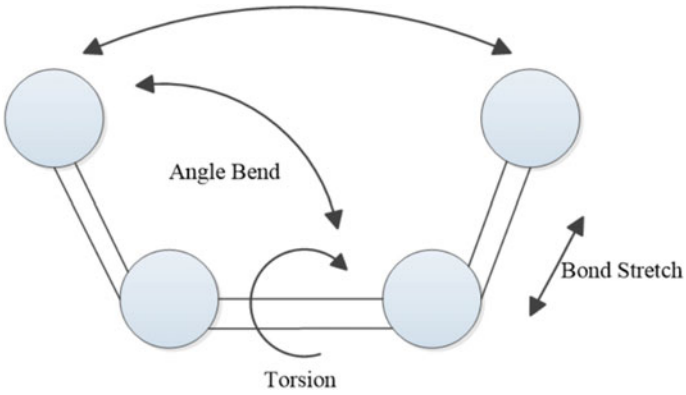


Fig. 1 Representation of the non-bonded interactions in a typical force field

- Cross terms Most interaction terms we just studied are generally not present independently in biomolecules but affect each other. Cross term accounts for all such interactions affecting others, including bend-bend, bend-torsion, stretch-bend, stretch-stretch, and stretch-torsion.

2.3 Periodic Boundary Conditions

Differential equations along with additional constraints called boundary conditions that are chosen for the approximation of a large system by using a corresponding smaller part called unit cell are known as periodic boundary conditions. Now let us understand this in simpler terms. Imagine the simulation of a system within a box-shaped container [9]. Since the system is physically fluid, it is very likely that a few particles flow out of the box due to its dynamic nature. We can apply a small trick to overcome this issue. We generate a replica of the box such that it covers the original box from all sides: whenever a particle tries to go out of the central box, there is a particle from the adjoining replica going at the same speed into the central box. The balance is maintained precisely.

Summarizing, periodic boundary conditions allow a simulation to deal with comparatively fewer particle numbers. With this operation, particles experience forces as if they are in the bulk of the liquid.

2.4 Langevin Dynamics

A real-world molecular system in a general and biomolecular system, in particular, is not likely to be present in a vacuum. Rather, it lies in such an environment, the cellular environment, where they constantly experience frictional forces. Jostling of biomolecules in such an environment causes perturbation of the system [15]. Langevin Dynamics, based on the Langevin Equation, a kind of stochastic differential equation, allows computational simulation methods to incorporate these effects. We can perceive it as an approach that necessarily imitates the solvent's viscosity but excludes its electrostatic and hydrophobic effect [9]. Suppose a system made up of N particles with mass M and coordinates X = X (t). We can use the Langevin Equation as follows:

$$MX = -\Delta U(X) - YMX + \sqrt{2\mathrm{Yk_B}\mathrm{TMR(t)}}$$

where the notations have their standard meaning.

2.5 Time Series Calculation

To study the dynamicity of a biomolecular system based on a temporal scale, getting a proper understanding of time-dependent statistical mechanics is a vital requirement because of the recently proposed algorithms for molecular dynamics and that dynamics critically capture equilibrium time-correlation functions, particularly those corresponding to transport coefficients [16].

Components comprising time series analysis include but are not limited to the root mean square deviation (RMSD), root means square fluctuation (RMSF), surface accessibility (SA), and the radius of gyration (RGYR) [9]. These calculations help us develop an idea about the biomolecular changes that occur gradually over time. System stability during simulation is calculated by root mean square deviation. Root mean square fluctuations provide us with an overview of the residue's flexibility understudy on a determined time scale. The radius of gyration, defined as the root mean square distance of the system from its center of mass, concerns system fluctuations. Surface accessibility applies the conformation of the biomolecule so that can describe the part of the biomolecule accessible to solvent.

2.6 MD Simulation Algorithms

To study the evolution of biological systems on a temporal scale, we will use Newton's motion laws:

$$F = \frac{dP}{dT} = m\frac{d^2r}{dt^2}$$

These classical equations of motion are integrated by applying the finite difference method. Finite difference methods are nothing but techniques applied to generate MD trajectories with continuous potential models. The basic idea behind this is that the complete integration is divided into smaller steps, so the total force of a given particle is calculated on a time scale as the vector sum of interaction between the particle being studied and other particles. Algorithms are available for integrating the equation of motion using the finite difference method, and the main assumption made by all the available algorithms is that the dynamic property obeys Taylor's theorem as follows [17–19]:

$$r(t+\partial t) = r(t) + v(t)\partial(t) + \frac{1}{2}a(t)\partial t^2 + \ldots$$

$$v(t+\partial t) = v(t) + a(t)\partial t + \frac{1}{2}b(t)\partial t^2 + \ldots$$

$$a(t+\partial t) = a(t) + b(t)\partial t + \ldots$$

where r = position, v = velocity (the first derivative of the position concerning time), and a = acceleration (the second derivative of the position concerning time).

i. Verlet algorithm

It is an algorithm commonly used for integrating the equation of motion [20]. It is a two-thirds order algorithm that applies Taylor series expansion for the position of molecule $r(t)$, one forward and the other reverse in time. Verlet algorithm employs a method that does not involve explicit velocities, and for this, it relates to an "explicit central difference method." The position of the previous step will be $r(t-dt)$, and to calculate a new position, we can write down the following equation:

$$r(t+\partial t) = r(t) + v(t)\partial(t) + \frac{1}{2}a(t)\partial t^2$$

$$r(t-\partial t) = r(t) - v(t)\partial(t) + \frac{1}{2}a(t)\partial t^2$$

If we add the above two equations, we obtain

$$r(t+\partial t) = 2r(t) + r(t-\partial t) + a(t)\partial t^2$$

The advantages of this algorithm are that it is straightforward and self-starting, and the new positions can easily be obtained from the current and previous positions. Another advantage is that it requires less computer memory.

The disadvantage is that due to the lack of an explicit velocity term, it is difficult to obtain the velocity at the current position until the position has been computed for the next step.

ii. Velocity Verlet Algorithm

As we mentioned earlier, the Verlet method does not involve velocity. Though it is unnecessary during the actual simulation process, it is needed for the calculation of the kinetic energy in order to test the total energy conservation [21]. The advantage is that this step helps verify whether the simulation process is proceeding correctly or not. It shows a better use of the basic Verlet algorithm discussed above. The calculation of positions, velocities, and accelerations are as follows:

$$r(t+\partial t) = r(t) + v(t)\partial t + \frac{1}{2}a(t)\partial t^2$$

$$r(t+\partial t) = v(t) + \frac{1}{2}[a(t) + a(t+\partial t)]\partial t$$

Because at a particular time, all three parameters: position, velocity, and acceleration are considered, there is no compromise on the precision.

The advantage is the same as the Verlet Algorithm, i.e., it is storage efficient. The disadvantage is related to the error we can find in the range of Δt^2.

iii. Leapfrog algorithm

It was developed to control the error obtained associated with the method of Velocity Verlet. The leapfrog method calculates velocities and positions at interleaved time intervals in a way that the position r is integral time step $(t+\delta t)$ and velocity incorporates an extra half step and is defined as integral time plus a half step $(t+1/2\delta t)$ [22, 23]. Therefore, velocities do leap over positions, and vice versa. We can write it as:

$$r(t+\partial t) = r(t) + v\left(t+\frac{1}{2}\partial t\right)\partial t$$

$$v\left(t+\frac{1}{2}\partial t\right) = v\left(t-\frac{1}{2}\partial t\right) + a(t)\partial t$$

It is a way to explicitly calculate velocities, but there is no room to simultaneously calculate positions.

As a result, it uses a different formula to estimate total energy at any point in time:$v(t) = \frac{1}{2}\left[v\left(t-\frac{1}{2}\partial t\right) + v\left(t+\frac{1}{2}\partial t\right)\right]$.

iv. Beeman's algorithm

It is very close to the method of Verlet and can be described by:

$$r(t+\partial t) = r(t) + v(t)\partial t + \frac{2}{3}a(t)\partial t^2 - \frac{1}{6}a(t-\partial t)\partial t^2$$

$$v(t+\partial t) = v(t) + v(t)\partial t + \frac{1}{3}a(t)\partial t + \frac{5}{6}a(t)\partial t - \frac{1}{6}a(t-\partial t)\partial t$$

It allows more accurate treatment of velocities and energy [24]. However, even with this method, the calculations are not perfectly made, but instead, it is a computationally expensive algorithm and therefore is not practical in the real world.

A few of the many software used for MD simulation are AMBER (Assisted Model Building with Energy Refinement), CHARMM (Chemistry at HARvard Molecular Mechanics), GROMOS (GROningen MOlecular Simulation), and NAMD (Nanoscale Molecular Dynamics).

3 Method

There are several softwares available for performing the molecular dynamic simulation of biomolecules like GROMACS, Open Babel, VMD, UCSF Chimera, etc. We can select the software of our choice and perform the task but always remember that different software uses different force fields.

MD simulations are performed in three main steps, which further consist of smaller steps: model selection; energy minimization, heating, and equilibration; and production run and analysis.

If we talk about Chimera, MD simulation can be thought of as a link to minimization and molecular dynamics routines provided by Molecular Mechanics Toolkit (MMTK), which is incorporated with it. Standard residues are assigned Amber parameters, while non-standard residues are assigned parameters using Chimera's Antechamber module.

i. Model Selection: A model system of interest should be chosen. Most of the time, complete models are available for use, but in case complete models are unavailable, the missing segments are secured, and the protonation states are conditioned. All atoms of interest should be considered and included in this step because models not included here will be ignored. The prepared molecule should be read in the pdb and psf files.

   Obtaining Files- Simulations generally begin with a crystal structure one can obtain from the Protein Data Bank. The information about atoms of use is the atom names (N, C, CA), Residue name and ID, Occupancy, Coordinates, Beta factor or Temperature Factor, and Segment ID.

ii. Energy minimization, heating, and equilibration: This step includes the equilibration of the model structure that depends on the force field of choice (T = 0). We also decide the number of equilibration steps (default 5000). Then the systemis heated by rescaling the velocities, and its stability is ensured until the system's properties stop changing with time and the system reaches a particular temperature. Preparing the system for energy minimization It includes energy searching by force-field methods, and accordingly, generating low-energy conformations. Different strategies can be used to complete the so-called minimization step, a few of which are listed below:

   - Steepest Descent—Used for highly restrained systems.
   - Conjugate Gradient—Used for large systems; applies intelligent choices of search direction; efficient.
   - Broyden–Fletcher–Goldfarb–Shanno (BFGS)—Quasi-newton variable metric method.
   - Newton–Raphson Method—Calculates both slopes of energy as well as rate of change.

Periodic Boundary conditions should be used whenever a solvent box is added. The cut-off distance should not exceed half of the smallest box dimension for maintaining periodic boundary conditions.

Fixed Atoms help specify whether one needs to freeze some atoms in a position during the calculations. Such atoms to be frozen in place are highlighted by the selection, but one must always remember that all atoms in the desired model will be considered in the energy calculations, whether they are fixed or not.

Translation Remover aids in subtracting out a global translational motion during MD and also decides which steps, by default the first, third, fifth, etc., through the end.

Rotation Remover aids in subtracting out a global rotational motion during MD and also decides which steps, by default the first, third, fifth, etc., through the end.

Topology files: It assumes that each element contains different atoms and corresponds to molecular orbital environments that result from the interaction between the atoms and their charges and orbitals. Topology files include information about atoms, atomic charges and orbitals, and atom representations in elements.

Parameter files force constants necessary to predict the bond energy, non-bonded interactions (Van der Waals and electrostatic), angle energy, and torsion energy are available in these files along with parameters proposed for energy calculations.

Solvation Some biochemical processes take place in aqueous systems and therefore, the impact of solvation is significant on the determination of molecular conformation, binding energies, and electronic properties [25]. There are two methods of model solvation: the explicit method works on solvents as being explicitly introduced to the system, while the implicit method models the solvent molecule as a continuum dielectric.

iii. Production Run and Analysis: The model is then simulated under desired conditions of NVT, NPT, etc. Finally, a production run is performed for a relevant time to get the output trajectories. The 'include production phase' helps us decide whether to include the production MD in a phase and if so, how many steps should be included. We also need to mention the time steps at which we write the trajectory files, which are further analyzed to obtain the desired properties of interest [26].

The steady advancement of potential computational sampling methods now lets us carry out the simulation process on a time scale of seconds to microseconds and even milliseconds. Here it should be significantly noted that in simulation, these millisecond scales are believed to be enormous that contradicts the in vitro experiment because, in a computer simulation, coordinates are produced at the femtosecond level. When moving from femtoseconds to milliseconds scale, we have many conformations emerged to unravel biological problems that otherwise remain unsolved.

4 Applications

MD simulations have a wide broad of applications not only in the field of biological science but in any field one can imagine ranging from physics, chemistry, biology to climatology and meteorology, video games, to film industries. Let us focus on the applications of MD simulations in biological complexes.

i. Determination of Structures and Movements of Biomolecules

   As already mentioned, we now know that the most common application of MD simulation in biomolecules is to study, analyze and mimic the flexibility, movements, and interactions of and among the different proteins. Structures determined by experimental studies by X-Ray Crystallography or NMR studies reveal only an average approximation of what the real thing could be. However, with computational simulation techniques, one could make an even more precise approximation of the types of structural fluctuations the molecules undergo. By just scrutinizing a simulation of these structures, one can quantify the movements of different regions of the molecule at equilibrium and the types of structural fluctuations that occur [27]. Such simulations also can show the dynamic behavioral properties of water molecules and salt ions, the effects of which are often critical for the proper functioning of protein and also for ligand binding.

ii. Assessment of accuracy and Refinement of modeled structures
   This method can also be used to assess the accuracy of already modeled structures or even to refine the structures built using molecular modeling techniques or experimentally in the lab. For example, it is frequently seen that experimentally determined X-Ray crystal structures are refined by a computational MD simulated annealing protocol and fit the model to the experimental data even more precisely while simultaneously maintaining a physically stable structure [28]. One advantage of this approach is that it has been shown to control model errors otherwise present. Let us consider another example. A membrane protein may suffer from artifacts due to the absence of a lipid bilayer or crystal structure suffers from such errors due to the crystal lattice packing but owing to the lucidity of the near accuracy of the simulated structures. It is now possible to correct such artifacts by performing a simulation of inappropriate solvation environments as per the requirements of the structures one is working with. Though MD simulations are extremely useful in the refinement of existing homology models, several attempts to do this have been unsuccessful [29]. MD simulations have also been applied to retrieve ensembles of conformations, against a single structure, from NMR data [30]. In each of these cases, the molecular mechanic's force field is augmented by terms that have to be taken from experimental data, which results in lower energy for structures (or structural ensembles) that are more suitable.

iii. The flexibility of Molecules
The flexibility directly modulates the association of a molecule with its neighboring atoms, molecules, and ions, and thus plays an active role in cellular function. We have already studied that the molecular dynamic system gives us clear insights into the dynamic evolution of any system. It can also be seen as reflecting its flexibility to an extent. Recently developed techniques such as Anisotropic Network Model (ANM), Elastic Network Model (ENM), Principal Component Analysis (PCA), among others, have allowed the extrapolation of prime motions in the system [31].

iv. Another interestingly important use of MD simulation is to ascertain the mechanism in which a biomolecular system will respond to perturbation. Say, for example, someone changes the molecular environment of the protein like the salt concentration or lipid composition, or adds a ligand where there was originally no ligand present or replaces a bound ligand with a different ligand, or changes the amino acid residues present in a particular protein by mutating them or by changing the protonation state of the amino acid [32, 33]. In all the cases mentioned above, simulations help one in getting a thorough understanding of the system under study. One thing to be kept in mind while performing such simulations is that one should perform it several times by using both perturbed and unperturbed systems to get clear insights into the consistent differences in the results and thus ascertain one's results.

v. Analysis of results of MD simulation of different systems helps one to answer such questions about the role of structure, flexibility, and the interactions among different biomolecules that are experimentally very difficult to address. Since simulations can occur at the scale of femtoseconds, we can observe such biological processes that occur in a jiffy, like the order in which the substructures form during protein folding [34, 35]. One can also perform a thorough study of processes like ligand binding, protein folding, conformational changes, membrane transport, etc. They also help us understand the factors controlling ligand binding and dissociation kinetics, the process of assembly of disordered proteins to form fibrils [36, 37]. Simulations may capture an entire process in one go, or they may capture it in parts, which can then be used to reconstruct the entire process [38–41].

vi. Modeling of Drug Receptor Interactions
Experimental studies help us determine the 3D conformation of ligands. Ligands, also known as drug molecules, bind to receptors, that do not have a known structure and for this, cannot be targeted directly. Instead, ligand-receptor interactions are candidates amenable to drug design. It is significant to know the structure of both the receptor and ligands before carrying out further simulations by making drug design modifications. Recent work has emphasized structure-based drug design (SBDD) and ligand-based drug design (LBDD) approaches for modern drug discovery. SBDD uses the 3D structure of drug target and free energy techniques for the task of approximating the absolute and relative binding free energy (RBFE). RBFE or alchemical approaches allow the application of MD simulations to

the initially sampled ligand and consider this first application for calculations of binding free energy differences between structurally similar ligands. This is immensely important from the point of view of drug discovery. In short, simulations are used to determine the molecule's location to bind to its receptor and how it changes the binding strength and affinity of molecules that bind elsewhere. This information, along with other geometrical, physical, chemical, and thermodynamic properties, is used to alter the structure as many times as possible to design a drug that fulfills one's needs. Once this computational task is done, the experimental scientists take over, and after its testing and approval, clinical trials take place, and if it passes the clinical trial, the drug is ready to be launched in markets.

vii. Gives Insights into Molecular Interactions on a Temporal Scale
Molecular dynamics simulation generates pictures of atomic-level details of the dynamic evolution of the biomolecular system. This property clubbed with temporal scale for MD simulations, enables us to predict different feasible cellular interactions and behaviors based on which modifications in existing structures can be made and seen and is immensely helpful in studying the properties of these samples, which can be used as potent drugs in the market.

viii. Docking
Docking is the process by which two or more molecular structures orient themselves so that they bind to each other to form a stable complex. MD simulation techniques offer approaches for monitoring different types of interactions, including DNA-ligand, DNA–protein, protein–ligand, and protein–protein. More interestingly, they enable us to probe emerging types of molecular interactions that participate in the formation of more complex structures. For this reason, docking is performed first, and then MD simulation is done so that one can know the effect of interactions on a temporal scale. Docking software like Glide, GOLD, AutoDock, etc., make use of different algorithms to calculate a docking score based on different parameters like the surface of contact, electrostatics, etc. [42], and a good score is considered as one which has a good binding affinity. Then, experimental and clinical studies are necessary to test the findings in a realistic platform, although it is a challenging task in terms of time and cost to obtain all the docked structures. This calls a need to filter out structures based on thermodynamic and structural characteristics as explained by MD simulation techniques.

ix. Protein folding
It is an interesting topic in biology. Though the 3D structure of proteins has been studied pretty well, less is known about protein folding. MS simulation could help to identify the folding mechanism using computational power. For example, the MD simulation of a sub-domain of villin protein was done, which gave significant insights and a glimpse of hope toward the proper understanding of the protein folding mechanism [43].

x. To understand how mutation affects Interactions
To understand and realize the effects of a given amino acid and its binding, a mutation is introduced to the residue and the difference in the simulated trajectories both before and after the process is studied well. The ligand can also undergo refinements and modifications so that its affinity is improved and insights into its structural properties and interactions become certified [9].

5 Future Scope

As students of science, we all know that the macroscopic properties of elements owe a great deal to the time-dependent underlying microscopic properties and interactions among atoms and molecules. Molecular dynamics simulation has unbolted an incredibly huge number of doors for the research enthusiasts in biological sciences, from the study of protein folding to the physics and chemistry behind the interactions among biomolecules, molecular docking to drug design, etc. The entire technique of MD simulation relies solely on the trustworthiness of the model, force field calculations, and the thermodynamic property calculation, and the ability of particular software to be able to mimic a process with as much reality as possible. Even though so many computer simulation techniques have been developed, there is always the scope of improvement, and simulation techniques do not show cent percent accuracy in their results, so better and more accurate techniques can always be developed. With better results of the simulation process, the foundation for future studies on ion exchange is being laid. Also, we need to develop more robust algorithms and one with a shorter number of steps. There is also scope for the development of lightweight and free software. Also, algorithms that are computationally less intensive are the need of the hour.

6 Conclusion

MD simulation technique is now more than sixty years old, yet it is such an incredibly amazing technology that there is still much excitement in the scientific community about this technology, and also it has maintained its spot in the limelight since then. But it was only recently that MD achieved time scales compatible with that of the biological system. Today, conformational changes and assembly of proteins and ligand–protein or protein–protein or ligand-ligand interactions can be studied with such ease with effective simulation. It also provides knowledge about the interactions at the atomic level, which directly affect the functions and behavior of the molecule, uncovers how different energy components make different contributions to molecular binding and stability. Therefore, it can rightfully be said that molecular dynamic simulation is a technique which is the best instance of

productive research being carried out where scientists from the background of Physics, Chemistry, Biology, and Computer Science come forward and work together on a common platform and is thus an exemplary technique developed in the era of interdisciplinary research.

Core Messages

- Molecular dynamic (MD) simulation offers computer-aided techniques to apprehend the dynamic behavior of biomolecules by visualizing atoms and molecules when interacting with each other over a period of time and analyzing their physical movement and chemical interactions.
- MD simulation aims to predict the behavior and movement of atoms as a time-dependent function with the hope of comprehending the properties of molecules concerning their structure and conduct under different conditions.
- Though Alder and Wainwright first introduced MD simulation in 1957–1959, the first simulation dates back to 1964 when Rahman et al. initiated developing real-world liquid argon.

Acknowledgment We would like to acknowledge Director (Admin), **Mr. Syed Sharique Hussain**, Salfia Paramedical Institute Darbhanga, (SPI) Darbhanga, (director_admin@salfiainstitute.edu.in), **Dr. Naima Saman** (pursuing MD), **Dr. Syed Abdul Hakeem**, **Mrs. Siddiqua Khatoon**, all members of Farid/Salfia family, Faridia Hospital, **Dr. Faiyaz Ahmad**, Chairman, Madhubani Medical College, Madhubani, **Dr. Faraz Fatmi** (JD(U)), **Mr. Ashraf Ali Fatmi** (JD(U)), **Dr. Shakeel Ahmad** (Congress), **Mr. Salman Akhtar**, Chief Functionary, Nalanda Minority Educational and Welfare Trust, **Dr. Muhammad Nezamuddin (Jawed)** and **Dr. Arifa Nishat**, Ruston, LA, USA, **Mr. Prashant Kishore**, (I-PAC), **Dr. David K Mills** and **Dr. Thomas C Bishop** (Louisiana Tech University, Ruston, LA, USA), **Dr. Manzoor Alam**, Chairman, AIMC, Brother Aman, Aligarh Bachchon Ka Ghar, Aligarh, **Hafiz Hassan Badar**, Secretary, Hazrat Umar Farooque Academy, Nazra-Ranipur, **Mrs. Shahnaz Badar**, and motivator for Abdur Razzaque Family, **Mr. Saeeduzzafar**, (IDB Jeddah) and my life partner, **Mrs. Rana Kamal Sufian**.

References

1. Durrant JD, McCammon JA (2011) Molecular dynamics simulations and drug discovery. BMC Biol 9(1):1–9
2. Ma B, Kumar S, Tsai C-J, Nussinov R (1999) Folding funnels and binding mechanisms. Protein Eng 12(9):713–720
3. Ma B, Shatsky M, Wolfson HJ, Nussinov R (2002) Multiple diverse ligands binding at a single protein site: a matter of pre-existing populations. Protein Sci 11(2):184–197
4. Fischer E (1894) Einfluss der configuration auf die wirkung der enzyme. Ber Dtsch Chem Ges 27(3):2985–2993
5. Teague SJ (2003) Implications of protein flexibility for drug discovery. Nat Rev Drug Discovery 2(7):527–541

6. Tsai CJ, Kumar S, Ma B, Nussinov R (1999) Folding funnels, binding funnels, and protein function. Protein Sci 8(6):1181–1190
7. Feynman RP, Leighton RB, Sands M (2011) The Feynman lectures on physics, Vol. I: the new millennium edition: mainly mechanics, radiation, and heat, vol 1. Basic books
8. Hollingsworth SA, Dror RO (2018) Molecular dynamics simulation for all. Neuron 99 (6):1129–1143
9. Karaa WBA (2015) Biomedical image analysis and mining techniques for improved health outcomes. IGI Global
10. González M (2011) Force fields and molecular dynamics simulations. École thématique de la Société Française de la Neutronique 12:169–200
11. Jorgensen WL, Tirado-Rives J (1996) Monte Carlo versus molecular dynamics for conformational sampling. J Phys Chem 100(34):14508–14513
12. Ishida H, Agag T (2011) Handbook of benzoxazine resins. Elsevier
13. Mayo SL, Olafson BD, Goddard WA (1990) DREIDING: a generic force field for molecular simulations. J Phys Chem 94(26):8897–8909
14. Pearlman DA, Kollman PA (1991) The overlooked bond-stretching contribution in free energy perturbation calculations. J Chem Phys 94(6):4532–4545
15. Tamar S (2002) Molecular modeling and simulation. Interdisciplinary applied mathematics. Springer, NY
16. Gullingsrud JR, Braun R, Schulten K (1999) Reconstructing potentials of mean force through time series analysis of steered molecular dynamics simulations. J Comput Phys 151(1):190–211
17. Akimov AV, Kolomeisky AB (2011) Recursive taylor series expansion method for rigid-body molecular dynamics. J Chem Theory Comput 7(10):3062–3071
18. Kutteh R (1999) New methods for incorporating nonholonomic constraints into molecular dynamics simulations. J Chem Phys 111(4):1394–1406
19. Streett W, Tildesley D, Saville G (1978) Multiple time-step methods in molecular dynamics. Mol Phys 35(3):639–648
20. Grubmüller H, Heller H, Windemuth A, Schulten K (1991) Generalized Verlet algorithm for efficient molecular dynamics simulations with long-range interactions. Mol Simul 6(1–3):121–142
21. Martys NS, Mountain RD (1999) Velocity Verlet algorithm for dissipative-particle-dynamics-based models of suspensions. Phys Rev E 59(3):3733
22. Toxvaerd S (1991) Algorithms for canonical molecular dynamics simulations. Mol Phys 72 (1):159–168
23. Van Gunsteren WF, Berendsen HJ (1988) A leap-frog algorithm for stochastic dynamics. Mol Simul 1(3):173–185
24. Van Gunsteren W, Berendsen H (1982) Algorithms for Brownian dynamics. Mol Phys 45 (3):637–647
25. Rosenthal SJ, Jimenez R, Fleming GR, Kumar P, Maroncelli M (1994) Solvation dynamics in methanol: experimental and molecular dynamics simulation studies. J Mol Liq 60(1–3):25–56
26. Jabbarzadeh Kaboli P, Ismail P, Ling K-H (2018) Molecular modeling, dynamics simulations, and binding efficiency of berberine derivatives: a new group of RAF inhibitors for cancer treatment. PLOS ONE 13(3): e0193941
27. Miao J, Hodgson KO, Sayre D (2001) An approach to three-dimensional structures of biomolecules by using single-molecule diffraction images. Proc Natl Acad Sci 98(12):6641–6645
28. Terwilliger TC, Read RJ, Adams PD, Brunger AT, Afonine PV, Grosse-Kunstleve RW, Hung L-W (2012) Improved crystallographic models through iterated local density-guided model deformation and reciprocal-space refinement. Acta Crystallogr D Biol Crystallogr 68(7):861–870
29. Mirjalili V, Feig M (2013) Protein structure refinement through structure selection and averaging from molecular dynamics ensembles. J Chem Theory Comput 9(2):1294–1303

30. Lindorff-Larsen K, Best RB, DePristo MA, Dobson CM, Vendruscolo M (2005) Simultaneous determination of protein structure and dynamics. Nature 433(7022):128–132
31. Bakan A, Bahar I (2009) The intrinsic dynamics of enzymes plays a dominant role in determining the structural changes induced upon inhibitor binding. Proc Natl Acad Sci 106 (34):14349–14354
32. Fribourg M, Moreno JL, Holloway T, Provasi D, Baki L, Mahajan R, Park G, Adney SK, Hatcher C, Eltit JM (2011) Decoding the signaling of a GPCR heteromeric complex reveals a unifying mechanism of action of antipsychotic drugs. Cell 147(5):1011–1023
33. Peng Y, McCorvy JD, Harpsøe K, Lansu K, Yuan S, Popov P, Qu L, Pu M, Che T, Nikolajsen LF (2018) 5-HT2C receptor structures reveal the structural basis of GPCR polypharmacology. Cell 172(4): 719–730. e714
34. Lindorff-Larsen K, Piana S, Dror RO, Shaw DE (2011) How fast-folding proteins fold. Science 334(6055):517–520
35. Snow CD, Nguyen H, Pande VS, Gruebele M (2002) Absolute comparison of simulated and experimental protein-folding dynamics. Nature 420(6911):102–106
36. Dedmon MM, Lindorff-Larsen K, Christodoulou J, Vendruscolo M, Dobson CM (2005) Mapping long-range interactions in α-synuclein using spin-label NMR and ensemble molecular dynamics simulations. J Am Chem Soc 127(2):476–477
37. Nguyen HD, Hall CK (2004) Molecular dynamics simulations of spontaneous fibril formation by random-coil peptides. Proc Natl Acad Sci 101(46):16180–16185
38. Bernardi RC, Melo MC, Schulten K (2015) Enhanced sampling techniques in molecular dynamics simulations of biological systems. Biochimica et Biophysica Acta (BBA)-General Subjects 1850(5): 872–877
39. Harpole TJ, Delemotte L (2018) Conformational landscapes of membrane proteins delineated by enhanced sampling molecular dynamics simulations. Biochimica Et Biophysica Acta (BBA)-Biomembranes 1860(4):909–926
40. Hertig S, Latorraca NR, Dror RO (2016) Revealing atomic-level mechanisms of protein allostery with molecular dynamics simulations. PLoS Comput Biol 12(6):e1004746
41. Schwantes CR, McGibbon RT, Pande VS (2014) Perspective: markov models for long-timescale biomolecular dynamics. J Chem Phys 141(9): 09B201_201
42. Kroemer RT (2007) Structure-based drug design: docking and scoring. Curr Protein Pept Sci 8(4):312–328
43. 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

Mohammad Sufian Badar PhD has served as a Teaching Faculty in the Department of Bioengineering at the University of California, Riverside, CA, USA. He served as an Analytics Architect in CenturyLink for more than a year in Denver, CO, USA. Currently, he is an Director (Academic), Salfia Paramedical Institute Darbhanga, India, (www.salfiainstitute.edu.in). He possesses an excellent academic record with an MS degree in Molecular Science and Nanotechnology and Ph.D. in Engineering from Louisiana Tech University Ruston, LA, USA, respectively. Before joining the Ph.D. program at Louisiana Tech University, he graduated with an MSc in Bioinformatics from Jamia Millia Islamia University, New Delhi, India. Dr. Sufian has over 14 years of teaching, research, and industry experience. He has published his research in conferences and high reputed International journals. He has authored many chapters in the area of Artificial Intelligence/Machine Learning, Blockchain/IoT. Currently, he is Editing two books: 1. Diagnosis & Analysis of

COVID-19 using Artificial Intelligence and Machine Learning based Techniques, 2. COVID- 19: Cause, Transmission, Diagnosis, and Treatment. He has developed an algorithm for Face Detection, Recognition, and Emotion Recognition. He is currently in the process of developing a device that, using Biosensors, can correlate the physiology of the human body with the emotion recognition algorithm, giving us a clear measure of the amount of stress hormones in the body. Currently, my group and I have developed an ML model which predicts Covid infection based on symptoms only, and we are now working on increasing the accuracy of our model.

Shazmeen Shamsi has an M.Sc. degree in Bioinformatics from Jamia Millia Islamia, New Delhi, India, 2020 and B.Sc. (Honors) in Botany from the University of Calcutta, West Bengal, India, in 2018. She has a Diploma in PC Application, and certifications in "Blockchain & Applications" and "Python Programming". She has worked as a research intern in the School of Computer and System Sciences, Jawaharlal Nehru University, New Delhi, and has co-authored three book chapters. She is presently a PGD (Big Data Analytics) student at Jawaharlal Nehru University, New Delhi, and is also working as a Research Trainee under Dr. Mohammad Sufian Badar, Senior Teaching Faculty, Department of Bioengineering, University of California, Riverside, CA, USA. Her current research interests include Artificial Intelligence, Internet of Things, Network Security, Data Analytics, Social Networking, and Simulation Techniques used in Bioinformatics.