Molecular Dynamics Simulations: Concept, Methods, and Applications

Authors:

Mohammad Sufian Badar, et al.

Contents

Abstract:

Molecular dynamics (MD) is a computer simulation method used in the theoretical study of biological molecules, such as proteins and nucleic acid, to analyze the physical movements of the constituent atoms and molecules. In the computer simulation, these atoms and molecules interact over time and give a sense of the dynamic evolution of the system. MD simulation mimics the changes in the structures of biological molecules over a given period of time, giving us atomic insights about the change in structure. This data helps us understand biological functions. These simulations give us detailed information about the fluctuations and conformational changes of the proteins and nucleic acids under study. These methods are applied to thoroughly study the organization and dynamics of biological molecules, their complexes, and the conformational changes of 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 study protein-protein interaction, protein-ligand docking, the effects of mutation on interactions, protein folding, and flexibility of the biological molecules. Currently, it is being used to determine the tertiary structure of proteins from x-ray crystallography and NMR experiments.

Keywords: Molecular Dynamic Simulation, Conformational Changes

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

-Stephane Nappo

Paraphrasing Stephane Nappo, the Global Head of Information Security for Société Générale International Banking pole since 2011, we shall learn in this chapter, how the world of research in biological science is heading towards 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.

1. Introduction

The ultimate goal of man is the sheer modernization of lifestyle for his comfort. Science aims at developing tools to fulfil this objective. Everything in the universe is evolving- be it nature or some man-made technology. The only difference is that nature evolves gradually while man-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 structure of biomolecules under different conditions is necessary in order to develop a proper understanding of the interactions between them so as to develop drugs for different diseases as per the need of the situation. It is experimentally not feasible to precisely calculate the time-dependent behavior of biological molecules under real-time laboratory conditions, but developing a proper understanding of the complex dynamic biological processes such as protein stability and folding, conformational changes, ion transport, central dogma of life, enzymatic reactions, etc 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 span of time, take for example the development of drugs against Covid-19. Researchers need to do this task as soon as possible, but 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 a lot of time and keeping in mind the number of new cases that are emerging each day, the scientists just cannot afford to waste time, and 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 by 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 as well as the energy of 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 kind of 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 an approach that uses computer techniques to apprehend the dynamicity of biological molecules by allowing the atoms and molecules to interact for a secure period of time and analyzing their physical movement and chemical interactions. They provide detailed information on the structure, fluctuations and conformational changes, dynamics, and thermodynamics of biological molecules and 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 motionless, fixed receptor was assumed to house a small molecule without going through any conformational rearrangements, has now been abandoned to accept new binding models that consider not only the conformational changes but also the random motions of receptors and ligands [2-6], thus proving Richard Feynman's statement true. He was a recipient of the Nobel Prize (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 devoted field which aims at comprehending the true essence of this jiggling and wiggling of biological molecules.

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 understand the properties of molecules

with respect to their structure and their conduct under different conditions. It serves as an important suffix to the lab experiments thus saving time, cost, and labour 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 impossible in the laboratory. We should always keep in mind that it is possible that one might not necessarily have a flawlessly realistic molecular model but 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.

2. Brief History

Molecular Dynamic Simulation was first introduced by Alder and Wainwright in 1957-1959 to study interactions of hard spheres. Even though proper simulation was first performed in 1964 by Rahman et.al. in using a realistic potential for liquid argon, the numerical methods used for this process were developed much before, preceding the use of computers as well. In 1969, Barker and Watts first performed the Monte Carlo simulation of water, following which McCammon, *et al* in 1977, performed the first MD protein simulation of the bovine pancreatic trypsin inhibitor (BPTI). Duan and Kollman in the 1990s made an amazing revelation by discovering the folding mechanism of villin protein using molecular dynamics simulation and this achievement is considered as 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, in addition to which there are other composite techniques which 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 technique of choice for the simulation of liquids [11]. Further discussion on MC is beyond the scope of this chapter.

2. Concepts

Computer simulation for the study of the dynamic behavior of molecules to understand the enigma behind the complexity of the biological world is a demanding task. It necessitates the need of optimally developed models capable of mimicking the cellular environment, physical forces that 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 modelling, energy calculations, algorithms to simulate the chemical aspect of the real systems, docking-scoring techniques, etc, thereby making the whole technique robust. To make the simulation realistic, 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.

1. Molecular Modelling

Molecular Modelling is one of the fastest spreading techniques in the field of 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 dependent on these structures. As per recent studies, the modelled molecules should be able to simulate their behavior taking into account the equations of classical and quantum physics [12]. Figure 1.a and 1.b show graphs of the number of entries of structures that have been successfully modelled till date and stored in UniProtKB/TrEMBL and RCSB PDB database respectively. We see that at present the total number of entries in UniProtKB/TrEMBL database is 184,998,855 while in PDB it is 166891. Table 1(information taken from UniProtKB/TrEMBL database) shows the existence of proteins identified at different levels as on 25th July 2020:

Protein Existence	Number of Entries
Evidence at protein level	169117
Evidence at transcript level	1314759
Inferred from homology	46812266
Predicted	136702713
Uncertain	0

Table 1: Table representing the number of entries of proteins in UniProtKB/TrEMBL database at different levels of its existence.

MD Simulation considers molecules as a ball-on-spring model. This model is apt to simulate the dynamic behavior of the molecules. Molecular Modelling helps generate the structures of biomolecules by supplying the geometrical coordinates of biomolecules available as NMR or X-Ray crystallographic structures but in case 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 major methods used for modelling are the ab-initio method, threading, and homology modelling.

2. Molecular Interaction and Force Field

MD simulation requires the step-by-step numerical solution of the classical equations of motion, which can in the simplest form be written as

$$m_i r_i = f_i$$

$$\text{where,} \quad f_i = -\frac{\delta}{\delta_{r_i}} u$$

For this calculation we should numerically know the forces f_i acting on the atoms, which are in turn generally derived from a potential energy U (r^N), where $r^N = (r_1, r_2, \ldots r_N)$ stands for the complete set of the geometrical 3N atomic coordinates.

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 related to 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,

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

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

Bond Stretching ($E_{Stretch}$)- It 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})- It 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 ($E_{Torsion}$)-It originates through space and accounts for the rotation of covalent bonds. This approximation of this term can be done with the help of a series of geometric functions.

Figure 2 is a representation of the non-bonded interactions that we just studied

Electrostatic Term ($E_{Electrostatic}$)-It is calculated using Coulomb's Law with the inclusion of partial charges which are calculated by Quantum Mechanics. For better calculations, static partial charges and polarisable charges can also be taken into consideration as per one's needs.

Van der Waals Term ($E_{Van\ der\ Waals}$)-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 ($E_{Hydrogen\ Bond}$)- 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.

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. Cross terms include stretch-stretch, stretch-bend, bend-bend, bend-torsion, stretch torsion types of interactions.

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, there is a high possibility 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. What happens now is that whenever a particle tries to move out from the central box another will enter from the adjacent replica with the same speed in order to maintain a balance in the system.

Summarizing, periodic boundary conditions enable a simulation to be performed using a relatively small number of particles in such a way that particles experience forces as if they are in bulk fluid.

4. Langevin Dynamics

A real-world molecular system in a general and biomolecular system, in particular, is not likely to be present in vacuum rather it lies in such an environment, here, the cellular environment where they constantly experience frictional forces. Jostling of biomolecules in such an environment causes perturbation of the system [15]. Langevin Dynamics which is based on 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 viscosity of the solvent but excludes its electrostatic and hydrophobic effect [9]. Considering a system of N particles, having mass M and coordinates X= X (t), the Langevin equation will be:

$$M\ddot{X} = -\Delta U(X) - YM\ddot{X} + \sqrt{2Yk_BTMR(t)}$$

, where the notations have their standard meaning.

5. Time Series Calculation

To study the dynamicity of a biomolecular system based on a temporal scale, getting proper knowledge of time-dependent statistical mechanics is an important step because firstly, in recent years there have been notable advances in the understanding of molecular dynamics algorithms and secondly, a comprehension of equilibrium time correlation functions in relation to their dynamical properties particularly their connection with transport coefficients, is a crucial step [16].

Time Series Calculation includes the root-mean-square deviation (RMSD), root means square fluctuation (RMSF), surface-accessibility (SA), the radius of gyration (RGYR), etc [9]. These calculations help us develop an idea about the biomolecular changes that occur gradually over time. Root mean square deviation gives us an idea whether the system is stable or unstable over the process of simulation. Root mean square fluctuations give us an idea about the flexibility of the residue understudy on a determined time scale. The radius of gyration is in simplest terms described as the measure of the root mean square distance of the system from its center of mass and it is also used to generate an idea of fluctuations of the system. Surface accessibility

gives us an idea about which portion of the biomolecular system being studied can be accessible to the solvent-based on its three-dimensional conformation.

6. MD Simulation Algorithms

MD Simulation to view the dynamic evolution of biological systems on a temporal scale can be determined by employing Newton's Laws of motion.

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

These classical equations of motion are integrated using the finite difference method. Finite difference methods are nothing but techniques used to generate MD trajectories with continuous potential models. The basic idea behind this is that the complete integration is broken down into smaller steps so the total force of each particle is calculated on a time scale as the vector sum of its interaction with other particles.

Algorithms are available for integrating the equation of motion using finite difference method and the main assumption made by all the available algorithms is that the dynamic property can be approximated by Taylor Series Expansion Method as [17-19]:

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

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

$$a(t+\delta t) = a(t) + b(t)\delta t + \cdots$$

where r = position,

v = velocity (the first derivative with respect to time),

a = acceleration (the second derivative with respect to time), etc.

i. Verlet Algorithm

It is the most basic and widely used algorithm used for integrating the equation of motion in molecular dynamics [20]. This is actually the two-third order Taylor Series expansion for the position of molecule r(t), one forward and another backward in time without using explicit velocities. The position of the previous step will be r(t-*dt*), and to calculate a new position we can write down the following equation:

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

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

Adding these two equations we get,

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

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

In the Verlet algorithm, the velocity got cancelled during the summation of the two equations. Though it is not required during the actual simulation process, but it is necessary to calculate the kinetic energy for testing the conservation of the total energy [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 positions, velocities, and accelerations are calculated as:

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

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

Because at a particular time all three parameters that are 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 the error that might occur is in the range of Δt^2.

iii.Leap Frog Algorithm

It was developed to control the error obtained due to the use of Velocity Verlet Algorithm. In this algorithm, positions and velocities are calculated at interleaved

time intervals. Here position r, is integer time step (t+δt), while velocity is integer plus half time step (t+1/2 δt) [22, 23]. Here, the velocities leap over the positions, then the positions *leap* over the velocities. The equation can be represented as:

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

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

The advantage of this algorithm is that the velocities are explicitly calculated.

The disadvantage is that they are not calculated at the same time as the positions because of which it is difficult to calculate total energy at any time point but an estimation can be made from the following equation:

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

iv. Beeman's algorithm

This algorithm is closely related to the Verlet algorithm and can be described by the equations:

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

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

Its advantage is that it provides a more accurate expression for the velocities and better energy conservation [24].

The disadvantage of this algorithm is that the calculations are computationally more expensive owing to the complex equations.

7. Software Used:

Following is a list of a few of the many software used for MD Simulation:

i.AMBER

AMBER is the abbreviation for Assisted Model Building with Energy Refinement and is a suite of molecular dynamics and energy minimization

program which can either refer to a set of molecular mechanical force fields or to a package of molecular simulation programs.

ii. CHARMM

CHARMM stands for Chemistry at HARvard Molecular Mechanics. It is also a molecular dynamic as well as an energy minimization program and can be used for molecular modelling as well.

iii. GROMOS

GROMOS or GROningen MOlecular Simulation is a general-purpose molecular dynamics computer simulation package for the simulation of biomolecular systems.

iv.NAMD

NAMD is the acronym for Nanoscale Molecular Dynamics and is an object-oriented MD simulation program.

3. Method

There are a number of software available for performing the molecular dynamic simulation of bio-molecules like GROMACS, Open Babel, VMD, UCSF Chimera, etc. We can select a 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.

i.Model Selection

ii.Energy Minimization. Heating and Equilibration

iii.Production Run and Analysis

If we talk about Chimera, Molecular Dynamics 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 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 pdb and psf file.

Obtaining Files- Simulations generally start with a crystal structure from the Protein Data Bank, in the standard PDB file format. The information about atoms of use to us are 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 observed structure with the force field used(T=0) by solving Newton's equations of motion. We also decide the number of equilibration steps (default 5000). Then the system is heated by rescaling the velocities and its stability ensured until the properties of the system no longer change with time and the system reaches a particular temperature.

Preparing the System for Energy Minimization- The energy of the system can be found using the force field calculations. The conformation of the system can be changed to locate lower energy conformations through the process called minimization. There are different minimization algorithms available, like

- steepest descent (slowly converging – used for highly restrained systems)
- conjugate gradient (efficient, uses intelligent choices of search direction – used for large systems)
- Broyden-Fletcher-Goldfarb-Shanno (BFGS) (quasi-newton variable metric method)
- Newton-Raphson (calculates both slopes of energy and rate of change)

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

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 selection, but one

must always remember that all atoms in the desired model will be included 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: In them, the atom types are assigned, for the identification of different elements and different molecular orbital environments. Also, charges are assigned to each atom, and connectivity between them is established.

Parameter files contain force constants requisite to describe the bond energy, nonbonded interactions (van der Waals and electrostatic), angle energy, torsion energy, etc. and also the suggested parameters for setting up the energy calculations.

Solvation is an important step because most biological processes occur in aqueous solutions and solvation effects play a decisive role in the determination of molecular conformation, electronic properties, binding energies, etc [25]. There are two methods of model solvation- explicit treatment and implicit treatment. In explicit treatment, solvent molecules are added to the molecular system whereas, in implicit treatment, the solvent is modelled 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 period of 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 development potential computational sampling methods now let us carry out the simulation process on a time scale of seconds to microsecond and even millisecond. Here it should be significantly noted that in simulation, these millisecond scales are believed to be very large that contradict the in vitro experiment owing to the fact that in computer simulation, coordinates are generated at the femtosecond level. Moving from femtoseconds to milliseconds

scale gives a large ensemble of conformations that have the potential to reveal many underlying biological mysteries.

4. Applications

MD simulations have a wide range 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, but with the usage of computational simulation techniques, one could make an even more precise approximation of the types of structural fluctuations the molecules undergo. By just examining a simulation of these structures, one can quantify the movements of various regions of the molecule at equilibrium and the types of structural fluctuations that occur [27]. Such simulations also have the capacity to show the dynamic behavioural 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 modelled structures-

This method can also be used to assess the accuracy of already modelled structures or even to refine the structures built using molecular modelling 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 fits the model to the experimental data even more precisely while simultaneously maintaining a physically stable structure [28]. One advantage of this approach is that this approach has been shown to control model errors that are

otherwise present. Let us consider another example. It is possible that a membrane protein suffers from artifacts due to the absence of a lipid bilayer or crystal structure suffers from such errors as a result of 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 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, quite a number of attempts to do this have been unsuccessful [29]. MD simulations have also been used 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 agree better with the

iii. Flexibility of Molecules-

The flexibility of biomolecules directly modulates its association with the neighbouring 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. Techniques such as Anisotropic Network Model (ANM), Elastic Network Model (ENM), Principal Component Analysis (PCA) among few others have recently been developed which have the ability to extrapolate the prime contributing motions of the system under study [31].

iv. Another interestingly important application 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 by 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 take place at the scale of femtoseconds, it is possible for us to observe such biological processes which 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. Modelling of Drug Receptor Interactions –

Experimental studies help us determine the 3D structures of most ligands (drug molecules) which interact with the receptors but the structures of the receptors are still unknown. The interactions among receptors and ligands are the focal step in most biological processes like pharmacological actions of drugs, regulatory mechanisms, toxic effects of particular chemicals, etc. It is significantly important to know the structure of both the receptor and ligands before carrying out further simulations by making modifications for drug design. In recent times, structure-based drug design (SBDD) and ligand-based drug design (LBDD) in computational forms have become important components of modern drug discovery. SBDD can be carried out with the available 3D structure of the target. Free-energy calculation methods in SBDD are used to calculate the absolute or relative binding free energy. The relative binding free energy methods, also known as alchemical calculations, use MD simulations to first sample and then compute 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 location for the molecule 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 are used to alter the structure as many times as possible so as to design a drug which fulfils 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 possible cellular interactions and behaviours 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 in such a way that they bind to each other to form a stable complex. MD simulation is widely used to study the protein-ligand, protein-protein, DNA-protein, and DNA-ligand interactions. Researchers in general are always excited to study the effect of new kinds of molecular interactions. 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 surface of contact, electrostatics, etc. [42], and a good score is considered as one which has a good binding affinity. It can be further studied by in-vitro, in-vivo, pre-clinical, and clinical trials. As each of these processes is time taking and expensive, the docked structures can help filter out structures thermodynamics and structural analyses over the trajectories obtained from molecular dynamics simulation and the streamlined results can be further studied in the labs or in vivo.

ix. Protein Folding-

Protein folding is one of the most exciting topics in biology. Though the three-dimensional structure of protein has been studied pretty well but the mechanism of protein-folding is still not accurately known. Molecular dynamics simulation is being used to study 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 thus a glimpse of hope towards the proper understanding of protein folding mechanism [43].

x. To understand how Mutation affects Interactions –

To understand and know the effects of a particular amino-acid in binding with its counterpart, the residue is mutated and the difference in the simulated trajectories both before and after the process is studied well. Refinements and modifications in the ligand can also be done so as to improve its affinity and get clear insights into the structural and interaction properties of the particular ligand [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 the field of 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 depends 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 obviously do not show cent percent accuracy in their results so better and more accurate techniques can always be developed. Also, 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 which 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 a lot of 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 the aid of effective simulation. It also provides knowledge about the atomic level interactions which directly affect the functions of the molecule. Uncovers the detailed information about the contributions of the different energy components which help in the binding of molecules and also maintaining its 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.

7. Chapter Digest

- Molecular Dynamic Simulation is an approach that uses computer techniques to apprehend the dynamicity of biological molecules by allowing the atoms and molecules to interact for a secure period of time and analyzing their physical movement and chemical interactions.
- The aim of Molecular Dynamic Simulation is to predict the behavior and movement of atoms as a time-dependent function with the hope to understand the properties of molecules with respect to their structure and conduct under different conditions.
- Though Molecular Dynamic Simulation was first introduced by Alder and Wainwright in 1957-1959, the first simulation was performed in 1964 by Rahman et.al. using and liquid argon, and the first MD simulation of protein was done by McCammon, et al in 1977 using bovine pancreatic trypsin inhibitor (BPTI).
- Molecular Modelling is the premier step in MD Simulation which encompasses all the tasks from visualization, derivation, manipulation and representation of the structures of molecules both theoretically and computationally to be able to mimic and study the structure and behaviour of molecules.

- Force Field is an empirical set of energy functions that is the summation of bonded and non-bonded terms or covalent and non-covalent interactions among the atoms and molecules. It can be represented as:

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

- periodic boundary conditions are a set of boundary conditions with additional constraints that enable a simulation to be performed using a relatively small number of particles in such a way that particles experience forces as if they are in bulk fluid.
- Time Series Calculation includes the root-mean-square deviation (RMSD), root means square fluctuation (RMSF), surface-accessibility (SA), the radius of gyration (RGYR), etc [9] and helps us develop an idea about the biomolecular changes that occur gradually over time.
- Algorithms available for performing MD Simulations integrate the equation of motion using finite difference method and the main assumption made by all the available algorithms is that the dynamic property can be approximated by Taylor Series Expansion Method as.
- AMBER, CHARMM, NAMD and GROMOS are a few notable software used for MD Simulation.
- Model Selection, Energy Minimization, Heating, Equilibration, Production Run and Analysis are the key steps in MD Simulation.

References:

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

2. Ma, B., Kumar, S., Tsai, C.-J., and Nussinov, R. (1999) Folding funnels and binding mechanisms. Protein engineering 12 (9), 713-720

3. Ma, B., Shatsky, M., Wolfson, H. J., and Nussinov, R. (2002) Multiple diverse ligands binding at a single protein site: a matter of pre-existing populations. Protein science 11 (2), 184-197

4. Fischer, E. (1894) Einfluss der Configuration auf die Wirkung der Enzyme. Berichte der deutschen chemischen Gesellschaft 27 (3), 2985-2993

5. Teague, S. J. (2003) Implications of protein flexibility for drug discovery. Nature reviews Drug discovery 2 (7), 527-541

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

7. Feynman, R. P., Leighton, R. B., and 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, S. A., and Dror, R. O. (2018) Molecular dynamics simulation for all. Neuron 99 (6), 1129-1143

9. Karaa, W. B. A. (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, W. L., and Tirado-Rives, J. (1996) Monte Carlo vs molecular dynamics for conformational sampling. The Journal of Physical Chemistry 100 (34), 14508-14513

12. Ishida, H., and Agag, T. (2011) *Handbook of benzoxazine resins*, Elsevier

13. Mayo, S. L., Olafson, B. D., and Goddard, W. A. (1990) DREIDING: a generic force field for molecular simulations. Journal of Physical chemistry 94 (26), 8897-8909

14. Pearlman, D. A., and Kollman, P. A. (1991) The overlooked bond-stretching contribution in free energy perturbation calculations. The Journal of chemical physics 94 (6), 4532-4545

15. Tamar, S. (2002) Molecular Modeling and Simulation. Interdisciplinary Applied Mathematics. NY: Springer

16. Gullingsrud, J. R., Braun, R., and Schulten, K. (1999) Reconstructing potentials of mean force through time series analysis of steered molecular dynamics simulations. Journal of Computational Physics 151 (1), 190-211

17. Akimov, A. V., and Kolomeisky, A. B. (2011) Recursive taylor series expansion method for rigid-body molecular dynamics. Journal of chemical theory and computation 7 (10), 3062-3071

18. Kutteh, R. (1999) New methods for incorporating nonholonomic constraints into molecular dynamics simulations. The Journal of chemical physics 111 (4), 1394-1406

19. Streett, W., Tildesley, D., and Saville, G. (1978) Multiple time-step methods in molecular dynamics. Molecular Physics 35 (3), 639-648

20. Grubmüller, H., Heller, H., Windemuth, A., and Schulten, K. (1991) Generalized Verlet algorithm for efficient molecular dynamics simulations with long-range interactions. Molecular Simulation 6 (1-3), 121-142

21. Martys, N. S., and Mountain, R. D. (1999) Velocity Verlet algorithm for dissipative-particle-dynamics-based models of suspensions. Physical Review E 59 (3), 3733

22. Toxvaerd, S. (1991) Algorithms for canonical molecular dynamics simulations. Molecular Physics 72 (1), 159-168

23. Van Gunsteren, W. F., and Berendsen, H. J. (1988) A leap-frog algorithm for stochastic dynamics. Molecular Simulation 1 (3), 173-185

24. Van Gunsteren, W., and Berendsen, H. (1982) Algorithms for Brownian dynamics. Molecular Physics 45 (3), 637-647

25. Rosenthal, S. J., Jimenez, R., Fleming, G. R., Kumar, P., and Maroncelli, M. (1994) Solvation dynamics in methanol: Experimental and molecular dynamics simulation studies. Journal of Molecular Liquids 60 (1-3), 25-56

26. Jabbarzadeh Kaboli, P., Ismail, P., and 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, K. O., and Sayre, D. (2001) An approach to three-dimensional structures of biomolecules by using single-molecule diffraction images. Proceedings of the National Academy of Sciences 98 (12), 6641-6645

28. Terwilliger, T. C., Read, R. J., Adams, P. D., Brunger, A. T., Afonine, P. V., Grosse-Kunstleve, R. W., and Hung, L.-W. (2012) Improved crystallographic models through iterated local density-guided model deformation and reciprocal-space refinement. Acta Crystallographica Section D: Biological Crystallography 68 (7), 861-870

29. Mirjalili, V., and Feig, M. (2013) Protein structure refinement through structure selection and averaging from molecular dynamics ensembles. Journal of chemical theory and computation 9 (2), 1294-1303

30. Lindorff-Larsen, K., Best, R. B., DePristo, M. A., Dobson, C. M., and Vendruscolo, M. (2005) Simultaneous determination of protein structure and dynamics. Nature 433 (7022), 128-132

31. Bakan, A., and Bahar, I. (2009) The intrinsic dynamics of enzymes plays a dominant role in determining the structural changes induced upon inhibitor binding. Proceedings of the National Academy of Sciences 106 (34), 14349-14354

32. Fribourg, M., Moreno, J. L., Holloway, T., Provasi, D., Baki, L., Mahajan, R., Park, G., Adney, S. K., Hatcher, C., and Eltit, J. M. (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, J. D., Harpsøe, K., Lansu, K., Yuan, S., Popov, P., Qu, L., Pu, M., Che, T., and Nikolajsen, L. F. (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, R. O., and Shaw, D. E. (2011) How fast-folding proteins fold. Science 334 (6055), 517-520

35. Snow, C. D., Nguyen, H., Pande, V. S., and Gruebele, M. (2002) Absolute comparison of simulated and experimental protein-folding dynamics. Nature 420 (6911), 102-106

36. Dedmon, M. M., Lindorff-Larsen, K., Christodoulou, J., Vendruscolo, M., and Dobson, C. M. (2005) Mapping long-range interactions in α-synuclein using spin-label NMR and ensemble molecular dynamics simulations. Journal of the American Chemical Society 127 (2), 476-477

37. Nguyen, H. D., and Hall, C. K. (2004) Molecular dynamics simulations of spontaneous fibril formation by random-coil peptides. Proceedings of the National Academy of Sciences 101 (46), 16180-16185

38. Bernardi, R. C., Melo, M. C., and 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, T. J., and 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, N. R., and Dror, R. O. (2016) Revealing atomic-level mechanisms of protein allostery with molecular dynamics simulations. PLoS computational biology 12 (6), e1004746

41. Schwantes, C. R., McGibbon, R. T., and Pande, V. S. (2014) Perspective: Markov models for long-timescale biomolecular dynamics. The Journal of chemical physics 141 (9), 09B201_201

42. Kroemer, R. T. (2007) Structure-based drug design: docking and scoring. Current Protein and Peptide Science 8 (4), 312-328

43. Duan, Y., and Kollman, P. A. (1998) Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution. Science 282 (5389), 740-744