Chapter 6
Molecular Dynamics Methods in Simulations of Macromolecules

Molecular dynamics simulations at atomic level have widely been used in studying macromolecular systems, such as protein, DNA and their complexes, mainly because the laws of classical statistical mechanics can largely govern the processes involved at the experimental conditions. Macromolecules, such as proteins, are characterized by dynamics with time scales ranging from nanoseconds to milliseconds. In this chapter, we discuss the molecular dynamics method as one of the most common computer simulation approach used to study molecular systems. In particular, we will present the equations of motion in the most relevant statistical ensembles used in the molecular dynamics simulations of molecular systems.

6.1 Introduction

Molecular dynamics (MD) method is playing a major role in studying macromolecular systems (Karplus and McCammon 2002) in part because the laws of classical statistical mechanics can mainly govern the processes involved at the experimental conditions (van Gunsteren et al. 2006). MD is a computer simulation approach used to numerically integrate the equations of motion of atoms and molecules by approximations of known physics (Allen and Tildesley 1989; Hoover 1991; Frenkel and Smit 2001). Ciccotti and Vanden-Eijnden (2015) argue that MD is an engine that is used to sample both time-independent and time-dependent statistical mechanical properties of molecular systems. In general, molecular systems may consist of a large number of particles. Thus it is impossible to find the properties of such complex systems analytically (Abraham 1986). When the number of bodies is more than two no analytical solutions can be seen and result in chaotic motion (Posch et al. 1986). MD simulation circumvents this problem by using numerical methods (Allen and Tildesley 1989; Hoover 1991). The first large-scale

H. Kamberaj, *Molecular Dynamics Simulations in Statistical Physics: Theory and Applications*, Scientific Computation,
https://doi.org/10.1007/978-3-030-35702-3_6

atomistic molecular dynamics simulations include the work of Abraham and co-workers (Abraham et al. 1984).

MD represents an interface between the experiment and theory and can be understood as a virtual experiment (Allen and Tildesley 1989; Frenkel and Smit 2001). MD probes the relationship between molecular structure, movement, and function (Karplus and McCammon 2002; Karplus and Kuriyan 2005). Molecular dynamics is a multidisciplinary method, and its laws and theories originate from different fields, such as mathematics, physics, and chemistry. Also, it implements algorithms from computer science and information theory (van Gunsteren et al. 2006). It was initially conceived within theoretical physics (Alder and Wainwright 1959), but today applies to other fields too, such as the computer simulations of the materials science and biomolecular systems. In the beginning, before the computers were used to perform molecular dynamics simulations, simple, but that required a lot of hard work, physical models were used, such as macroscopic spheres to prove the concepts (Bernal 1964; Hoover and Ree 1968; Alder et al. 1968).

Molecular dynamics as a discipline includes molecular modeling and computer simulations based on statistical mechanics laws. The main justification of MD method, as a specialized discipline of molecular modeling and computer simulation based on statistical mechanics, is that statistical ensemble averages are equal to time averages of the system, known as the ergodic hypothesis (Allen and Tildesley 1989; Leach 2001):

$$\langle \mathcal{A} \rangle = \lim_{T \to \infty} \int_0^T \mathrm{d}t \, \mathcal{A}(t) \tag{6.1}$$

where $\langle \cdots \rangle$ is an ensemble average and $\mathcal{A}$ can be any physical property of the system.

MD has also been termed *statistical mechanics by numbers* and *Laplace's vision of Newtonian mechanics* of predicting the future by animating nature's forces (Bernal 1964; Schlick 1996) and allowing insight into molecular motion on an atomic scale. However, long MD simulations are mathematically limited, due to cumulative errors in the numerical integration of equations of motion. However, these numerical errors can be minimized using more sophisticated algorithms and parameter selections (Martyna et al. 1994; Minary et al. 2004), but not eliminated (Piana et al. 2014). Furthermore, current potential functions may not, in all cases, be sufficiently accurate to reproduce the dynamics of molecular systems, so the much more computationally demanding Ab-Initio Molecular Dynamics method must be used combined with recent advances in parallel supercomputing (Hardy et al. 2011; Stone et al. 2011, 2013; Scarpazza et al. 2013; Phillips et al. 2014). Nevertheless, molecular dynamics method is successfully used to provide a detailed time and space resolution into representative behavior in phase space when combined with appropriate algorithms for optimizing the conformation search (Andricioaei and Straub 1996b).

MD approach has been used for a very long time to study macromolecular systems. These studies include many applications, such as the internal atomic motion (Amadei et al. 1993; Karplus and McCammon 2002), protein folding (Rogal and Bolhuis 2008), transition path sampling (Bolhuis et al. 2002), free energy calculations (Seyler and Beckstein 2014) and protein structure prediction (Perez et al. 2016). In all these studies, an observed limitation of standard MD method is the time and size scales covered in the simulations. For instance, in studying slow conformation motions of macromolecular systems (Palmer 1982; Clarage et al. 1995). To date there exist MD simulations that are far longer than standard MD simulation timescales, for instance, in Arkhipov et al. (2008) there are reported up to ten-microsecond MD simulation of a fast folding protein that was made possible through an improved scaling and parallel performance of the MD engine, or even longer, up to milliseconds time scale by computer engineering (Friedrichs et al. 2009; Dror et al. 2011). Long MD simulations are also possible using multiple time step integrator algorithms (Tuckerman et al. 1992; Tuckerman and Martyna 2000; Minary et al. 2004). Different advanced methods have been introduced to enhance conformation sampling of MD simulations (see the review Kamberaj 2019).

6.2 Equations of Motion

In this section, we discuss the equations of motion used in molecular dynamics simulations of different statistical ensembles.

6.2.1 Microcanonical Ensemble

The simplest form of equations of motion used in molecular dynamics corresponds to the microcanonical ensemble (namely NVE ensemble), in which the system is isolated and does not allow changes in number of particles (N), volume (V) and energy (E) (Allen and Tildesley 1989; Frenkel and Smit 2001). It corresponds to an adiabatic process with no heat exchange. During generation of a molecular dynamics trajectory in microcanonical ensemble, the energy is exchanged between potential and kinetic energy, but the total energy is constant. For an N particles system with coordinates $\mathbf{r} = \{\mathbf{r}_1, \cdots, \mathbf{r}_N\}$ and momenta $\mathbf{p} = \{\mathbf{p}_1, \cdots, \mathbf{p}_N\}$, the first order differential equations for each atom i with mass m_i are in Newton's notation as follows (Goldstein 2002):

$$\begin{aligned} \dot{\mathbf{p}}_i &= -\nabla U(\mathbf{r}), \\ \dot{\mathbf{r}}_i &= \mathbf{p}_i/m_i, \quad i = 1, \cdots, N \end{aligned} \tag{6.2}$$

The potential energy function $U(\mathbf{r})$ of the system is a function of the particle coordinates $\mathbf{r}$. In Physics, it is simply called the potential and in Chemistry as the force field (Leach 2001). The first equation comes from Newton's second law, where the force $\mathbf{F}_i$ acting on each particle i in the system is the negative gradient of the potential.

In a MD simulation run, every time step the positions and momenta of particles may be determined numerically using a numerical integration algorithm, such as Verlet or Leap-frog (Allen and Tildesley 1989). The time evolution of $\mathbf{r}$ and $\mathbf{p}$ defines a trajectory in the phase space. Given the initial positions (e.g. from theoretical knowledge) and momenta (e.g. according to Maxwell-Boltzmann distribution), the positions and velocities of the particles advance in time using Eq. (6.2).

It is interesting to explain the meaning of temperature in molecular dynamics simulations. From the macroscopic point of view, the temperature involves a huge number of particles. From the statistical point of view, the temperature is a statistical quantity, and as such only if there is a large enough number of atoms, statistical temperature can be estimated from the instantaneous temperature $\mathcal{T}$, which is found by equating the kinetic energy, E_k, of the system to $gk_B\mathcal{T}/2$ where g denotes the number of degrees of freedom of the system (Allen and Tildesley 1989):

$$\begin{aligned} T &= \langle \mathcal{T} \rangle, \\ \mathcal{T} &= 2E_k/gk_B \end{aligned} \tag{6.3}$$

If the number of atoms used in MD simulations is small, then the instantaneous temperature may show high fluctuations and may not converge to the thermodynamic temperature, known as the *temperature-related phenomenon* (Allen and Tildesley 1989). Something similar happens in biophysical simulations. The temperature of the system in NVE increases naturally when macromolecules, such as proteins, undergo exothermic conformation changes and binding.

6.2.2 Canonical Ensemble

The canonical (NVT) ensemble corresponds to the ensemble with the number of the atoms (N), volume (V) and temperature (T) constant. Often, it is called constant temperature molecular dynamics. In NVT, the energy of the system exchanges with the external bath also called as the *thermostat*.

There exist different thermostat methods introduced to absorb or dissipate energy at the boundaries of an MD system with a surrounding thermostat using the laws of physics, approximating in this way the canonical ensemble. In the MD simulations of macromolecular systems, the most popular methods to control temperature are the Nosé-Hoover thermostat, the Berendsen thermostat, and Langevin dynamics.

In the following, we will introduce all the methods that can be used to control the temperature in a molecular dynamics simulation.

6.2.2.1 Isokinetic Dynamics Method

The isokinetic dynamics method is the earliest technique used in the molecular dynamics simulations to re-scale the velocities sufficiently frequently for controlling the temperature of the system. In this approach, after each time step, Δt, of solving the microcanonical equations of motion (as given by Eqs. (6.2)), the particles momenta $\mathbf{p}(t)$ are updated to $\mathbf{p}(t + \Delta t)$, then each of the particle momentum is multiplied by a re-scaling factor giving in this way a new momentum $\mathbf{p}'(t + \Delta t)$ with the same kinetic energy as the previous time step (t):

$$\mathbf{p}_i'(t + \Delta t) = \sqrt{\frac{T_0}{\mathcal{T}}}\mathbf{p}_i(t + \Delta t), \quad i = 1, \cdots, N \tag{6.4}$$

where T_0 is the target temperature of the system and $\mathcal{T}$ is the instantaneous temperature calculated according to Eq. (6.3). The isokinetic dynamics method represents the simplest deterministic and time-reversible thermostat for sufficiently small integration time step Δt used in the molecular dynamics method.

6.2.2.2 Berendsen Thermostat

Another approach for controlling the temperature is the *Berendsen thermostat*, which represents a weak coupling of the system to an external heat bath at a fixed temperature T_0 (Berendsen et al. 1984). To control the temperature of system, each particle velocity is scaled at every timestep such that the rate change of temperature is proportional to the difference in temperature given as:

$$\frac{dT(t)}{dt} = \frac{1}{\tau}(T_0 - T(t)) \tag{6.5}$$

Here, τ represents a coupling parameter, which determines how strongly the bath and the system are coupled to each other. This method yields an exponential decay of the system's temperature towards the desired temperature with a change in the temperature at every time step given as:

$$\Delta T = \frac{\Delta t}{\tau}(T_0 - T(t)) \tag{6.6}$$

The scaling factor of the velocities is (Berendsen et al. 1984)

$$\lambda = \left[1 + \frac{\Delta t}{\tau}\left(\frac{T_0}{T(t - \Delta t/2)} - 1\right)\right]^{1/2} \tag{6.7}$$