

# Molecular dynamic simulation

Mamta Sagar

UIET, CSJMU Kanpur

# Introduction

- One of the principal tools in the theoretical study of biological molecules is the method of molecular dynamics simulations (MD).
- This computational method calculates the time dependent behavior of a molecular system.
- MD simulations have provided detailed information on the fluctuations and conformational changes of proteins and nucleic acids.

# Different type of motions

- Biological molecules exhibit a wide range of time scales over which specific processes occur; for example
- Local Motions (0.01 to 5 Å,  $10^{-15}$  to  $10^{-1}$  s)  
Atomic fluctuations ,Sidechain Motions ,Loop Motions
- Rigid Body Motions (1 to 10Å,  $10^{-9}$  to 1s)  
Helix Motions ,Domain Motions (hinge bending)  
,Subunit motions
- Large-Scale Motions ( $> 5\text{Å}$ ,  $10^{-7}$  to  $10^4$  s)  
Helix coil transitions ,Dissociation/Association ,Folding and Unfolding

Molecular dynamics simulations permit the study of complex, dynamic processes that occur in biological systems. These include, for example

- Protein stability
- Conformational changes
- Protein folding
- Molecular recognition: proteins, DNA, membranes, complexes
- Ion transport in biological systems

and provide the mean to carry out the following studies,

- Drug Design
- Structure determination: X-ray and NMR

## 2. Historical Background

- The molecular dynamics method was first introduced by Alder and Wainwright in the late 1950's (Alder and Wainwright, 1957,1959) to study the interactions of hard spheres.
- the next major advance was in 1964, when Rahman carried out the first simulation using a realistic potential for liquid argon (Rahman, 1964).
- The first molecular dynamics simulation of a realistic system was done by Rahman and Stillinger in their simulation of liquid water in 1974 (Stillinger and Rahman, 1974). The first protein simulations appeared in 1977 with the simulation of the bovine pancreatic trypsin inhibitor (BPTI) (McCammon, et al, 1977).

# 3. Statistical Mechanics

- Molecular dynamics simulations generate information at the microscopic level, including atomic positions and velocities.
- The conversion of this microscopic information to macroscopic observables such as pressure, energy, heat capacities, etc., requires statistical mechanics. Statistical mechanics is fundamental to the study of biological systems by molecular dynamics simulation.
- In this section, we provide a brief overview of some main topics. For more detailed information, refer to the numerous excellent books available on the subject.

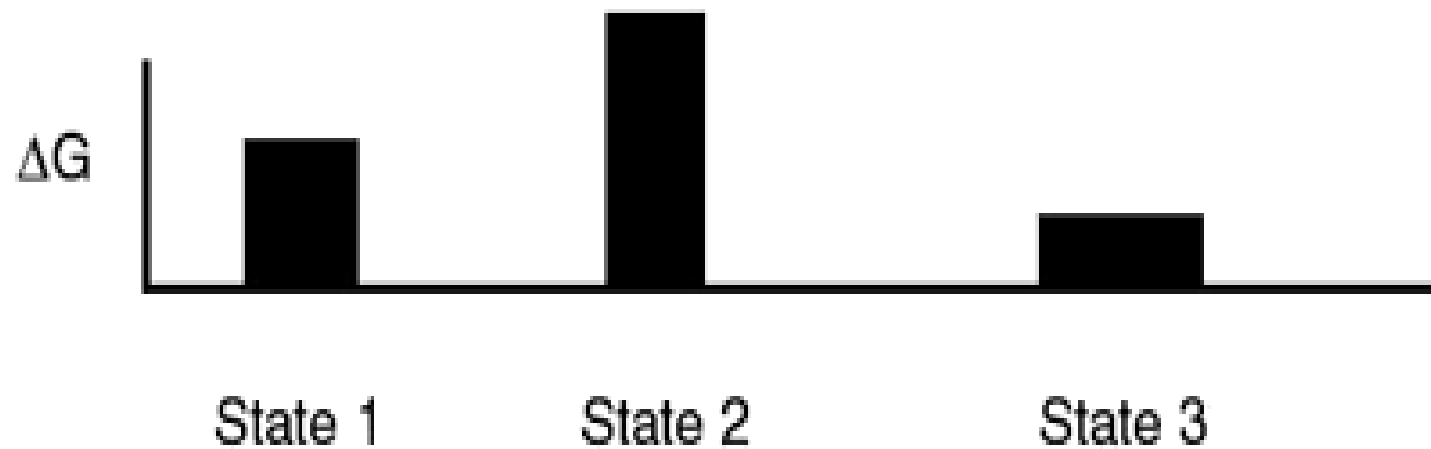
# Introduction to Statistical Mechanics:

- In a molecular dynamics simulation, one often wishes to explore the macroscopic properties of a system through microscopic simulations, for example, to calculate changes in the binding free energy of a particular drug candidate, or to examine the energetics and mechanisms of conformational change.
- The connection between microscopic simulations and macroscopic properties is made via statistical mechanics which provides the rigorous mathematical expressions that relate macroscopic properties to the distribution and motion of the atoms and molecules of the N-body system;

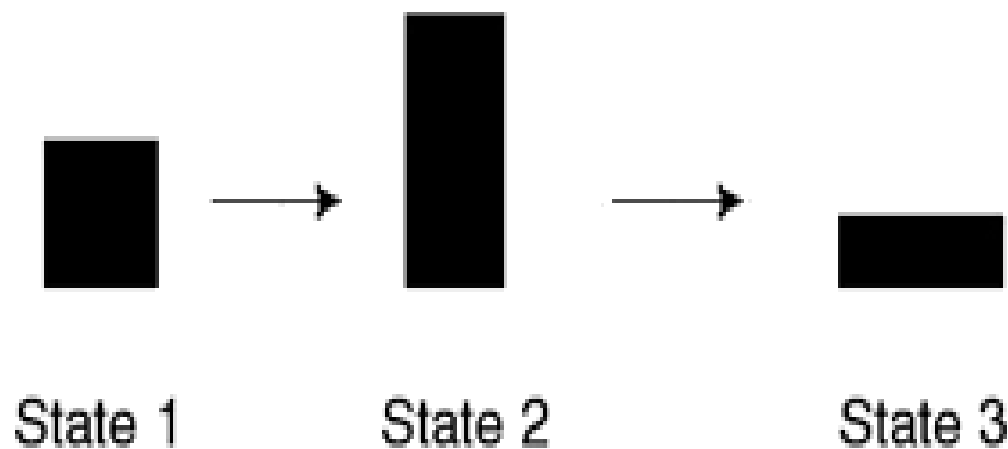
- molecular dynamics simulations provide the means to solve the equation of motion of the particles and evaluate these mathematical formulas.
- With molecular dynamics simulations, one can study both thermodynamic properties and/or time dependent (kinetic) phenomenon.



Thermodynamics describes the driving force for chemical processes



Kinetics describes the mechanism for the chemical process



# thermodynamic state & Phase space

- Definitions

The thermodynamic state of a system is usually defined by a small set of parameters, for example, the temperature,  $T$ , the pressure,  $P$ , and the number of particles,  $N$ . Other thermodynamic properties may be derived from the equations of state and other fundamental thermodynamic equations.

- The mechanical or microscopic state of a system is defined by the atomic positions,  $q$ , and momenta,  $p$ ; these can also be considered as coordinates in a multidimensional space called phase space.

- For a system of  $N$  particles, this space has  $6N$  dimensions. A single point in phase space, denoted by  $G$ , describes the state of the system. **An ensemble is a collection of points in phase space satisfying the conditions of a particular thermodynamic state.**
- A molecular dynamics simulation generates a sequence of points in phase space as a function of time; these points belong to the same ensemble, and they correspond to the different conformations of the system and their respective momenta. Several different ensembles are described below.

- An ensemble is a collection of all possible systems which have different microscopic states but have an identical macroscopic or thermodynamic state.

- There exist different ensembles with different characteristics.
- Microcanonical ensemble (NVE) : The thermodynamic state characterized by a fixed number of atoms,  $N$ , a fixed volume,  $V$ , and a fixed energy,  $E$ . This corresponds to an isolated system.
- Canonical Ensemble (NVT): This is a collection of all systems whose thermodynamic state is characterized by a fixed number of atoms,  $N$ , a fixed volume,  $V$ , and a fixed temperature,  $T$ .
- Isobaric-Isothermal Ensemble (NPT): This ensemble is characterized by a fixed number of atoms,  $N$ , a fixed pressure,  $P$ , and a fixed temperature,  $T$ .
- Grand canonical Ensemble ( $\mu VT$ ): The thermodynamic state for this ensemble is characterized by a fixed chemical potential,  $\mu$ , a fixed volume,  $V$ , and a fixed temperature,  $T$ .

In statistical mechanics, average values are defined as ensemble averages.

The ensemble average is given by

$$\langle A \rangle_{ensemble} = \iint dp^N dr^N A(p^N, r^N) \rho(p^N, r^N)$$

$$A(p^N, r^N)$$

is the observable of interest and it is expressed as a function of the momenta,  $p$ , and the positions,  $r$ , of the system. The integration is over all possible variables of  $r$  and  $p$ .

The probability density of the ensemble is given by

$$\rho(p^N, r^N) = \frac{1}{Q} \exp[-H(p^N, r^N) / k_B T]$$

where H is the Hamiltonian, T is the temperature,  $k_B$  is Boltzmann's constant and Q is the partition function

$$Q = \iint dp^N dr^N \exp[-H(p^N, r^N) / k_B T]$$

- This integral is generally extremely difficult to calculate because one must calculate all possible states of the system.
- In a molecular dynamics simulation, the points in the ensemble are calculated sequentially in time, so to calculate an ensemble average, the molecular dynamics simulations must pass through all possible states corresponding to the particular thermodynamic constraints.



Another way, as done in an MD simulation, is to determine a time average of  $A$ , which is expressed as

$$\langle A \rangle_{time} = \lim_{\tau \rightarrow \infty} \frac{1}{\tau} \int_{t=0}^{\tau} A(p^N(t), r^N(t)) dt \approx \frac{1}{M} \sum_{t=1}^M A(p^N, r^N)$$

where  $t$  is the simulation time,  $M$  is the number of time steps in the simulation and  $A(p^N, r^N)$  is the instantaneous value of  $A$ .

The dilemma appears to be that one can calculate time averages by molecular dynamics simulation, but the experimental observables are assumed to be ensemble averages. Resolving this leads us to one of the most fundamental axioms of statistical mechanics, **the ergodic hypothesis, which states that the time average equals the ensemble average.** Ergodic hypothesis state

$$\langle A \rangle_{ensemble} = \langle A \rangle_{time}$$

- The basic idea is that if one allows the system to evolve in time indefinitely, that system will eventually pass through all possible states.
- One goal, therefore, of a molecular dynamics simulation is to generate enough representative conformations such that this equality is satisfied.
- If this is the case, experimentally relevant information concerning structural, dynamic and thermodynamic properties may then be calculated using a feasible amount of computer resources.
- Because the simulations are of fixed duration, one must be certain to sample a sufficient amount of phase space.

# Some examples of time averages:

Average potential energy

$$\bar{V} = \langle V \rangle = \frac{1}{M} \sum_{i=1}^M V_i$$

where  $M$  is the number of configurations in the molecular dynamics trajectory and  $V_i$  is the potential energy of each configuration.

# Average kinetic energy

$$K = \langle K \rangle = \frac{1}{M} \sum_{j=1}^M \left\{ \sum_{i=1}^N \frac{m_i}{2} v_i \cdot v_i \right\}_j$$

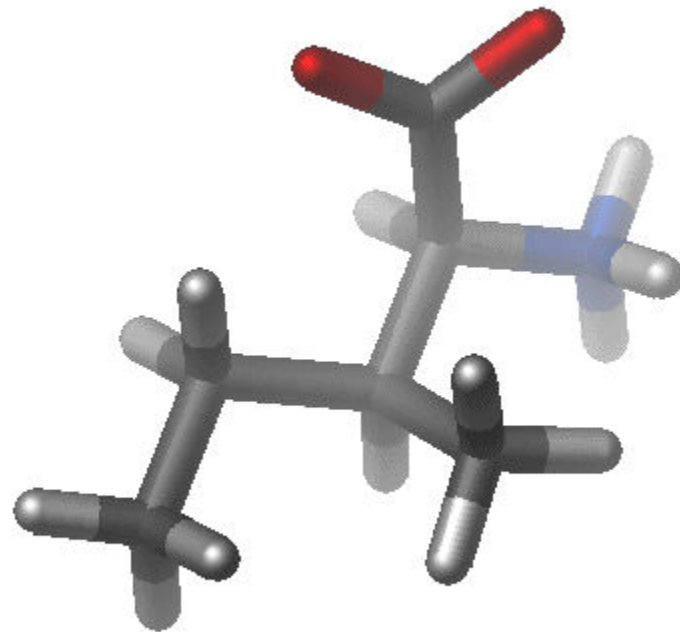
where  $M$  is the number of configurations in the simulation,  $N$  is the number of atoms in the system,  $m_i$  is the mass of the particle  $i$  and  $v_i$  is the velocity of particle  $i$ .

A molecular dynamics simulation must be sufficiently long so that enough representative conformations have been sampled.

# 4. Classical Mechanics

- The molecular dynamics simulation method is based on Newton's second law or the equation of motion,  $F=ma$ , where  $F$  is the force exerted on the particle,  $m$  is its mass and  $a$  is its acceleration.
- From a knowledge of the force on each atom, it is possible to determine the acceleration of each atom in the system.
- Integration of the equations of motion then yields a trajectory that describes the positions, velocities and accelerations of the particles as they vary with time. From this trajectory, the average values of properties can be determined

- Current generation force fields (or potential energy functions) provide a reasonably good compromise between accuracy and computational efficiency. They are often calibrated to experimental results and quantum mechanical calculations of small model compounds.
- Their ability to reproduce physical properties measurable by experiment is tested; these properties include structural data obtained from x-ray crystallography and NMR, dynamic data obtained from spectroscopy and inelastic neutron scattering and thermodynamic data.





- The development of parameter sets is a very laborious task, requiring extensive optimization. This is an area of continuing research, Among the most commonly used potential energy functions are the AMBER, CHARMM, GROMOS and OPLS/AMBER force fields.

## **Charm potential energy function**

- The energy,  $E$ , is a function of the atomic positions,  $R$ , of all the atoms in the system, these are usually expressed in term of Cartesian coordinates.

- The value of the energy is calculated as a sum of internal, or bonded, terms  $E_{\text{bonded}}$ , which describe the bonds, angles and bond rotations in a molecule, and a sum of external or nonbonded terms,  $E_{\text{non-bonded}}$ . These terms account for interactions between nonbonded atoms or atoms separated by 3 or more covalent bonds.

$$V(R) = E_{\text{bonded}} + E_{\text{non-bonded}}$$

The  $E_{\text{bonded}}$  term is a sum of three terms:

$$E_{\text{bonded}} = E_{\text{bond-stretch}} + E_{\text{angle-bend}} + E_{\text{rotate-along-bond}}$$

which correspond to three types of atom movement:

- The first term in the above equation is a harmonic potential representing the interaction between atomic pairs where atoms are separated by one covalent bond, i.e., 1,2-pairs. This is the approximation to the energy of a bond as a function of displacement from the ideal bond length,  $b_0$ . The force constant,  $K_b$ , determines the strength of the bond.

$$E_{\text{bond-stretch}} = \sum_{1,2 \text{ pairs}} K_b (b - b_0)^2$$

- The second term in above equation is associated with alteration of bond angles  $\theta$  from ideal values  $\theta_0$ , which is also represented by a harmonic potential. Values of  $\theta_0$  and  $K_\theta$  depend on chemical type of atoms constituting the angle.

$$E_{\text{bond-bend}} = \sum_{\text{angles}} K_\theta (\theta - \theta_0)^2$$

- These two terms describe the deviation from an ideal geometry; effectively, they are penalty functions and that in a perfectly optimized structure, the sum of them should be close to zero.

- The third term represents the torsion angle potential function which models the presence of steric barriers between atoms separated by 3 covalent bonds (1,4 pairs). The motion associated with this term is a rotation, described by a dihedral angle and coefficient of symmetry  $n=1,2,3$ , around the middle bond. This potential is assumed to be periodic and is often expressed as a cosine function.

$$E_{\text{rotate-along-bond}} = \sum_{1,4 \text{ pairs}} K_{\phi} (1 - \cos(n\phi))$$

- In addition to these terms, the CHARMM force field has two additional terms; one is the Urey-Bradley term, which is an interaction based on the distance between atoms separated by two bonds (1,3 interaction). The second additional term is the improper dihedral term (see the section on CHARMM) which is used to maintain chirality and planarity

- The parameters for the these terms,  $K_b$ ,  $K_q$ ,  $K_f$ , are obtained from studies of small model compounds and comparisons to the geometry and vibrational spectra in the gas phase (IR and Raman spectroscopy), supplemented with ab initio quantum calculations.

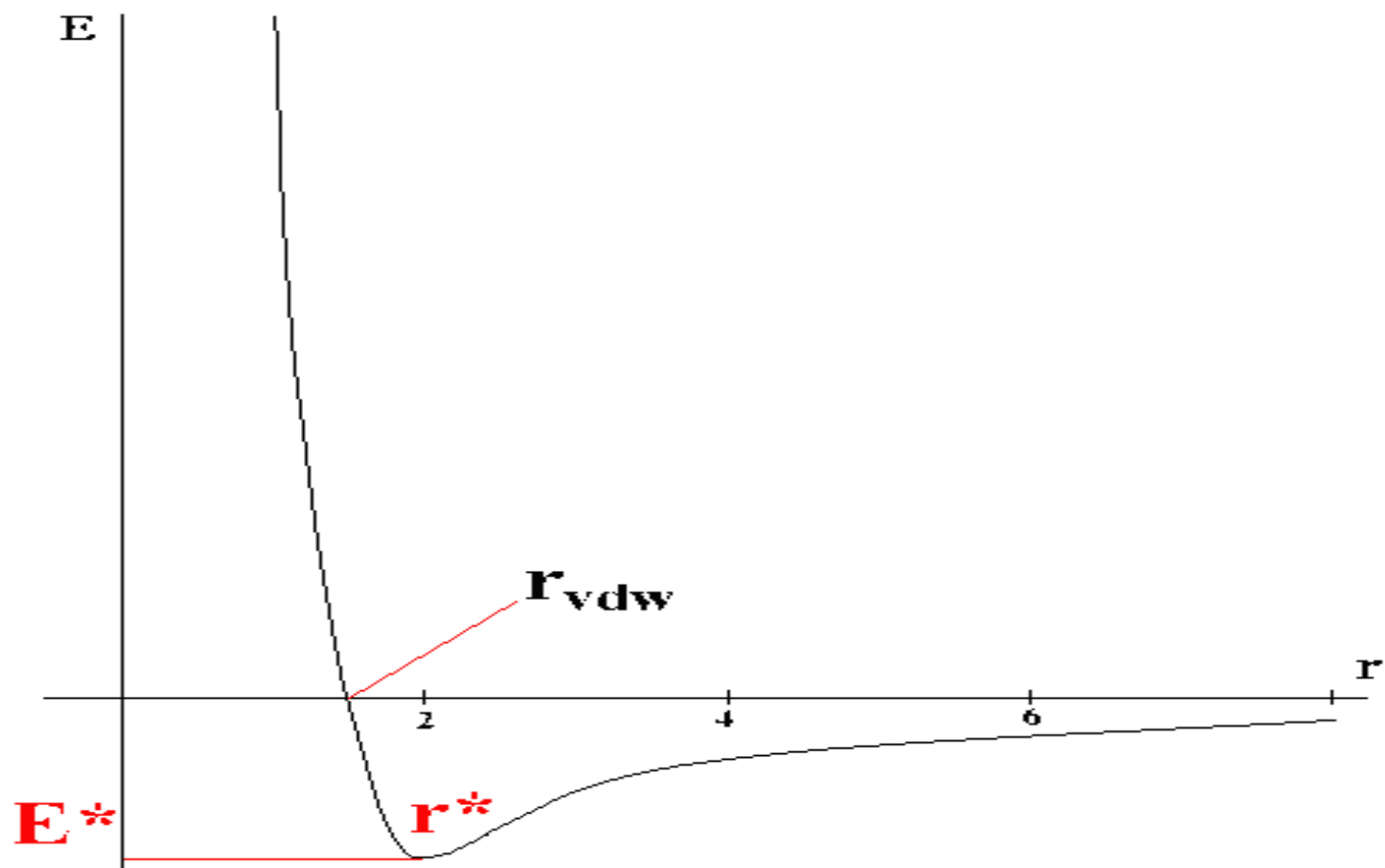


- The energy term representing the contribution of non-bonded interactions in the CHARMM potential function has two components, the Van der Waals interaction energy and the electrostatic interaction energy. Some other potential functions also include an additional term to account for hydrogen bonds. In the CHARMM potential energy function, these interactions are account for by the electrostatic and Van der Waals interactions.

$$E_{non-bonded} = E_{van-der-Waals} + E_{electrostatic}$$

- The van der Waals interaction between two atoms arises from a balance between repulsive and attractive forces. The repulsive force arises at short distances where the electron-electron interaction is strong. The attractive force, also referred to as the dispersion force, arises from fluctuations in the charge distribution in the electron clouds.
- The fluctuation in the electron distribution on one atom or molecules gives rise to an instantaneous dipole which, in turn, induces a dipole in a second atom or molecule giving rise to an attractive interaction. Each of these two effects is equal to zero at infinite atomic separation  $r$  and become significant as the distance decreases.

- The attractive interaction is longer range than the repulsion but as the distance become short, the repulsive interaction becomes dominant. This gives rise to a minimum in the energy. Positioning of the atoms at the optimal distances stabilizes the system. Both value of energy at the minimum  $E^*$  and the optimal separation of atoms  $r^*$  (which is roughly equal to the sum of Van der Waals radii of the atoms) depend on chemical type of these atoms.



- The van der Waals interaction is most often modelled using the Lennard-Jones 6-12 potential which expresses the interaction energy using the atom-type dependent constants A and C. Values of A and C may be determined by a variety of methods, like non-bonding distances in crystals and gas-phase scattering measurements

$$E_{\text{van-der-Waals}} = \sum_{\substack{\text{nonbonded} \\ \text{pairs}}} \left( \frac{A_{ik}}{r_{ik}^{12}} - \frac{C_{ik}}{r_{ik}^6} \right)$$

- The electrostatic interaction between a pair of atoms is represented by Coulomb potential;  $D$  is the effective dielectric function for the medium and  $r$  is the distance between two atoms having charges  $q_i$  and  $q_k$ . The empirical potential energy function is differentiable with respect to the atomic coordinates; this gives the value and the direction of the force acting on an atom and thus it can be used in a molecular dynamics simulation.

$$E_{\text{electrostatic}} = \sum_{\substack{\text{nonbonded} \\ \text{pairs}}} \frac{q_i q_k}{D r_{ik}}$$

- The empirical potential function has several limitations, which result in inaccuracies in the calculated potential energy.

- One limitation is due to the fixed set of atom types employed when determining the parameters for the force field. Atom types are used to define an atom in a particular bonding situation,
- An approximation introduced to decrease the computational demand is the pair-wise additive approximation, i.e., interaction energy between one atom and the rest of the system is calculated as a sum of pair-wise (on atom to one atom) interactions, or as if the pair of atoms do not see the other atoms in the system

- Another important point to take into consideration is that the potential energy function does not include entropic effects.
- Thus, a minimum value of  $E$  calculated as a sum of potential functions does not necessarily correspond to the equilibrium, or the most probable state; this corresponds to the minimum of free energy.



- Because of the fact that experiments are generally carried out under isothermal-isobaric conditions (constant pressure, constant system size and constant temperature) the equilibrium state corresponds to the minimum of Gibb's Free Energy,  $G$ .
- While just an energy calculation ignores entropic effects, these are included in a molecular dynamics simulations