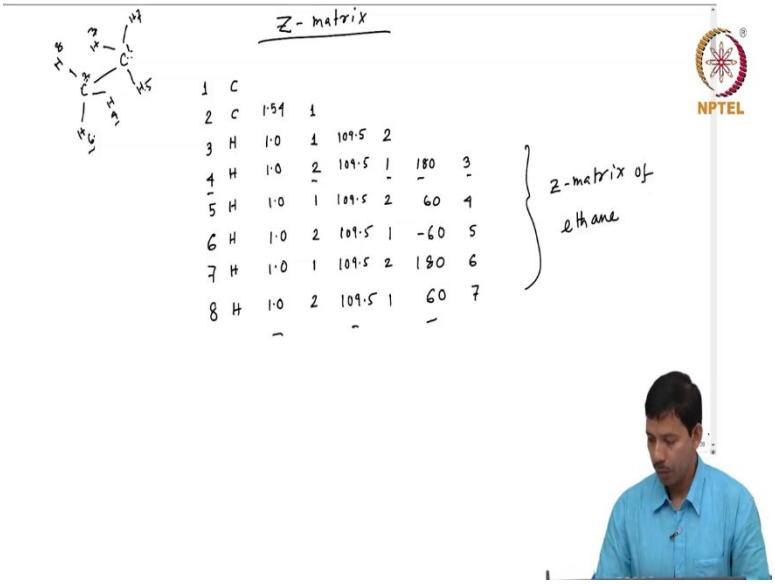
Thermodynamics for Biological Systems:
Classical and Statistical Aspects
Prof. Sanjib Senapati
Department of Biotechnology
Indian institute of Technology - Madras

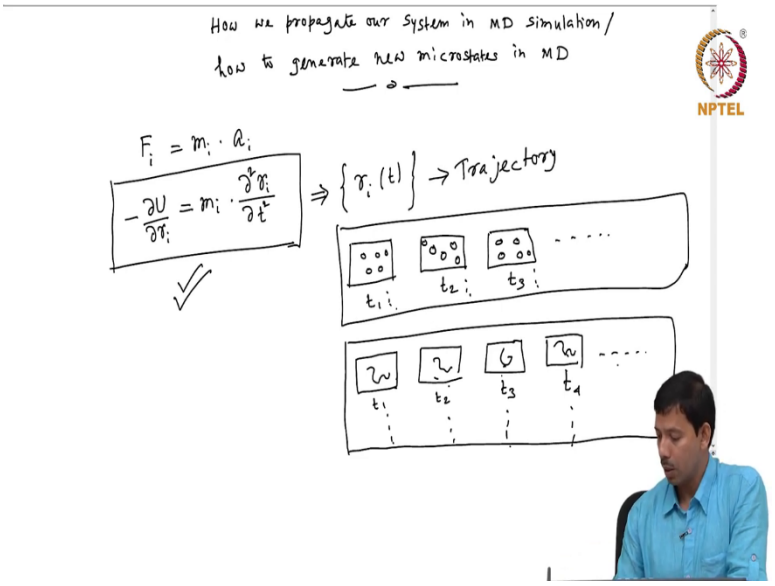
Lecture – 81
Basics of MD Simulations

(Refer Slide Time: 00:18)



So, starting from that initial structure how I generate a new structure or a new microstate or the new conformation the protein.

(Refer Slide Time: 00:31)



And so, now the question is, how we propagate our system in MD simulation, MD is molecular dynamics simulation, or, in other words, how to generate new microstates in MD? This is what

we will be discussing. It is very simple. The way we generate new microstate in molecular dynamics simulation is by solving Newton's equation of motion which is Newton's second law. So Newton's equation of motion or the second law as you know is

$$F_i = m_i \, . \, a_i$$

So, here i denotes the ith particle, a_i is the acceleration and in acceleration that time factor is involved. So, a_i I can write as

$$a_i = \frac{\partial^2 r_i}{\partial t^2}$$

so here comes the time.

And how can I write force? Force we can write as a gradient negative gradient of potential

$$-\frac{\partial U}{\partial r_i} = m_i \cdot \frac{\partial^2 r_i}{\partial t^2}$$

this is second-order differential equation. If we can solve this second-order differential equation this will give us a set of r_i as a function of time.

So, a set of r as a function of time is nothing but a trajectory of different structures of your molecule. A simple example, at time t_1 let us say my five particle distribution was like this, if I solve this, I get at time t_2 a different distribution, at time t_3 I get another distribution and so on, so forth. Basically you keep on generating as many microstates as you can and that is basically your ensemble.

So, I can keep on making lot of microstates. If it was my biomolecular problem, so then, I started from this confirmation of the protein at time t_2, I generated another conformation of the iprotein at time t_3, I generated another confirmation of the protein at t_4, I got back the confirmation we started from and so on, so forth. I carry on this simulation as long my available facility allows me.

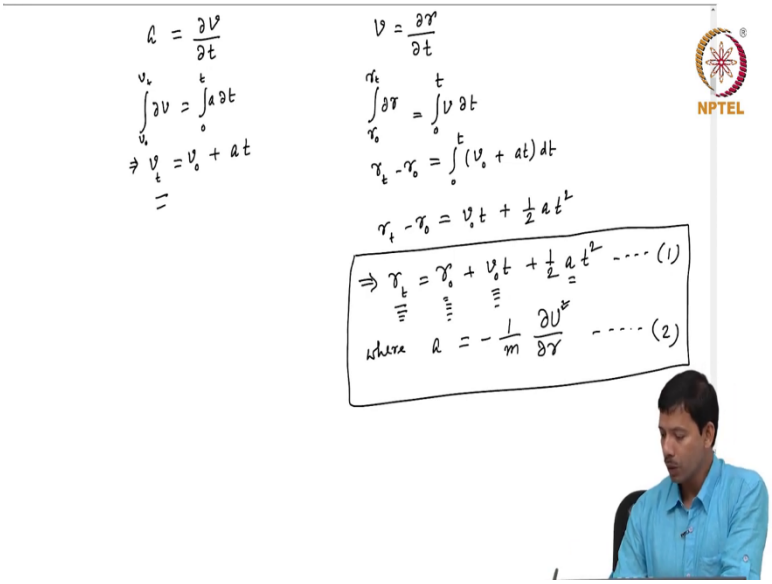
And then, I basically created the ensemble. So, this is called the Trajectory. So, Trajectory is basically the thread in which I have all these different microstates available and stored in my computer hard disk. So, if I take the average over all these structures, I get down ensemble average or I get the time average. And since our ensemble average or the time average had a

probability factor, so the particular microstate, which will appear more and more time, in the average the contribution of that microstate will be the largest.

And that is called the most probable state or the most probable distribution. So, that is the idea of getting the new microstate in my simulation. Now, in principle, we can solve this second-order differential equation and get the new microstates from MD. But in practical solving this second-order differential equation is difficult, is not easily tractable and that is why the easier process is to solve two coupled first order equations.

So, instead of solving the second order differential equation now we will be seeing how we can solve us a pair of first-order equations.

(Refer Slide Time: 07:48)



So, as you know, your acceleration is

$$a = \frac{\partial v}{\partial t}$$

v is the velocity. So I can write this as

$$\partial v = a \,.\, \partial t$$

if I integrate with v_0 to v_t and 0 to t, I will get

$$\int_{v_0}^{v_t} \partial v = \int_{0}^{t} a \, \partial t$$

$$v_t = v_0 + at$$

Here, I have assumed that a is not changing and that assumption is not bad because in simulation the delta t is basically the difference- t_2 minus t_1, so this time gap between two microstates is taken to be very small.

Otherwise there is a divergence in the algorithm, the algorithm diverges. And therefore this delta t in simulation is taken to be very small, as small as one femto second and in that small time delta t we assume that the acceleration does not change. So, this is one first-order differential equation and the other first-order differential equation we can take as

$$v = \frac{\partial r}{\partial t}$$

So if I again integrate now, this from r_0 to r_t where my r_0 is the initial coordinate and r_t is the coordinate at time t_0 to t and I can put back my v_t value, it gives

$$\int_{r_0}^{r_t} \partial r = \int_0^t v\, \partial t$$

$$r_t - r_0 = \int_0^t (v_0 + at)dt$$

$$r_t - r_0 = v_0 t + \frac{1}{2} at^2$$

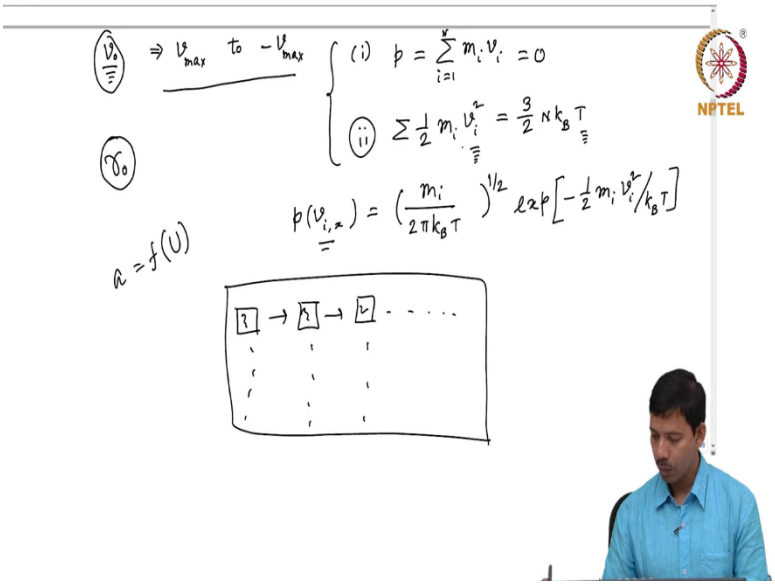
$$r_t = r_0 + v_0 t + \frac{1}{2} at^2 \qquad eqn\ (1)$$

Where a is

$$a = -\frac{1}{m}\frac{\partial U}{\partial r} \qquad eqn\ (2)$$

If I solve these two equations, what I get is basically the new set of coordinates r_t at time t from my starting confirmation of r_0. So, more number I integrate these two equations, I get more number of microstates. So, that is how we basically generate the microstates in MD simulations. Now, so r_0 is known to you which is the initial set of coordinates you get from PDB or homology modeling or chem draw, a is known to you because a is minus 1 by m del u del r and the potential you get from force field library so what you need to know to start with is initial set of velocities.

(Refer Slide Time: 12:55)



So, the initial set of velocities v_0 usually we start from Maxwell Boltzmann or Gaussian distribution. So, we choose the initial velocities of particles in such a way that certain conditions fulfil. So, we choose initial velocity from let us say from Vmax to -V max, in such a way that the total momentum of the system is zero.

$$(i) \quad p = \sum_{i=1}^{N} m_i v_i = 0$$

The total momentum of the system is zero. So, particles are moving but the total momentum of all particles in the simulation box is zero. The other condition we take is that

$$(ii) \quad \sum \frac{1}{2} m_i v_i^2 = \frac{3}{2} N k_B T$$

So, this equality condition has to be fulfilled because the temperature of the system is fixed, n is fixed. The total kinetic energy has to be same that has to be fixed and to maintain that we assign the initial velocities in such a way that my total kinetic energy 3 by 2 n KBT is fulfilled. And we do that with a probability of particle I, velocity in the x component is

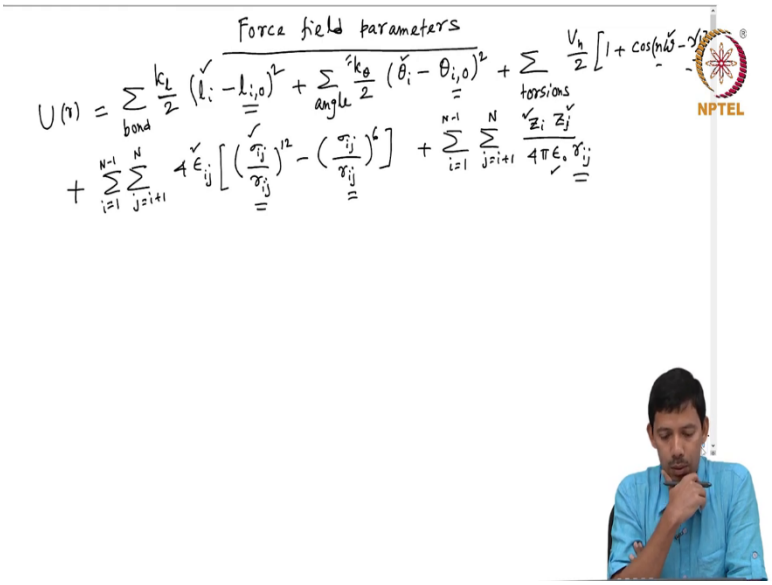
$$p(v_{i,x}) = (\frac{m_i}{2\pi k_B T})^{\frac{1}{2}} \exp[\frac{-\frac{1}{2} m_i v_i^2}{k_B T}]$$

So, this is how we have assigned our velocity till the ith particle.

So, this is the x component velocity of ith particle with a probability which basically follows the Maxwell-Boltzmann or the Gaussian distribution and my velocity would be from $+V_{Max}$ to $-V_{Max}$ with these two conditions to be fulfilled. huh so with those initial velocity of V_0, from the initial set of coordinates r_0 and from U, I start my MD simulation and from the MD simulation I generate one confirmation to another confirmation to another and so on, so forth.

And I have an ensemble of microstates and I can take time average over them to get the average property, which I can compare it to experimental data. On the top of that the advantage here is I have all these microstates stored in my system and therefore I can always go back and trace the transition of my system at different time points. So, not only I will be able to reproduce the average properties by MD but also I will be able to provide useful insights in terms of the transition or in terms of the microstructure, which will be difficult to obtain from experimental data.

(Refer Slide Time: 18:52)



If you now look at the expression of U, so expression of U does not look that simple.

$$U(r) = \sum_{bond} \frac{k_l}{2}(l_i - l_{i,0})^2 + \sum_{angle} \frac{k_\theta}{2}(\theta_i - \theta_{i,0})^2 + \sum_{torsions} \frac{V_n}{2}[1 + \cos(n\omega - \gamma)] + \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6} \right] - \sum_{i=1}^{N-1} \sum_{i=1}^{N-1} \frac{Z_i Z_j}{4\pi\varepsilon_o r_{ij}}$$

Expression of U and then when your bio molecular simulation boxes having so many atoms, so many protein residues and you need lot of water molecules around the protein to solvate the protein very well into the system, your calculation is really going to be difficult.

So, even though we have used two coupled first order equations and got equation one to get the microstates, but because of this complex nature of U, it is very difficult to get an analytical solution of r_t. Since it is difficult to get an analytical solution of r_t we often use numerical methods to obtain new values new rt values or the new microstates. So