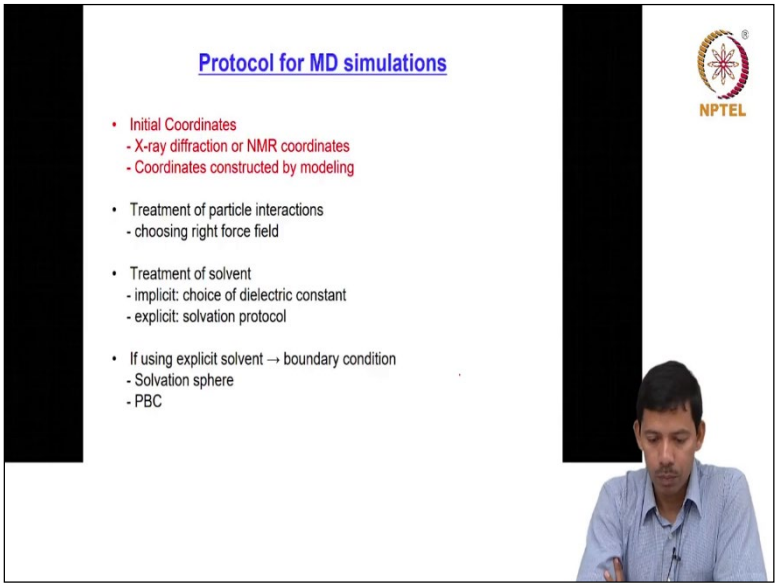
Thermodynamics for Biological Systems:
Classical and Statistical Aspects
Prof. Sanjib Senapati
Department of Biotechnology
Indian institute of Technology - Madras

Lecture – 77
MD Protocol

(Refer Slide Time: 00:17)



So, huh what are the important steps in MD simulation? So, to start your simulation the first thing you have to know it is a starting microstate, you have to start from some point what is that point. So, you can choose our random microstate, so when we mean the microstate of our classical system we basically look at the distribution of particles or the atoms or basically you have to know XYZ coordinates of each of the atom that make the system.

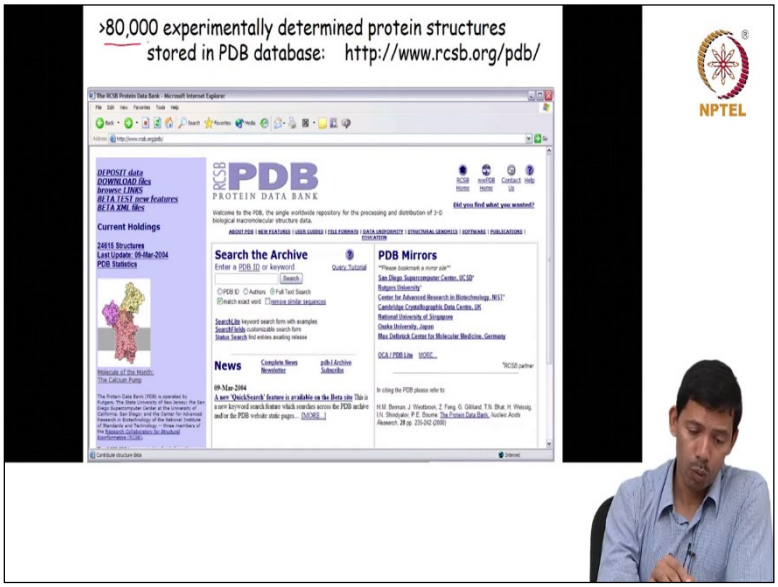
So if it is a biological system let us say your protein or DNA you if you are lucky then you might have a structure of your biomolecule available in something called protein data bank. So, as you know protein data bank has lot of x-ray or NMR structure of various proteins, so you can go to protein data bank and search for such for it is pdb search for its coordinates. So, if you are lucky so you get you get the extra structure or the NMR structures and from there you get the coordinate.

So that is your starting structure of the protein so you can start from there and from there you start your computer simulation and that will generate you different microstates over time. If your system of interest does not have a initial coordinate does not have x-ray structure or NMR

structure then you can or you can as I said that you can make the coordinates rough coordinate by modelling.

So, you just have to make a approximate structure of the system and then you are ready to go, you are said to go from there in carry out computer simulation and simulation will basically iterate over to get a better and better confirmation of your molecule of interest.

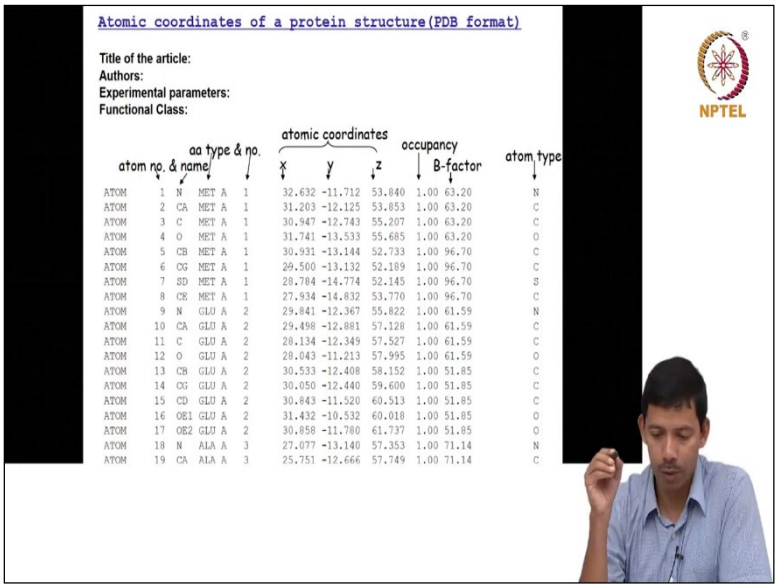
(Refer Slide Time: 03:09)



So, as I said that the initial coordinates you can get from protein data bank so at the moment there are more than 80,000 experimentally determined protein structures available in protein data bank. So, in the protein data bank you can search here with a pdb ID so each of the deposited protein has a ID given by the bank by this protein data bank. So, you can either search with a PDB ID or you can if you do not know the PDB ID of that particular protein of interest then you can also search by a keyword.

But the problem of keyword is if you just say HIV protease then for each a protease there are about 600-700 structures all the structures will come, but if you want one particular structure give that one ABC and that will with that PDB ID you just get one protein of interest and you can start from there.

(Refer Slide Time: 04:17)



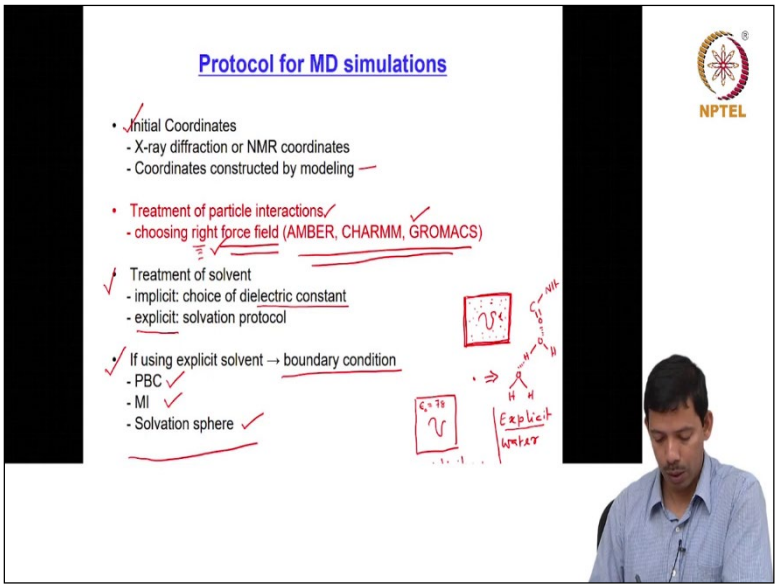
So this is a general format of a PDB file so as I said that the microstate in a computer simulation basically a set of ri pi. So, this ri is nothing but X Y Z, X Y and Z coordinates and in PDB you exactly have that you have the x coordinate y coordinate and z coordinate for each of atom. So, here so in the at the top you have the title of the article the journal where this particular work was published the name of the author's experimental parameter and resolution at which structure was deposited and what is the functional class in this case we are discussing about the protease.

So, now this molecule this protein to have multiple amino acids so this is amino acid number one which is methionine, so, in methionine we have this many atoms nitrogen C alpha carbon in carbon covering oxygen C beta C gamma and so on. So, for each atom I have x y and Z coordinates. Then we have the second residue glue and for each of these atoms we have X,Y and Z coordinates. So, if we put X Y Z coordinates of each atom so this is atom XYZ coordinates sorry X Y Z coordinate of atom 1, atom 2, atom 3 and so on so forth.

When we join them together we basically get a 3-dimensional structure of the protein and that is where we start from here. So, we start from here and this is my initial microstates the microstate number 1 and from here on I generate several other microstates to look how my protein conformation changes over time and if I want to get the average over then I take the time of average of ensemble average.

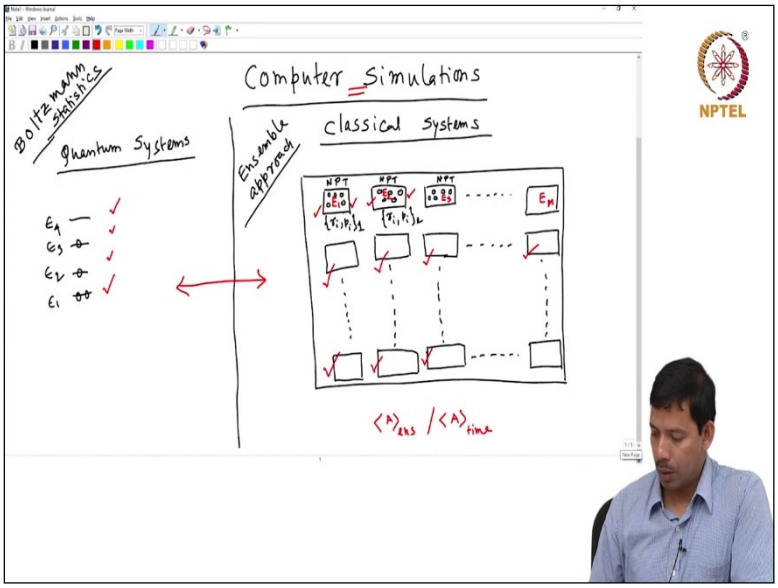
So this is about the structure but you can always calculate different thermodynamic quantity of the system.

(Refer Slide Time: 07:02)



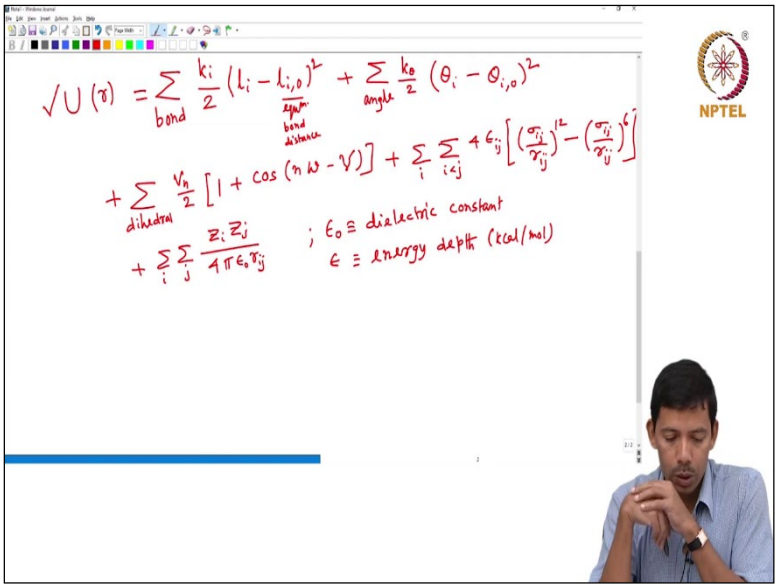
So, seen the protocol for md simulation the first step was initial coordinates you have to start from a set of initial coordinates and that you can get either from PDB protein data bank or if or if it is not available in PDB you can make a crude set of coordinates by modelling it. The second step is a very important step is called treatment of particle interactions and for that we need to go back to our discussion yesterday.

(Refer Slide Time: 07:43)



And if you recall we said that the inter particle interactions they are present or they are described by the potential energy of the system so U the potential energy it contains the inter particle interaction information.

(Refer Slide Time: 08:16)



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

And that you what U written yesterday was composed of several intramolecular and intermolecular interactions I need to write it again to explain you the protocol number two in the MD simulation. So the bond was here ki by 2 li - li 0 square where li series equilibrium, equilibrium bond distance or the bond length then we had angle theta i by 2 I am sorry k theta i by 2 the first constant theta i - theta equilibrium square plus we had some overall dihedral and the expression was Vn by 2 1 + cos n Omega - gamma and then we had Lennard-Jones sum over i sum over i less than j or epsilon is i j Sigma ij r ij to the power 12 - Sigma ij to the power 6 plus we had some of sum over i some over j zi zj by 4 pi epsilon 0 r ij.

Here epsilon 0 is the dielectric constant of the median and Epsilon epsilon 0 and epsilon is basically the energy depth. So, this is having a minute of kcal per mole so not to confuse epsilon 0 and epsilon. So, this was the functional form of the U which do which contains the inter particle interactions. Now when we saw about the treatment of particle interactions we need to look at these values because ki li k theta so all these quantities they basically decide the interactions among the particles.

So, when the first term denotes bond interaction second term denotes a angle angular interactions and dihedral interactions a third term and the last two terms indicate the inter molecular interactions. So, in this whole expressions as I said before that li is an instantaneous quantity so li varies when the bond stretches my li could be one 1.51 angstrom at some point in another instant my li at time t + delta t it could be 1.54 Armstrong.

So, li is basically the instantaneous value of the bond distance likewise theta i is the instantaneous value of the angle but the other quantities like ki li k theta k theta theta i to theta Vn sorry theta i n Omega epsilon the Sigma Zi Zj and epsilon 0 they are the fixed quantities in other words if you have a SP3, SP3 bond SP3 carbon SP3 carbon their equilibrium bond distance equilibrium is fixed this equilibrium distance is nothing 1.53 angstrom.

And likewise the k is also unknown value, so this equilibrium values they do not change what changes is this instantaneous value of the bond likewise for angle also that that equilibrium angle of water is a fixed quantity and around that equilibrium angle of water the water molecule vibrate. So, we can make a library where I have ki I have li 0 I have a k theta I have theta i 0 I have V n, n gamma I have epsilon I epsilon I have Sigma and then I make this library and I use this library whenever I see a molecule with CC .

So whenever I see a CC bond I basically instead of asking me to input the value my algorithm as soon as find carbon carbon SP3 SP3 bond it will get this value from the library and that library is called the force field. So, the force field is a very important term in computer simulation methods. So, forcefully basically contains the interaction parameters for various entities for various atoms and those force field parameters are kept in a library and the algorithm basically looks for the similar bond and angle at the dihedral and then gets those parameters from the library and use it for your system of interest.

So, in the MD simulation protocol the important step is choosing the right force field so when I say choosing the right force builds so the right depends on your system of interest. So, if you are simulating membrane protein versus water soluble protein then you need to choose different force field some forcefully are good for many proteins some force field are good for water-soluble proteins. So, dependent so that is again the conceptual understanding of you about the system.

So once you understand your system then you choose a write force field and that force field will give you the better description of the system and therefore you can reproduce the experimental data way better. So, here I use an some names of certain MD simulation programs who which basically can help you getting the different microstates some of them are free max is a free software for molecular dynamics and all of this software do have different force field and it is up to you to choose the right for speed for your system of interest.

Then we need to; this is an important step which is the treatment of solvent so far whenever I talked about generating the microstate I if you recall I always put my protein in a box say this is my protein I put my protein in a box and this we call a simulation box. So, simulation box is basically a dimension in which we put the protein we put the protein at the center of the box. So, that protein has some boundary and we put lot of water molecules around this protein to mimic the cellular environment.

Since our body is having 60 to 70% water we mimic that by putting the protein molecule in box full of water so now when you put water around the protein there are two ways of describing the protein oh the water so one way of describing the water is as I dream that I do just as simple as dot but we can we can describe this water molecule as HOH so where my each dot here corresponds to this where I have described the water molecule explicitly so this is the explicit description of water explicit water.

Where I have HOH and this water molecule would be able to make a hydrogen bond with another water molecule it will be able to form from H bond with the protein peptide bond and so on so forth. So, this is the description of explicit water you can have another way of describing water is that instead of described in the water explicitly you assume that I have a dielectric medium here of epsilon 0 78 so that is a implicit water model.

So, in the implicit water my water does not have the capacity of forming hydrogen bond with each other neither it has a capacity of up making hydrogen bond with the protein molecule. So, as you can see that when you have explicit description of the water so we have much more molecules than much more molecules and therefore the interactions to consider than implicit water where I have just a dielectric medium.

So, treatment of the solvent is also important so if you are using a implicit solvent you choose a dielectric constant of the medium if it is water we choose epsilon 0 as 78. If you are using the explicit water then this is the explicit description of water but then you have different model of water which is which is beyond the scope of this course. So, above the different explicit water models, so, if you are using explicit solvent then there are certain boundary conditions which you need to implement.

Some of those boundary conditions are called periodic boundary condition minimum image convention and salvation sphere which I will now describe one by one.