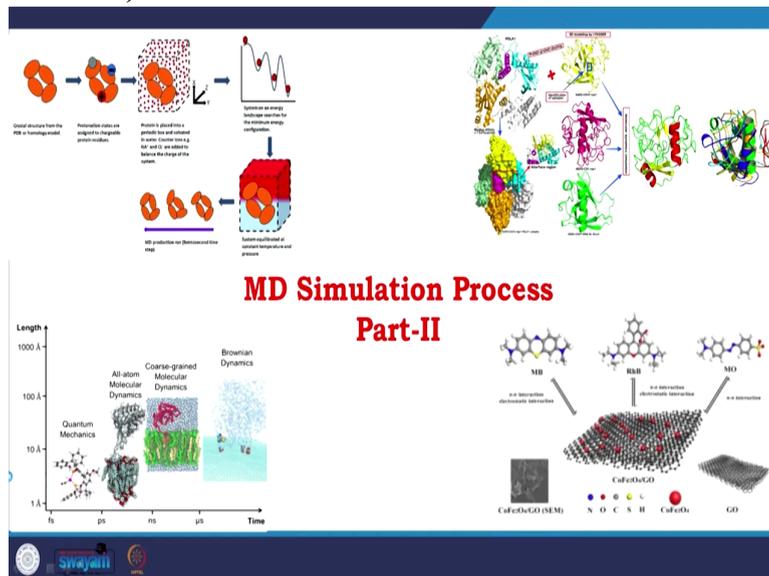


Structural Biology  
Prof. Saugata Hazra  
Department of Biotechnology  
Indian Institute of Technology, Roorkee

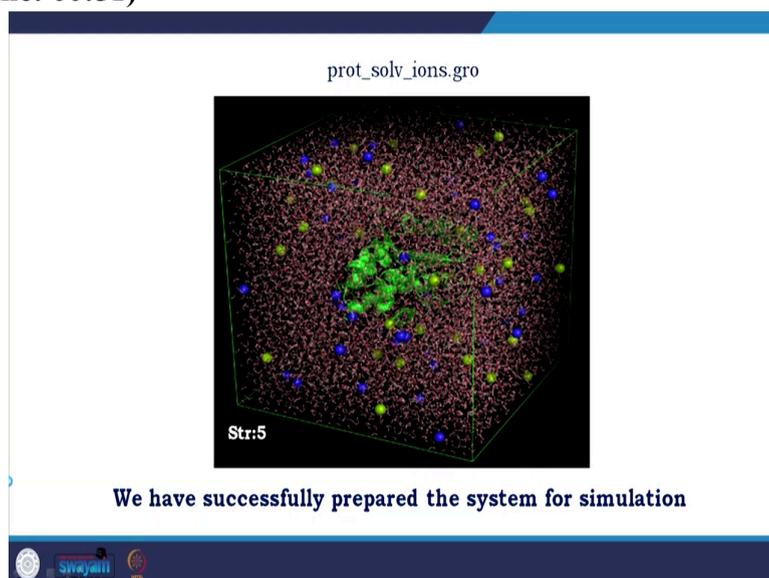
Lecture-48  
Molecular Dynamic Simulation Process Part II

(Refer Slide Time: 00:33)



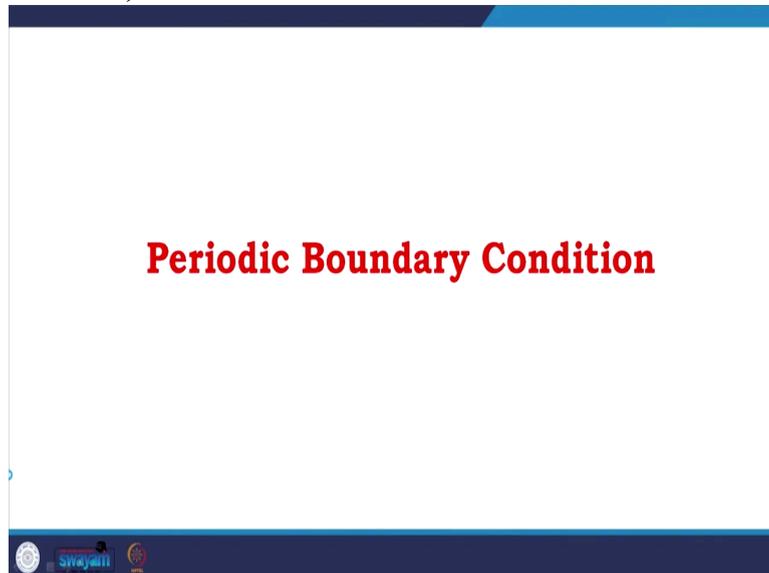
Hi everyone, welcome to the structural biologic course; again, we are continuing with MD simulation, and from the previous class, we are discussing the process of running a sample through MD simulation, which is the MD simulation process. So, today we will discuss the next part of the MD simulation process.

(Refer Slide Time: 00:51)

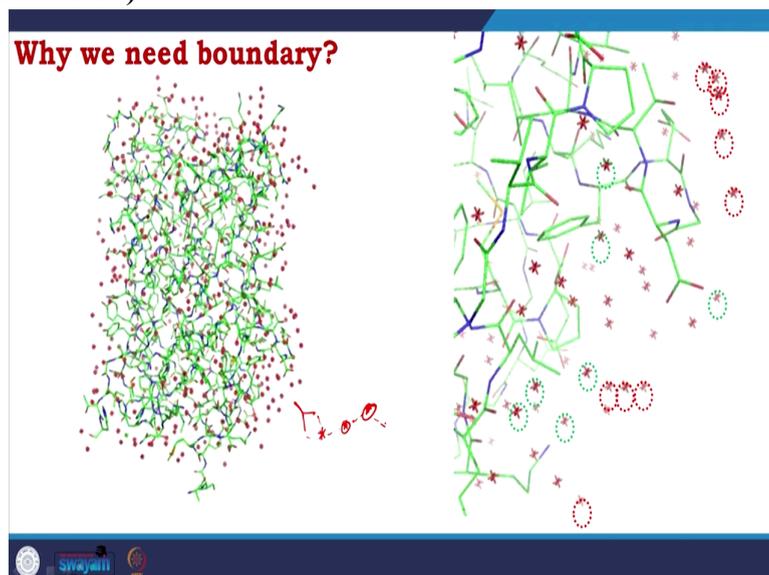


In the previous class, we prepared protein underscore to solve underscore ions dot gro. You see, that is the setup where you get the protein, the solvent, the box generated, and the ions now. We will proceed from that.

(Refer Slide Time: 01:21)



(Refer Slide Time: 01:23)

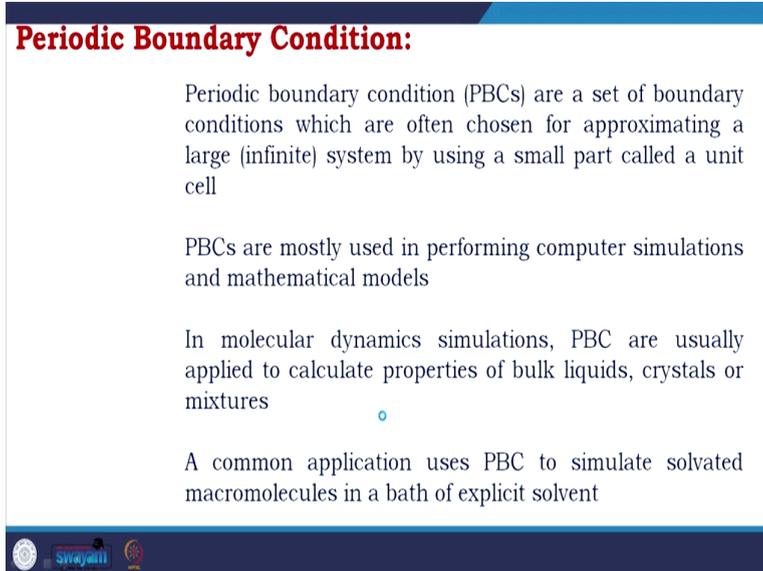


But, before going into this, certain things should be discussed in detail. One of them is the periodic boundary condition. So, first of all, why do we need boundaries? This represents a protein structure where you see the protein is there, but it is extensively solvated. This is a structure that is coming from the RCSB. So, if you look at the zoomed version, it is an experimental structure.

You will see that some water molecules are essential for the interaction with the macro molecule; I have circled other water molecules in red that is not essential. So, when you set

up a process, you have to create a cut-off; otherwise, if you see literally, one water molecule is supposed to interact with an amino acid. Now, some water molecule hydrogen bonding with this, then other water molecules bonding with that would continue. So, your limit would be infinity. But you need a finite system, so we need a boundary.

**(Refer Slide Time: 02:59)**



**Periodic Boundary Condition:**

Periodic boundary condition (PBCs) are a set of boundary conditions which are often chosen for approximating a large (infinite) system by using a small part called a unit cell

PBCs are mostly used in performing computer simulations and mathematical models

In molecular dynamics simulations, PBC are usually applied to calculate properties of bulk liquids, crystals or mixtures

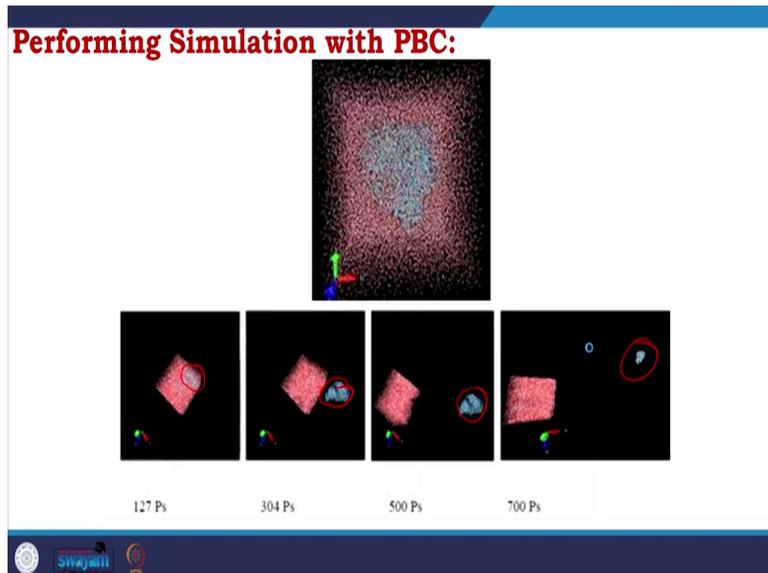
A common application uses PBC to simulate solvated macromolecules in a bath of explicit solvent

What is a periodic boundary condition? Periodic boundary conditions are a set of boundary conditions often chosen for approximating a large system using a small part called a unit cell. So, here as I told you do not want an infinite system to create because the larger the system, the more computational power. So, you want to make it smaller, and you cannot create a real boundary for making it smaller. So, you make an imaginary boundary by making some boundary conditions.

Periodic boundary conditions are mostly used in computer and mathematical models, as we will use here in molecular dynamics simulation. They are usually applied to calculate the properties of bulk liquid crystals or mixtures. A common application uses periodic boundary conditions to simulate solvated macromolecules in a bath of explicit solvent, and we have already discussed explicit solvent.

The explicit solvent and implicit solvent and the process we are going through the MD simulation process we are continuing here. If you remember, in the last class I talked about, we selected tip 3 p water molecule, which is part of the explicit solvent model.

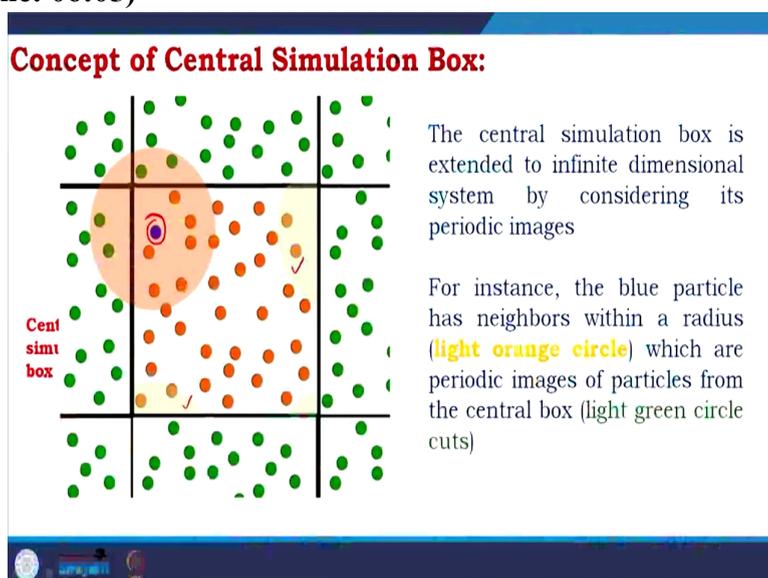
**(Refer Slide Time: 04:40)**



So, there are very interesting problems in choosing periodic boundary conditions. Let us say you choose periodic boundary conditions and a single unit cell. Here is the run where a single unit cell water molecule is chosen, a protein is kept in the center of the box, and the simulation run is developed. The protein reached the corner if you look at 127 picoseconds of the run. At 304 Pico seconds, it is already out of the solvated system. It is already in a vacuum.

In 500 Pico seconds, it is far away in 700 days, even mentoring more distance. So, this is not possible. So, what you have to do in addition to applying the boundary condition, you have to develop a central simulation box, and then you have to replicate that around you have to make layers.

(Refer Slide Time: 06:05)



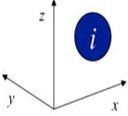
Now, you have to remember, as I told they could be going from the center to the other one. As you could have seen here, the central simulation box is extended to an infinite dimensional system by considering its periodic images. It is not the problem here if you look at the blue particle. This one has neighbors within a radius if you see which is represented by the light orange circle, which is a periodic image of particles from the boxes, these boxes, which are light green circle cuts. So that way, we set and simulate without a problem.

(Refer Slide Time: 06:55)

**Phase Space:**

In dynamical system theory, a **phase space** is a **space** in which all possible states of a system are represented, with each possible state corresponding to one unique point in the **phase space**

For a model system, the **phase space** usually consists of all possible values of **position** and **momentum** variables



A given atom is represented by its time-dependent position and momentum vectors

At a given instant, consider a plot of the positions and momenta of all N atoms in the system

**6 N-dimensional plotting domain** → **Phase-Space**

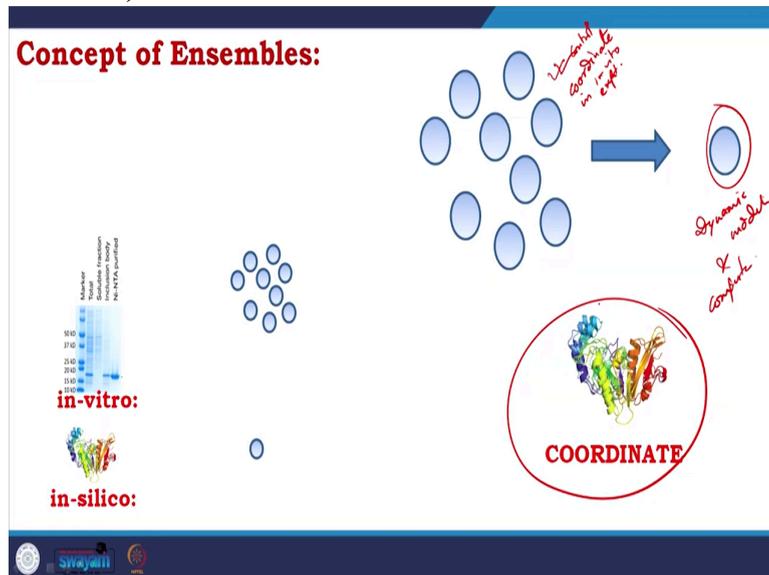
I have come to another important component of the simulation run, phase space. First, in dynamical system theory, a phase space represents all possible states of a system, with each possible state corresponding to one unique point in the phase space. Let me explain a little before that for a model system, the phase space usually consists of all possible position and momentum variables values. So, we are looking at the position and momentum of the model system.

So, what should be the starting one, a coordinate we have already discussed? Let us look at a model; Its time-dependent position and momentum vectors represent a given atom, the IMO atom, at a given instant. Consider a plot of the position and momenta for all the N atoms in the system. So, we start with the i atom. Now we are considering the N atom. So, there would be degrees of freedom regarding position and momenta.

So, there will be 6 N-dimensional plotting domains that 6 N-dimensional plotting domains define the phase space. Let me go into more you know details a little bit to know so that you

can understand. But before that, I have to tell you the concept of ensembles phase space and ensembles are very correlated.

**(Refer Slide Time: 08:45)**

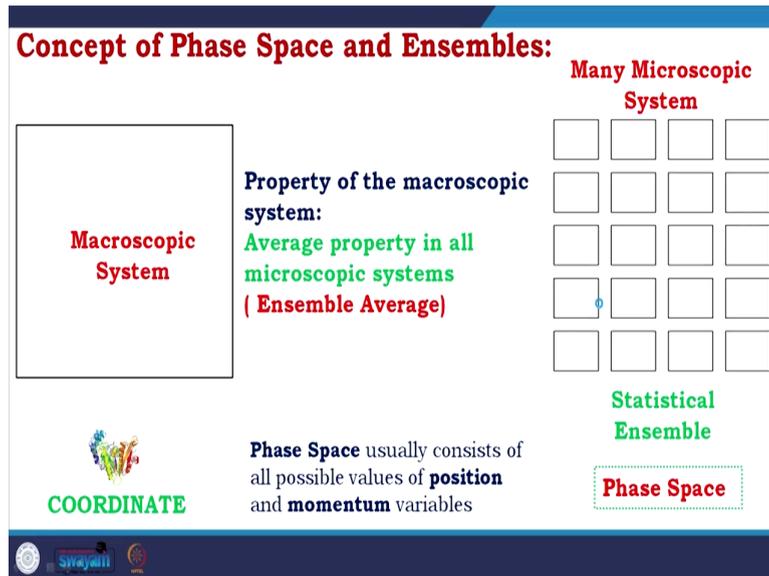


Before going, I will take you to a previous slide where I discussed the 3 systems in-vivo. I told you there is a mixture of proteins in-vitro where you have one protein, but in many numbers in-silico, where you have proteins depending on how many you want 1, and second, you have the coordinate information. So, we have removed the in-vivo system from this discussion because this is complex 1 2.

I have already discussed that when we are talking about the dynamic system, the MD simulation modeling, our goal is to mimic an in-vitro system and not an in-vivo system. So now, if we look at an in-vitro system, the many molecules there have a behavior trend, and here we are taking one, so we are making a journey of the ensemble. I am coming to discuss it in the next slide.

In addition, here, we mostly do not have coordinates. We cannot control coordinates in control or operate coordinates in in-vitro experiments, whereas, in dynamic modeling, or any computational study, we could control the coordinate change. So, as I told you now, I have come to the correlation between the phase space and ensembles.

**(Refer Slide Time: 10:52)**



So, if you think about a system, let us say about us, one person, in deadly we present a macroscopic system, and our cells represent a microscopic system even smaller the protein if it could be in a crystal, the crystal is a macroscopic system, individual protein is a microscopic system. So, a macroscopic system could be divided into many microscopic systems, and these are called statistical ensembles.

Now, the property of the macroscopic system is an average property in all the microscopic systems, called the ensemble average. As I already told you, we could have controlled the coordinate of our sample while doing a dynamic experiment. If you look at phase space, it usually consists of all possible values of position and momentum variables in the coordinate you have the position, and when a force is applied, it gets momentum. So, you have the position and your position change information. So, we could have gone into the phase space that way and connected the ensemble with the phase space.

**(Refer Slide Time: 12:41)**

## Definition and types of Ensembles:

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

**Different ensembles exist with different characteristics:**

**Microcanonical Ensemble (NVE):** The thermodynamic state is characterized by fixed number of atoms, (N), a fixed volume (V) and a fixed energy (E)

**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 fixed temperature (T)

**Grand canonical Ensemble ( $\mu$ VT):** The thermodynamic state is characterized by fixed volume (V), a fixed temperature (T) and a fixed chemical potential ( $\mu$ )

So, what is an ensemble? An ensemble collects all possible systems with different microscopic but identical macroscopic or thermodynamic states. I repeat the collection of all possible systems with different microscopic states but identical macroscopic or thermodynamic states. Different types of symbols exist with different characteristics.

The microcanonical ensemble is NVE; the thermodynamic state is characterized by a fixed number of atoms which is designated by N, a fixed volume V and a fixed energy E canonical ensemble NVT; this is the collection of all systems thermodynamic state is characterized by a fixed number of atoms in a fixed volume V and a fixed temperature T isobaric-isothermal ensemble isobaric is pressure thermal is T.

So, NPT, this ensemble is characterized by a fixed number of atoms, a fixed pressure, and a fixed temperature. The last one, grand canonical ensemble  $\mu$ VT the thermodynamic state is characterized by fixed volume fixed temperature VT and a fixed chemical potential defined by  $\mu$ . So, we have discussed periodic boundary conditions and their role, we have discussed phase space, and we have discussed symbols. Now, we will discuss another thing that is again very critical, probably the most critical thing in biology: energy minimization.

**(Refer Slide Time: 14:39)**

### Energy Minimization:

In the field of computational simulation, **energy minimization** or geometry optimization is the process of finding an arrangement in space of a collection of atoms where, according to some computational model, the net inter-atomic force on each atom is acceptably close to zero and the position on the **potential energy surface(PES)** is a stationary point

The collection of atoms might be a single molecule, an ion, a condensed phase, a biological macromolecule, a transition state or even a collection of any of these

The computational model of chemical bonding might, for example, be quantum mechanics

The computational model of protein movement would be Newtonian physics

Before going into biology in general, energy minimization in the field of computational simulation, also called geometric optimization, is a process to find an arrangement in the space of a collection of atoms where according to the computational model, the net inter-atomic force on each atom is acceptably close to 0, and the position on the potential energy surface is this stationary point.

So, I do not go into the second part now. I am going to explain it, but the first part is you have an atomic system, it had bond length and bond angle on all these things, it would be optimized in a way the configuration confirmation geometry that, the net inter-atomic force on each atom is acceptably close to 0, that is the major condition of energy minimization, the collection of atoms might be a single molecule.

Nowadays, a single molecule is becoming very popular for study. Iron, a condensed phase of a biological macromolecule, a transition state, or even a collection of any of these, the computational model of chemical bonding might, for example, come from calculating quantum mechanics. In contrast, the computational model of protein movement comes from Newtonian physics.

**(Refer Slide Time: 16:21)**

## Energy Minimization:

**Basic idea (Anfinsen's theory):** Protein native structure corresponds to the state with the lowest free energy of the protein-solvent system

### General procedures

#### Potential function

Evaluate the energy of protein conformation

Select native structure

#### Conformational search algorithm

To produce new conformations

Search the potential energy surface and locate the global minimum (native conformation)

When we come to protein, as I said, minimization has a huge role in studying protein, and we will return to Anfinsen's theory. When discussing protein, I talked about how Christian and Anfinsen have changed how the world looks toward protein folding through a simple experiment. Before Christian and Anfinsen, people thought protein folding was a biological phenomenon.

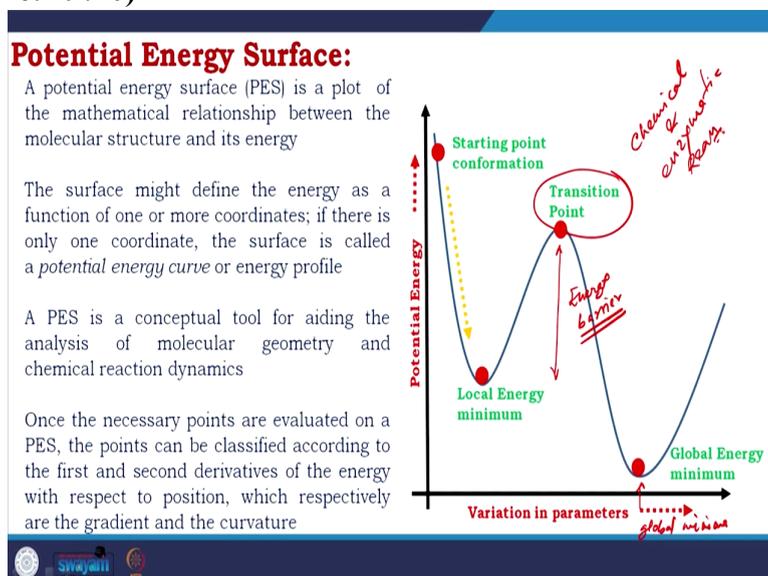
This means the interaction between the atoms, the charge, and the side. If you think molecularly, you will clearly understand there is physics, and there is chemistry, but before you have those molecular concepts, people used to think that this folding was a biological phenomenon with a simple experiment, he had proved that it is a physic-chemical affair and that information changed the way people had studied protein before people wanted to purify a protein only from the inside only from the biological system.

Suppose you look at what I talked about these before if you looked at before people have most in primitive times. When people started working on protein, they mostly worked on protein from the blood, myoglobin, and hemoglobin because they collected the blood and worked from them. But Christian and Anfinsen's theory, his experiment, which gave him a Nobel Prize, have changed the world's ideas.

So, how we could do energy minimization, one thing is to understand the potential function, then evaluate the energy of protein conformation, before people used to know that every protein at one conformation, but now we know that protein functioning changes their confirmation, so we have to evaluate we have to select the native structure and go from them

conformational search algorithm to produce new conformation search the potential energy surface and look at the global minima. So the potential energy surface I did not talk about is coming again.

(Refer Slide Time: 19:20)



So let us first discuss the potential energy surface. So potential energy surplus is a map. You will see here in the X coordinate, there is a variation in parameter, and in the Y coordinate, there is potential energy. So potential energy surface plots the mathematical relationship between the molecular structure and its energy. So on one side, energy in one side changes what is defined by the variation in the parameters.

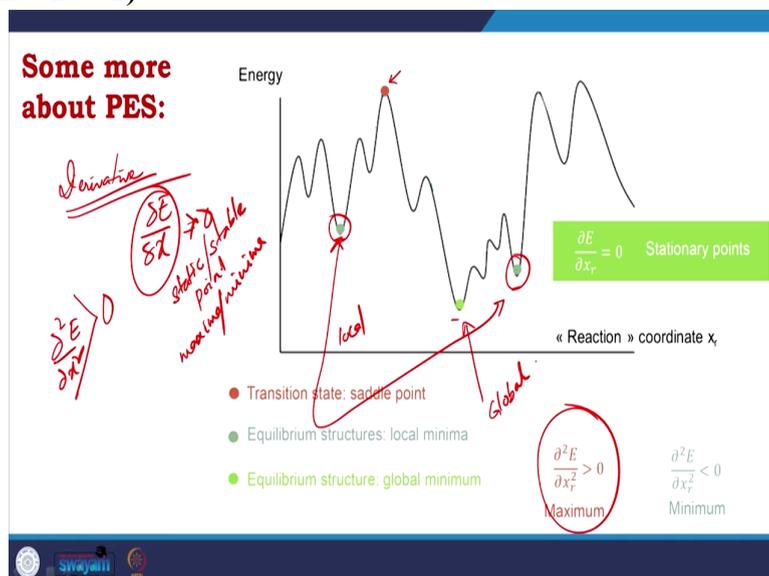
The surface might define the energy as a function of one or more coordinates. The surface is called a potential energy curve if only one coordinate exists. So, what you are looking at here, X and Y axis is given, so it is a 2D curve, but it could be multi-dimensional also. A potential energy surface is a conceptual tool for aiding the analysis of molecular geometry and chemical reaction dynamics.

Once the necessary points are evaluated on a potential energy surface, the points can be classified according to the first and second derivatives of the energy with respect to the position, which is the gradient and the curvature. So, you could have got the trend, how it is happening, and where it is happening. Let us discuss this here. So, at first, the position is the starting point of confirmation.

Now, if you apply minimization, if you do the lowering of energy, so, when we are talking about one thing, I should go back and talk again is the words I did not utter here. The protein's native structure corresponds to the state with the lowest free energy of the protein solvent system. That is a hypothesis that gives us a proper drive. So, if we have a card like that, the lowest point would be considered a global minimum.

So, returning to that, this is your starting point: apply energy minimization and reach the local energy minimum. There are other points; one I already talked about is global energy minima, and another is the transition point. Also, this is an energy barrier. The transition point is important for pupils studying chemical and enzymatic reactions. Identifying the transient state is critical to understanding a reaction and its proceedings in drug design.

**(Refer Slide Time: 23:01)**



So, if you see, I have shown you a very simple potential energy surface, it could be much much complicated, and in the case of protein, if the funnel like which we have already discussed, when you are talking about potential energy surface to measure it, we perform the derivative calculation the first derivative which is  $\frac{dE}{dx}$ , if that is 0 then we know that we have reached a static or stable point which you could also talk about as maxima and minima.

But then, when we go for double differentiation  $\frac{d^2E}{dx^2}$ , if  $\frac{d^2E}{dx^2}$  is greater than 0, it is maximum whereas, when you are getting  $\frac{d^2E}{dx^2}$  less than 0, then it is minimum, but among minimum if you see these are local minima whereas these represent global minima.

**(Refer Slide Time: 25:02)**

**The stationary points of the PES:**

Minimum  
 $\frac{\partial^2 f(x,y)}{\partial x^2} > 0; \frac{\partial^2 f(x,y)}{\partial y^2} > 0$

Maximum  
 $\frac{\partial^2 f(x,y)}{\partial x^2} < 0; \frac{\partial^2 f(x,y)}{\partial y^2} < 0$

Saddle point  
 $\frac{\partial^2 f(x,y)}{\partial x^2} > 0; \frac{\partial^2 f(x,y)}{\partial y^2} < 0$

- **Stable states:** minimum of the PES with respect to every coordinates
- **Transition states:** minimum in every normal direction but one, for which it is maximum
- Stationary points thus correspond to important states of the system

As I told the stable states, the minimum of the potential energy surface with respect to every coordinate, the transition state, or the saddle point is minimum in every normal direction, but one for which it is maximum stationary points that correspond to important states of the system. If you see, for minimum,  $\frac{\partial^2 f}{\partial x^2} > 0$ ,  $\frac{\partial^2 f}{\partial y^2} > 0$ , for maximum  $\frac{\partial^2 f}{\partial x^2} < 0$ ,  $\frac{\partial^2 f}{\partial y^2} < 0$ , for saddle point  $\frac{\partial^2 f}{\partial x^2} > 0$ , but  $\frac{\partial^2 f}{\partial y^2}$  is less than 0. So, by this double derivative, we could identify the maxima minima and the saddle point, which is the transition point.

**(Refer Slide Time: 25:57)**

- In practice, high dimensional surface
- Many local minima
- How to perform energy minimisation?

As I told you in practice, there are many dimensions on the surface and many local minima, so we must perform energy minimization.

**(Refer Slide Time: 26:11)**

## What it is going to describe:



Either a molecule or ensemble of molecules having constant atom composition

A system where a chemical reaction occurs

Relative energies for conformers

Example; The conformations of n-butane as the global minimum is the anti conformer; local minima are the gauche conformers, and the saddle points are the eclipsed conformations.

Energy minimization will describe either a molecule or ensemble of molecules with constant atom composition, a system where a chemical reaction occurs with relative energies for conformers, and an example I am not giving you now. Because we have already discussed that the confirmation of butane as the global minimum is the anti-conformer, local minima are the gauche conformer, and the saddle point or the eclipse confirmation. What they are talking about is, I hope you guys remember.

So, if you have a molecule like this, because these 2 sterically bulky groups are at the farthest position, they are showing the most stable state, which will be defined as the global minima. This would be represented as the global maximum when they are close.

(Refer Slide Time: 27:50)

EM is an numerical procedure to find a minimum on the potential energy surface starting from a higher energy initial structure

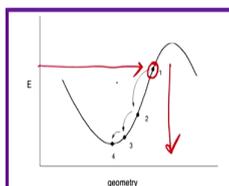


Figure 1. The process of energy minimization changes the geometry of the molecule in a step-wise fashion until a minimum is reached.

During EM the geometry would be altering in a stepwise fashion; energy of molecule is reduced from step 2 to 3 to 4.

Energy minimization is a numerical procedure for finding a minimum of the potential energy surface starting from a higher energy initial structure. In this picture, the geometry will alter stepwise energy during energy minimization. So, you see here, the first one is that high energy, and you are going down in energy with the alteration in the geometry. So, during energy minimization, the geometry would be altered in a stepwise fashion energy of the molecule will be reduced from 1 to 2 to 3 to 4.

(Refer Slide Time: 28:38)

**Determination of EM:**

Most of EM proceed by determining the energy & the slope of function at point 1.

- if slope is positive : it indicate the coordinate is too large (point 1)
- if slope is negative : the coordinate is too small

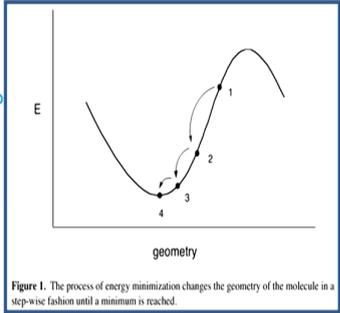


Figure 1. The process of energy minimization changes the geometry of the molecule in a step-wise fashion until a minimum is reached.

So, for determining energy minimization, most of the energy minimization method proceeds by determining the energy and the slope of the function at point 1. If the slope is positive, indicate the coordinate is too large. If the slope is negative, the coordinate is too small.

(Refer Slide Time: 28:59)

**Numerical Minimization technique:**

The numerical minimization technique then adjust the coordinate:

- if slope is positive : it indicate the value of coordinate is reduced (point2)
- if the slope is zero : a minimum has been reached
- if slope is still positive: coordinate reduced further (point 3) until a minimum is obtained.

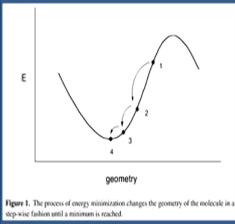


Figure 1. The process of energy minimization changes the geometry of the molecule in a step-wise fashion until a minimum is reached.

After that, you could optimize further with the numerical minimization technique, the numerical minimization technique at just the coordinate. If the slope is positive, indicate the value of the coordinate is reduced to point 2, it comes to point 2 if the slope is 0, a minimum has been reached. If the slope is still positive coordinate is reduced further to point 3 until a minimum is obtained. So this is the search for getting the minimum.

(Refer Slide Time: 29:32)

All the EM methods used to find a minimum on the potential energy surface of a molecule use an iterative formula to work in a step-wise fashion.

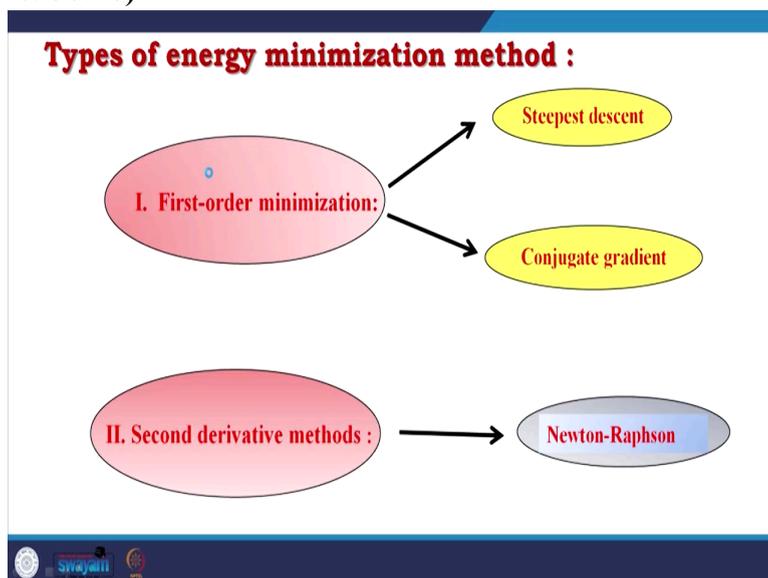
These are all based on formulas of the type:

$$X_{\text{new}} = X_{\text{old}} + \text{Correction}$$

In all these methods, a numerical test is applied to the new geometry ( $X_{\text{new}}$ ) to decide if a minimum is reached.

All that energy minimization method used to find a minimum on the potential energy surface of a molecule uses an iterative formula to work stepwise. If you remember, this is one thing I have talked about continuously. Iteration is what a computer does for you. These are all based on the formula of type  $X_{\text{new}} = X_{\text{old}} + \text{the correction}$ . A numerical test is applied to the new geometry in all these methods to decide if a minimum is reached.

(Refer Slide Time: 30:10)



There are 2 types of energy minimization, first-order minimization and second-derivative method second-order minimization. The first-order 2 popular methods are Steepest Decent and conjugate gradient, and the second-order is Newton Rapson.

(Refer Slide Time: 30:30)

**Steepest Decent Method:**

The second derivative is assumed to be constant, the equation to update the geometry becomes

$$X_{\text{new}} = X_{\text{old}} - Y E' (X_{\text{old}})$$

Where ; Y is a constant

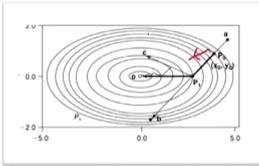


Steepest Decent, the second derivative here, is assumed to be constant. The equation to upgrade the geometric becomes  $X_{\text{new}} = X_{\text{old}} - Y E'$  where Y is a constant.

(Refer Slide Time: 30:48)

**Steepest descent algorithm:**

- The derivative vector from the initial point  $P_0(x_0, y_0)$  defines the line search direction
- The derivative vector does not point directly toward the minimum (0)
- The negative gradient of the potential energy (the force) points into the direction ( $P_0 \rightarrow b, P_1 \rightarrow c$ ) of the steepest descent of the energy hyper surface and is always oriented perpendicular to energy isosurfaces.



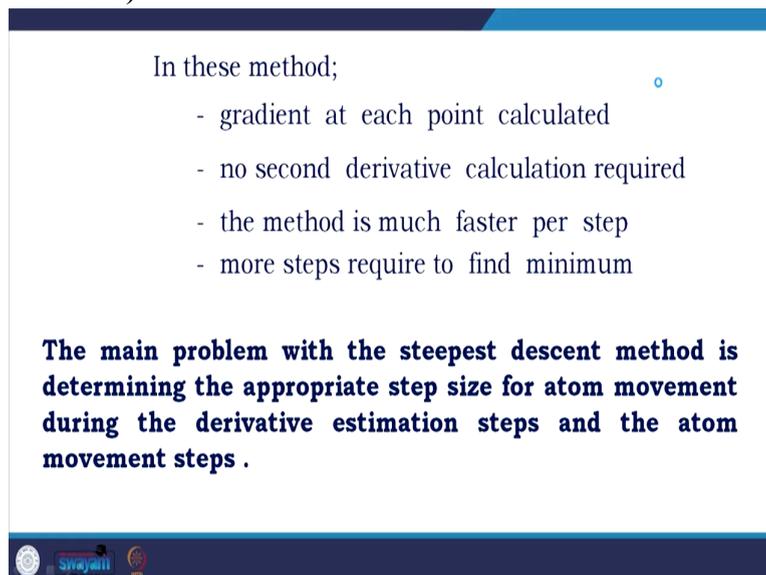
The method is named Steepest Descent because the direction in which the geometry is first minimized is opposite to the direction in which the gradient is largest (i.e., steepest) at the initial point.



In the steepest descent, if you see the picture, it starts from the initial point's derivative vector, defining the line search direction. So, it could directly go there, but it does not do that. The derivative vector does not point directly toward the minimum, the minimal point of the negative gradient of the potential energy. The force points into the direction  $p_0$  here to  $b$ , and then  $P_1$  reaches  $P_1$  by going in the direction of  $b$ ; it reaches  $P_1$  and then goes to  $c$ , which is perpendicular.

So, the negative gradient of the potential energy points into the direction P 0 to b, then P 1 to c of the Steepest Decent of the energy hypersurface and is always oriented perpendicular to the energizer surface. So, it comes in one direction, then go perpendicular, then comes in another direction go perpendicular. In that way, it reached the minima. The method is named steepest descent because the direction in which the geometric is first minimized is opposite to the direction in which the gradient is largest or the steepest at the initial point.

(Refer Slide Time: 32:15)



In these method;

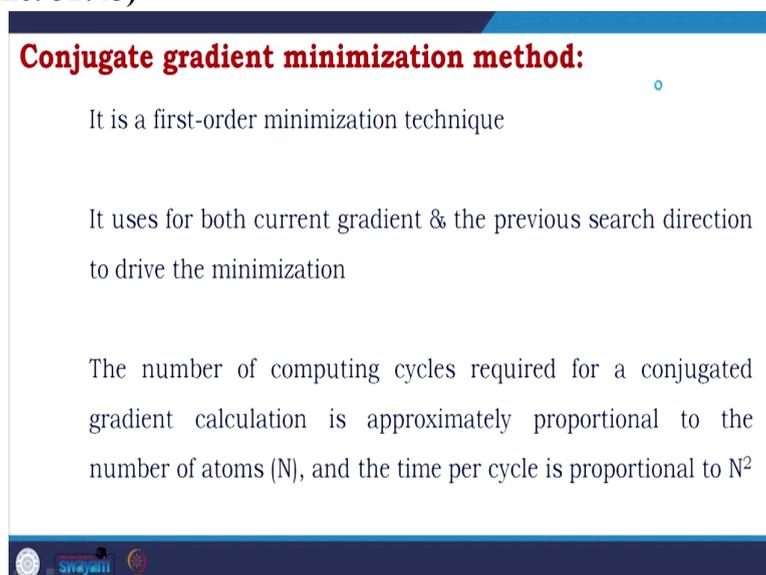
- gradient at each point calculated
- no second derivative calculation required
- the method is much faster per step
- more steps require to find minimum

**The main problem with the steepest descent method is determining the appropriate step size for atom movement during the derivative estimation steps and the atom movement steps .**

swayam

In this method gradient at each point is calculated. No second derivative calculation is required. The method is much faster per step; more steps are required to find the minimum. The main problem with the steepest descent method is determining the appropriate step size for atom movement during the derivative estimations steps and the atom movement steps.

(Refer Slide Time: 32:45)



**Conjugate gradient minimization method:**

It is a first-order minimization technique

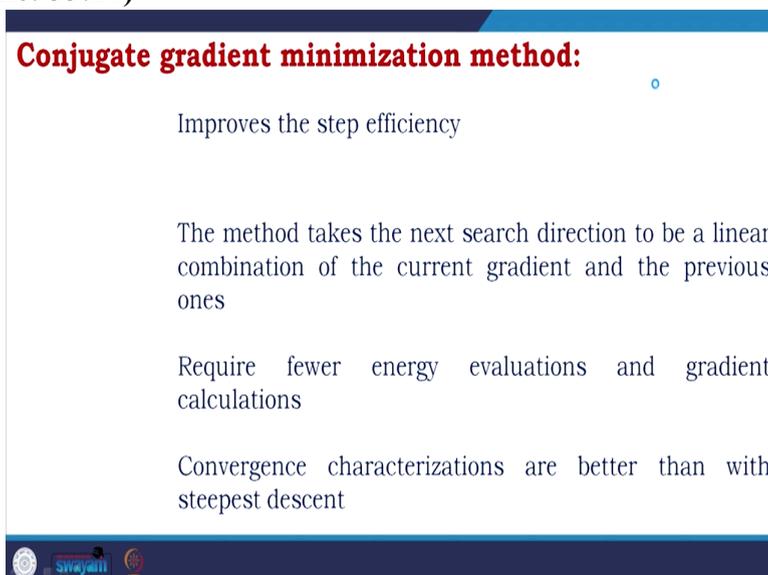
It uses for both current gradient & the previous search direction to drive the minimization

The number of computing cycles required for a conjugated gradient calculation is approximately proportional to the number of atoms (N), and the time per cycle is proportional to  $N^2$

swayam

Coming next the conjugate gradient: This is also a first-order minimization technique. It uses the current gradient and the previous search direction to drive the minimization. The number of computing cycles required for a conjugated gradient calculation is approximately proportional to the number of atoms. So, increasing the number and the time per cycle proportional to N square improves the step efficiency.

(Refer Slide Time: 33:17)

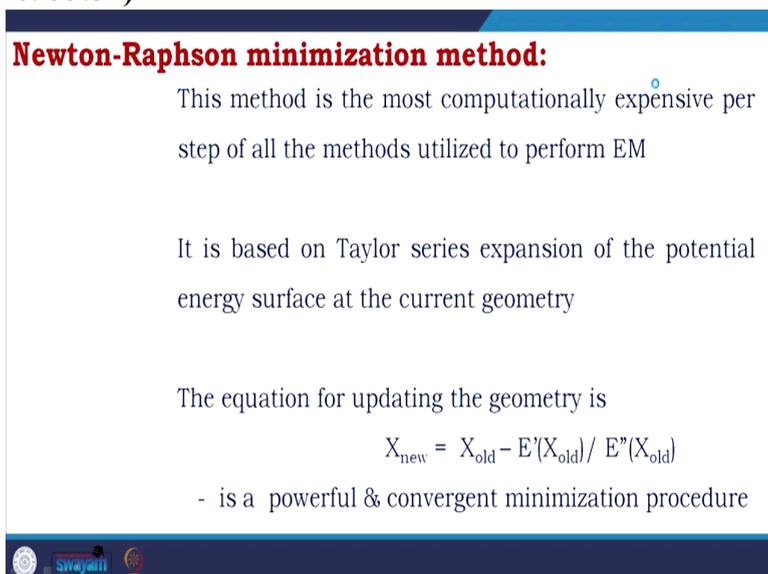


**Conjugate gradient minimization method:**

- Improves the step efficiency
- The method takes the next search direction to be a linear combination of the current gradient and the previous ones
- Require fewer energy evaluations and gradient calculations
- Convergence characterizations are better than with steepest descent

The method takes the next search direction as a linear combination of the current gradient, and the previous ones require fewer energy evaluations and gradient calculations. Convergence characterizations are better than with Steepest Descent.

(Refer Slide Time: 33:37)



**Newton-Raphson minimization method:**

This method is the most computationally expensive per step of all the methods utilized to perform EM

It is based on Taylor series expansion of the potential energy surface at the current geometry

The equation for updating the geometry is

$$X_{\text{new}} = X_{\text{old}} - E'(X_{\text{old}}) / E''(X_{\text{old}})$$

- is a powerful & convergent minimization procedure

Coming to Newton Raphson's minimization method; is the most computationally expensive per step of all the methods utilized to perform EM because of the secondary weighting. It is based on the Taylor series expansion of the potential energy surface at the current geometry,

and you have already seen how we use the Taylor series in different cases. The equation for updating the geometries  $X_{\text{new}} = X_{\text{old}} - E' X_{\text{old}} / E''$  where  $E'$  is the first derivative and  $E''$  is the second derivative. This is a powerful and convergent minimization procedure.

(Refer Slide Time: 34:15)

**Newton-Raphson minimization method:**

- The advantage of the Newton-Raphson procedure is that the minimization could converge in one or two steps.
- The major drawback is that this method requires the calculation of the second derivatives.
- The minimization can then become unstable when a structure is far from the minimum (or the energy surface is anharmonic).

The advantages of the Newton-Raphson procedure are that the minimization could converge in 1 or 2 steps. The major drawback is that this method requires the calculation of the second derivative. So, computationally expensive, the minimization can become unstable when a structure is far from the minimum or the energy surface is anharmonic.

(Refer Slide Time: 34:36)

**Advantage :**

- Only one iteration for quadratic functions
- Efficient (relative to first-order methods)  $N/N-1 = (N-1)/(N-2)^2$
- Better energy estimate

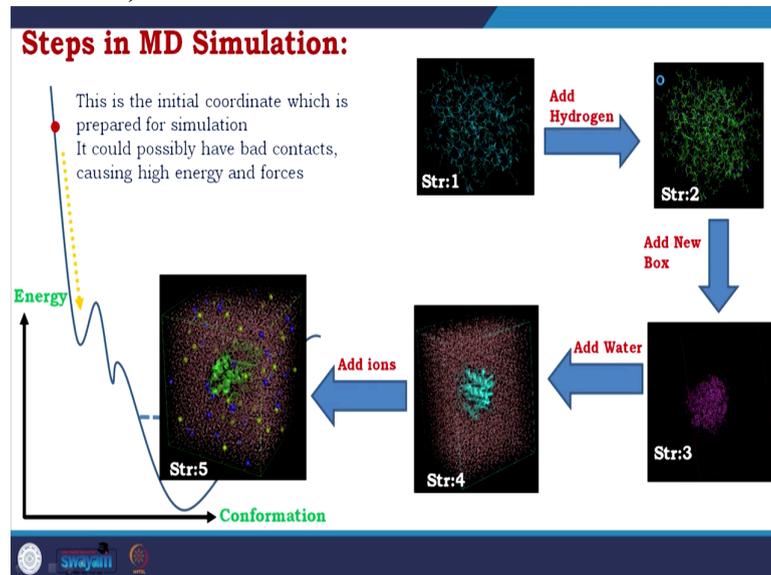
**Disadvantages :**

- $N^2$  storage requirements (compared to  $N$  for conjugate gradient)
- Involves calculating Hessian (~10 times time for gradient calculation)
- It is used in transition-structure searches (saddle point locator)

Advantage: only one iteration for quadratic functions, efficient relative to first-order methods, better energy estimate. Disadvantage:  $N^2$  storage requirement compared to  $N$  for conjugate gradient, because of the second derivative involves calculating Hessian which is 10 times more gradient calculation. It is used in transition structure searches, where we have

shown the saddle point. Saddle point, as I told is very important for identifying the transition state of reaction or the transition state or intermediate of a biochemical reaction.

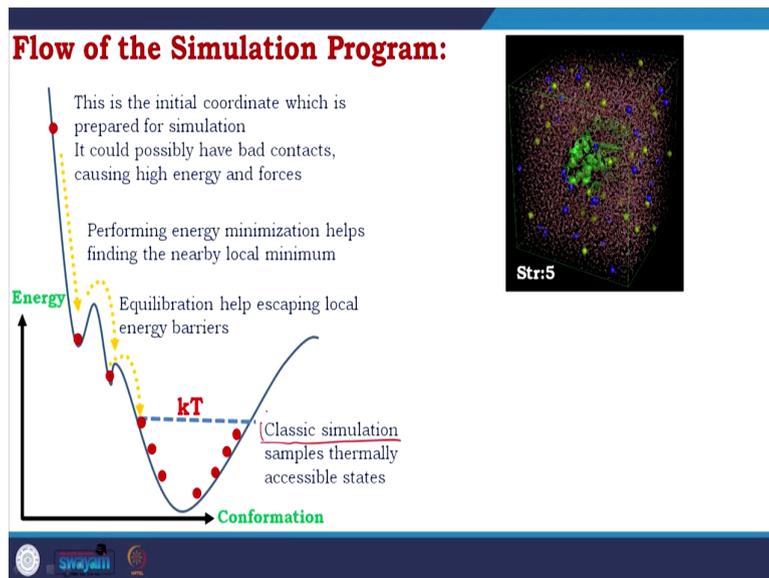
**(Refer Slide Time: 35:19)**



We are coming back to the process estimation. So, we will look at the steps in MD simulation. First, we talked earlier about the initial coordinate prepared for the simulation, but, as I told you earlier, you could not directly go to the simulation from this initial coordinate. What do you have to do? You have to do a lot of processes like adding hydrogen coming to structure 2, then you have to develop a new box, then you have to solve it by adding water, and finally, you have to add ions.

So, this is where we have finished the last lecture. When we came here, our structure could be processed now, but it could have bad contacts causing high energy and force. So, there is a requirement of energy minimization. So, the next step in MD simulation would be energy minimization.

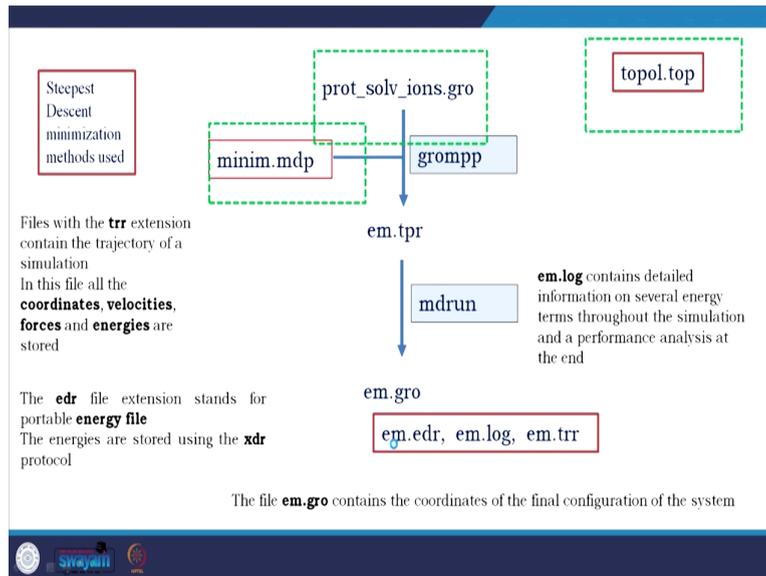
**(Refer Slide Time: 36:35)**



So, coming to the flow again, we have the initial coordinate. We process them we come to the structure where we have the protein. It is now in a box, and the periodic condition is applied solvated with tip 3 p ions to buffer eyes, so everything is there, but now, it could have bad contacts. So, we have to go for minimization. So, we have to perform energy minimization. It would help to find the nearby local minimum.

So, if you perform minimization, you reach here, what next from this local minima to other local minima, you need equilibration. We talk about ensembles now, which could be used to equilibrate the sample to escape the local energy barrier. Then we will come near the global minima, and the classic simulation samples would do this. We will get the thermally accessible states. So, these are mostly globally minimized. We could see the movements and all, but one thing you must understand here is that I see the classic simulation, which is why I will talk in the last class of this module.

**(Refer Slide Time: 38:20)**

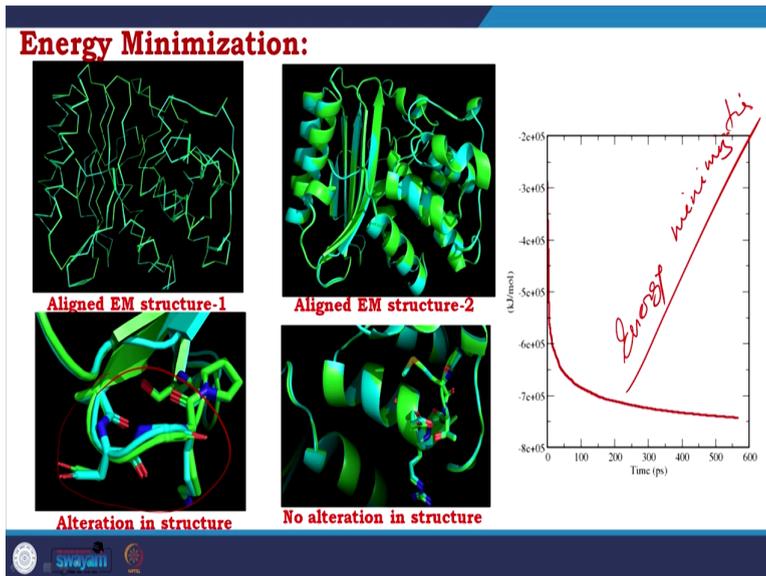


So, with protein underscore solver underscore ions dot grow, we have taken these 3. If you remember, it is a similar process to the minim dot MDP. The dot grows, and dot top files are converted into em dot tpr using the common grumpy here, Steepest Descent minimization method used. We have already discussed what the steepest descent minimization method is. As I told you, we are going to the MD run, the minimization run, and after the MD run.

We get em dot gro the gromacs competence file. But in addition to the em dot grow, we will also get em to dot edr, em dot log em dot trr the new files let us look at files with the trr extension containing the simulation trajectory. So trr are files that are giving you the trajectory. Remember I told you that when you like any sample, you put it on the force field and apply force? It would go up, up, up, and then come down.

So create a trajectory. This file stores all coordinates, velocities, forces, and energies. What is dot edr? The edr file extension stands for portable energy file, so it would record the energy change the energies are stored using the xdr protocols, m dot log file contents detail information on several energy terms throughout the simulation and then a performance analysis at the end, which we will see in the analysis part dot gro contains the coordinates of the final configuration of the system. So, now we minimize it.

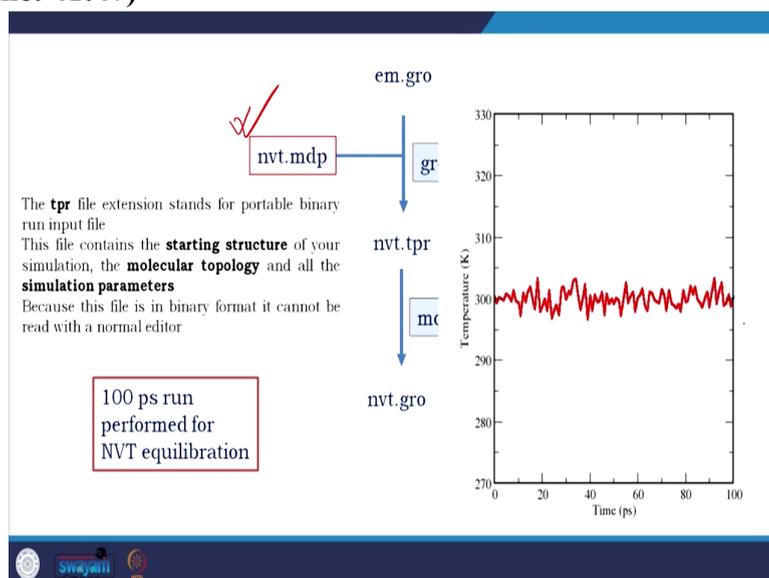
**(Refer Slide Time: 40:31)**



Let us take a look at what happened. So, here what we have done, we have aligned the energy-minimized structure with the initial structure, and we have started to look at what are the changes. And if you take a zoomed look, you will see that there are changes here, but especially when you pick up your know structures from the experimental database, you do not see many changes, but you see changes here.

You do not see many changes in other parts, but there are also little changes. So, changes are there but not much. This clearly shows you how the molecule is minimized from the high energy. So, our energy minimization is successful. In that way, we could check the energy minimization. What next?

**(Refer Slide Time: 41:47)**

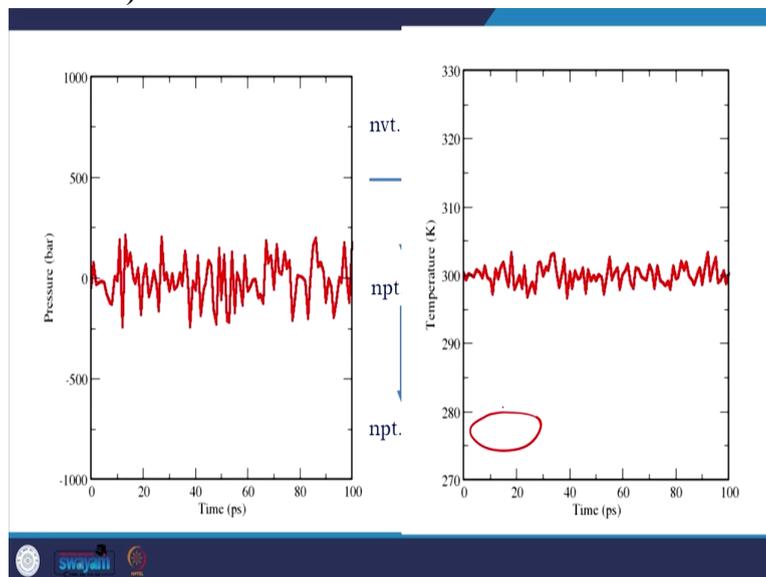


The equilibration so we get the minimized file m dot gro. We again do the same thing by applying grmpp; we come to nvt dot tpr. So now what we are doing, we are doing nvt, if you

remember these cell ensemble, what we are getting here after the MD run we get nvt dot gro, and it is a main file, additional files are nvt dot cpt, nvt dot edr, nvt dot log, nvt dot trr. For most of the files I have already talked about, the tpr file we already told also that it stands for a portable binary run input file.

It contains the starting structure of the simulation, the molecular topology, and all the simulation parameters, and because it is in binary format, it cannot be read with normal editors. You have performed 100 picoseconds run for nvt equilibration, as I told and this is what you get after nvt that how equilibrated this is a plot between temperature and time.

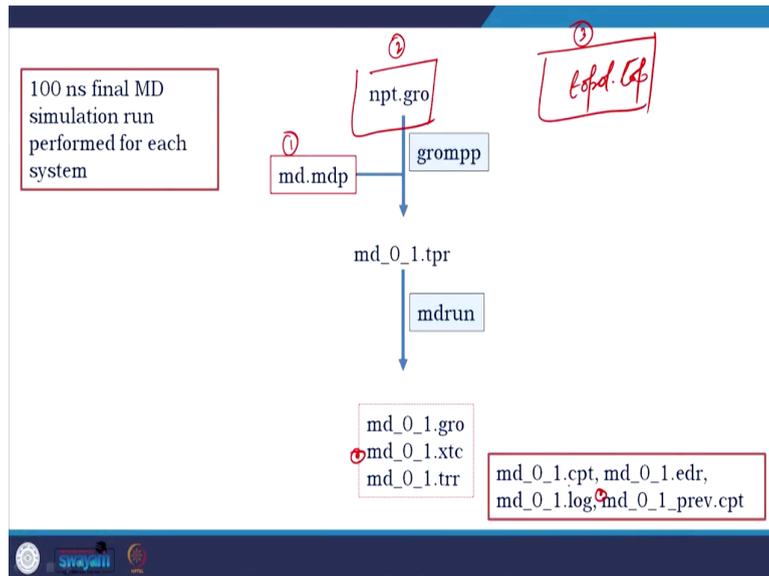
**(Refer Slide Time: 43:05)**



Now, you come here with nvt dot gro again. You do the same thing with npt dot mdp because you will do equilibration with the npt condition. So npt dot mdp nvt dot gro and topology top 3 files again march with grompp to give npt dot ppr, and then you again do the MD run you get npt dot gro with that npt dot cpt, npt dot edr, npt dot log, and npt dot trr. So, I talked about all those files except this npt dot gro. I will discuss that this is again 100 picosecond run perform for npt equilibration.

So this is the curve you get points to put pressure versus time to compare. I put the other file so you could see that this is npt. This is nvt the cpt file. We will discuss these in the last, but it is a checkpoint file. What is a checkpoint file? A checkpoint file helps you run again if your simulation is stopped for other reasons.

**(Refer Slide Time: 44:29)**

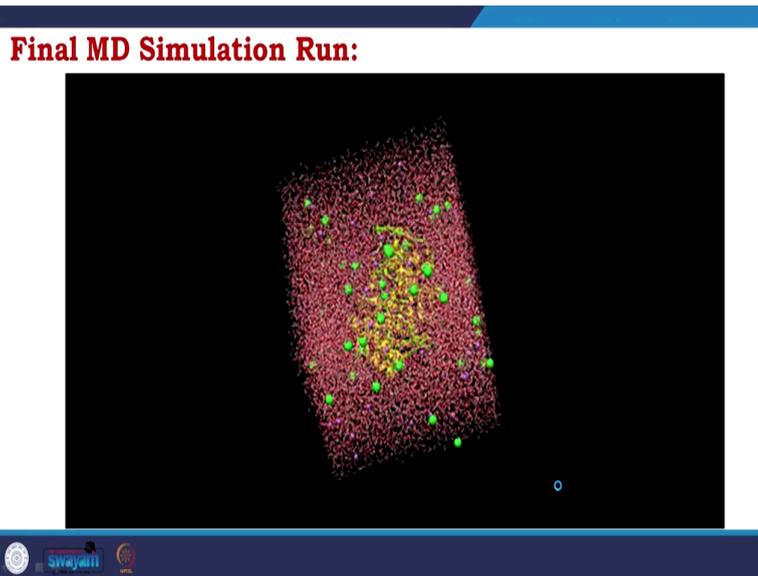


We are in the final stage and have this npt dot gro file. Again we have npt dot gro with updated to topal dot top. So npt dot gro to pull topal dot top and md dot mdb these 3 with grumpp makes md underscore 0 underscore 1 dot tpr, and when we do mdr here, we get gro file dot trr file dot cpp file dot edr file dot log file. In addition, we get dot xtc and previous dot cpt. This file was generated to take a backup because this is, you know, 100 nanoseconds run if you say 100 nanoseconds final MD simulation run performed for each system.

So if anything happened or there was any problem before, it was about 1 hour or something. Now it is like 6 days, 10 days after the end of 10 days, and if you have to rerun it, it is a loss of time. So they make a backup. So they make a checkpoint file, they make a previous checkpoint file, then think about if you are saving. Remember I talked about every 1 or 2 femtoseconds in 100 nanoseconds? How many files are created to make that dot trr the trajectory file huge?

So, while dot trr is a very big file now unzipped, you get dot xtc which has the same information but is zipped so much smaller in size. So, we have run the MD now. What is the result now,

**(Refer Slide Time: 46:39)**



You will see that it is guided me into a brand. So, you will see different sample conformers that are the end product of an MD simulation run. Now, from there, we will go to analyze the system. As I mentioned, we will get much information from this according to how we have set up the experiment. We have discussed the process. Always remember process details have no role in your result.

It is how you set up the experiment and you get the result out of that. I will talk about that in detail in the next class. Thank you very much for listening. I have taken this relatively longer period because I wanted to finish the MD simulation process. But you guys are amazing. Thank you very much for listening. Thank you.