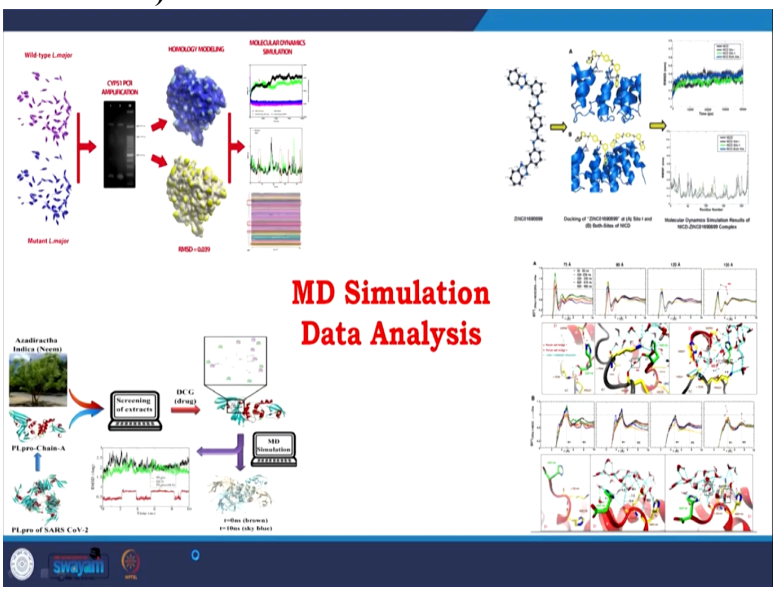
Structural Biology
Prof. Saugata Hazra
Department of Biotechnology
Indian Institute of Technology, Roorkee

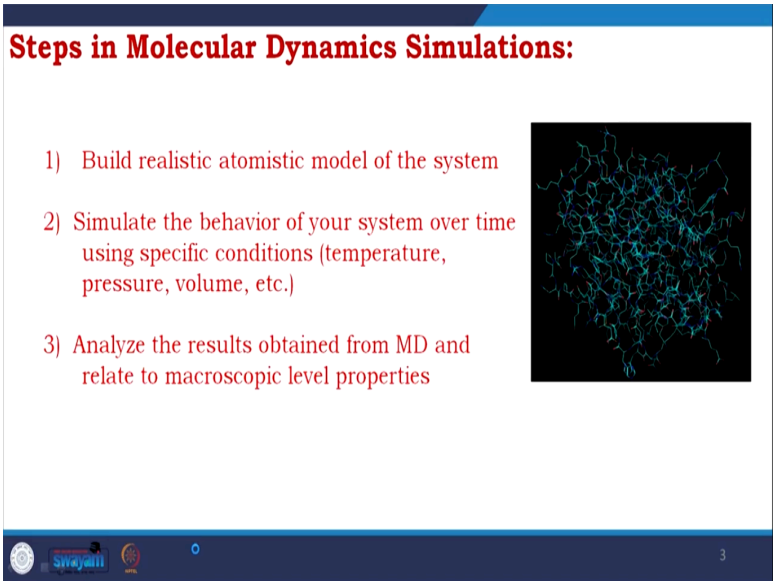
Lecture-49
Molecular Dynamic Simulation Process Part III

Hi everyone, again, welcome to the course of structural biology. Today, we are already in the model of MD simulation, and today we will talk about data analysis. So MD Simulation Data Analysis.

(Refer to Slide Time: 00:36)

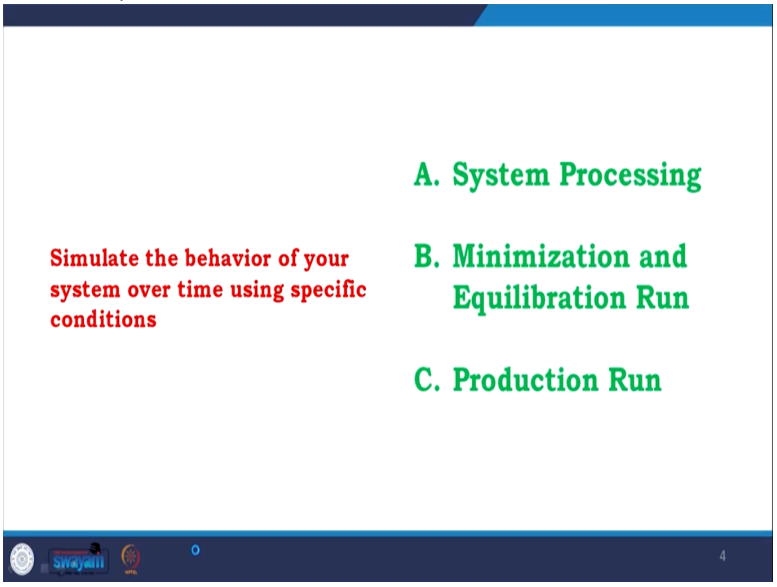


(Refer Slide Time: 00:42)



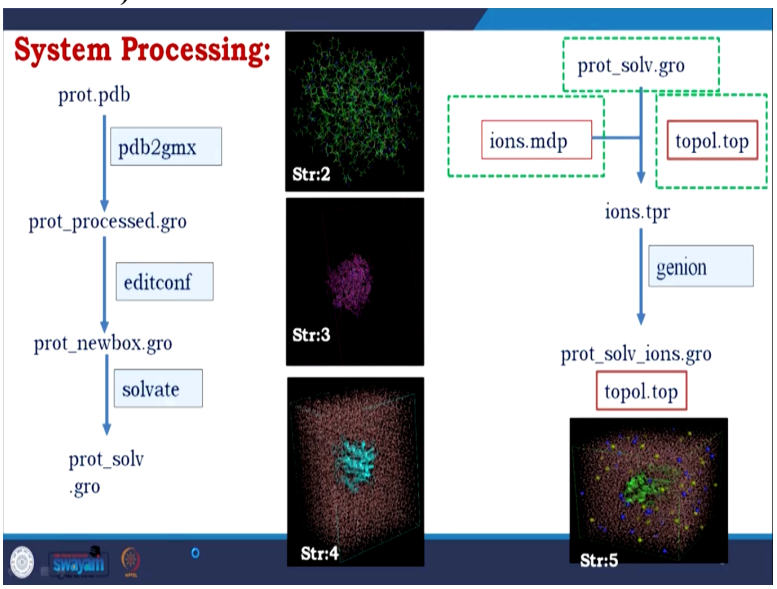
I have discussed the steps in molecular dynamics simulation multiple times. We have built a realistic atomistic model of the system, picked up the PDB, cleaned the PDB, and repaired if there was a necessary repair, and what I get is what is in front of you.

(Refer Slide Time: 01:06)



The next step was to simulate the simulation process. We divide it into 3 steps system processing, minimization and equilibration run, and final production run.

(Refer Slide Time: 01:26)



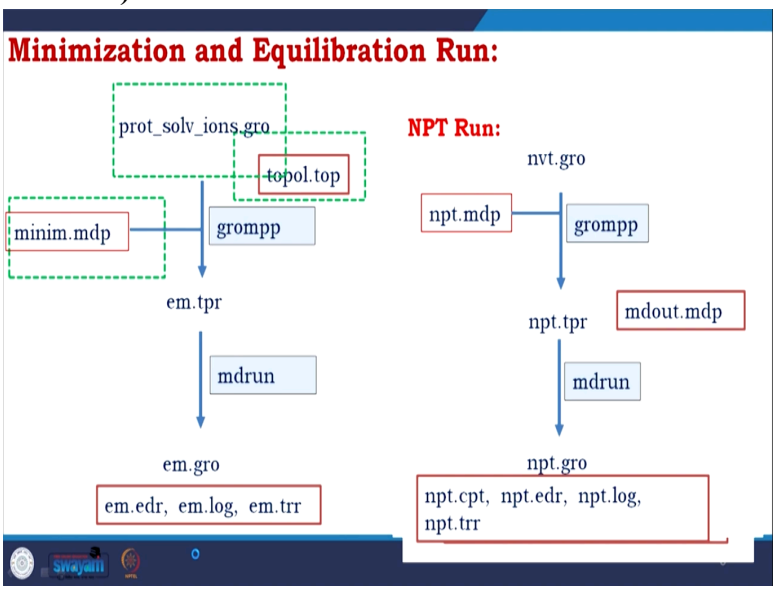
Let us take a quick look. So, in the system processing, we start with the clean pdb we got from the rcsb experimental or model and then work on it. In this one, we apply pdb to gmx, which is the conversion from a coordinate to the gromacs competing version with other software like Ember NMD charm. They have their format like that, and we got a dot gro file which is a gromacs competent file.

Then we apply editconf, which gives us a new box. So, you will see that you have the structure and where to add the hydrogens, then which at the new box, and finally, you solvated it. You see, the box is now solvated. Then, we go to the next step, where we have the

protein underscore salt dot grow file and the main file at topl top associated file and add ions dot mdp.

So, ions dot mdp dot gro dot top together they develop an ions dot tpr file, and then by genion, they put the ions which you see here. So, starting from a structure, we come somewhere where we could perform the real simulation.

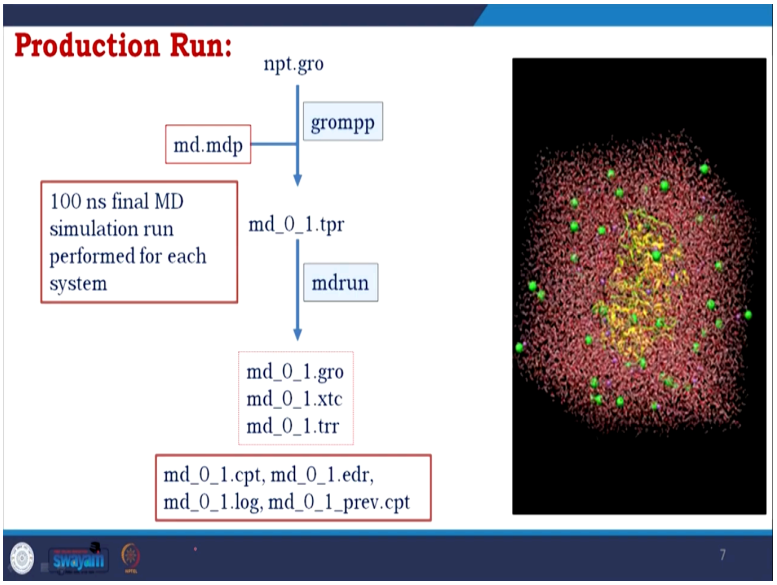
(Refer Slide Time: 03:10)



So the first step is minimization, where we take that dot gro file again minim dot mdp topol dot top file is coming we apply grompp which we apply in the last step either we got em dot tpr, and then we do the first MD run we got energy minimization dot gro em dot to grow as a main file as an associated file you got em dot edr which is energy file m dot log which is a log file having all the information and em dot trr which is a trajectory file.

With this em dot gro file, we do equilibration, an ensemble run with nvt condition. We got in nvt dot gro with the same procedure as the main file in nvt dot cpt, which is the checkpoint file in nvt dot edr which is the energy file, nvt.log is the log file, and nvt dot trr, which is the trajectory file same type of equilibration is again performed. Still, this time instead of nvt, we performed the npts ensemble condition we got in npt dot gro. With that, we got npt dot cpt, a checkpoint file in npt dot edr, an energy file. Npt dot log is a log file, and npt dot trr is a trajectory file.

(Refer Slide Time: 04:41)

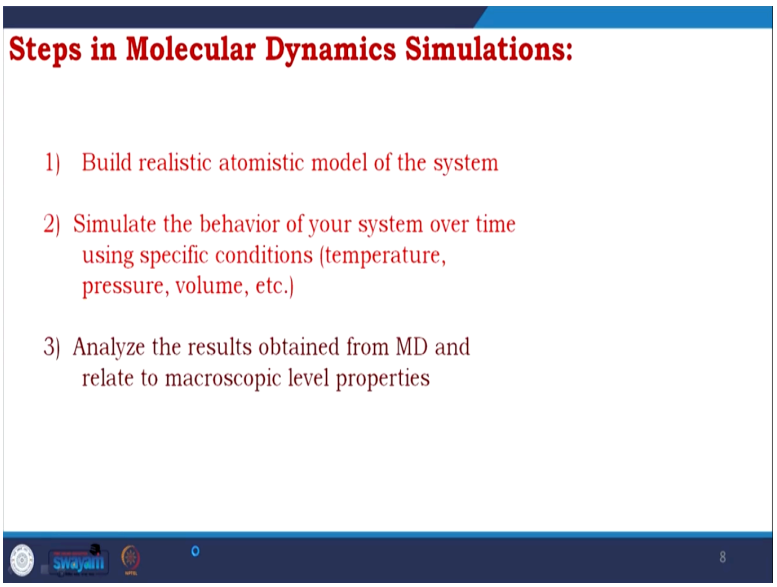


With minimization and equilibration run performed, now we come both of them. I did not talk about where 100 picoseconds and now we come to the final round with the file npt dot gro. Again, we converted the tpr file by the common grompp having dot mdp dot gro and dot top file, and the tpr file was performed MD run with 100 nanoseconds final MD simulation run performed for each system.

What we got we got md_0_1 dot gro, which is a gromacs competent file, md_0_1 dot xtc, and md_0_1 dot trr, so trr, I have already talked about which is a trajectory file, but because you know, after 100 nanosecond up run there would be so many or so high volumes of a huge volume of the file we make a zip version which is that dot xtc. I already talked about dot cpt, the checkpoint file. The checkpoint files are used if your MD simulation run is top for any reason, then you could start again from the checkpoint file without losing information.

Interestingly, because it is a 100 nanosecond run, we have done a double check here md_0_1 _ previous dot checkpoint another checkpoint file is there dot edr is the energy file is there and dot log and after the production run we get so many snapshots which give us a movie to look at this beautiful movie this is the result of the MD simulation run.

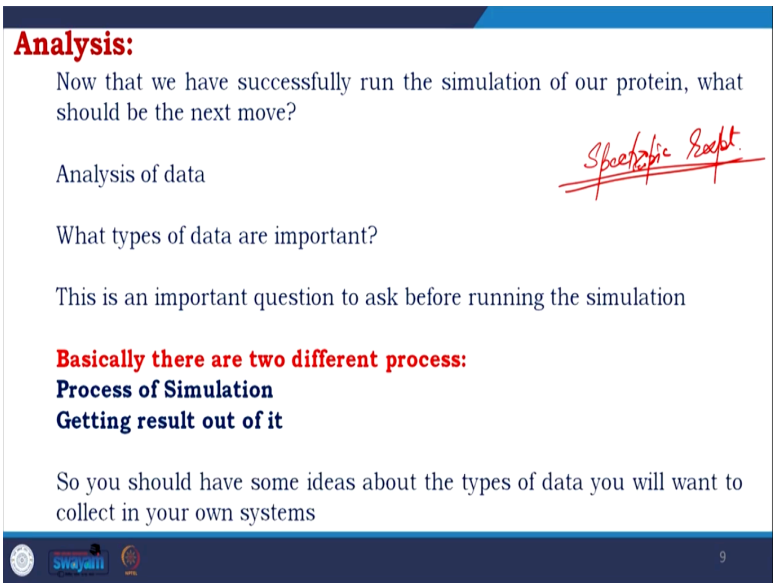
(Refer Slide Time: 06:41)



So again, coming to this slide, the steps in molecular dynamic simulation. The building of a realistic atomistic model of the system is done through simulation with 3 steps preparation, minimization, and equilibration, and production done is done. So, now we have to go for what we are here today, analyze the results obtained from MD, and relate them to macroscopic level properties. Remember, I talked about MD as the modeling and in this MD simulation system.

What we are trying is that inter converts the understanding of a model system and compare it with the experimental system. And then, from the compression, we put more input to the model system, and cyclically, it would be becoming relatively more realistic with more or multiple times you run the cycle, you get the result you compare the result you learn from it, and you update your MD simulation system.

(Refer Slide Time: 07:51)

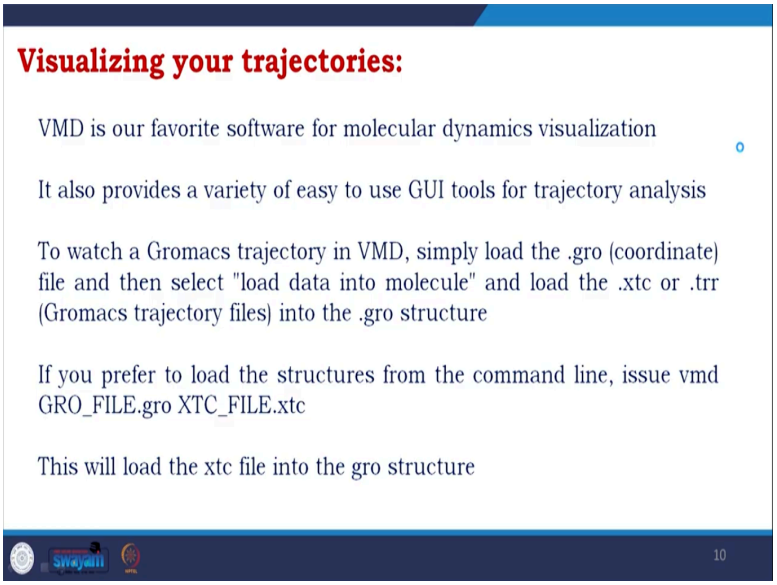


So, going to that data analysis, we have successfully simulated our protein. What should be the next move? The next move is analyzing data and determining what data types are important. Now, here I want to give a stop because this is a critical juncture that you should consider when discussing a procedure like the MD simulation process. This is unrelated to what you want to get.

What I mean is when you experiment, like if you perform a spectroscopic experiment, we need a spectroscope. It might be UV, and it might be fluorescence. It might be circular dichroism. Which one you need depends on what questions you are asking similarly, and a lot of people miss understand this tip. People think you perform MD simulation, and you start getting results.

The results are not dependent on the process of the simulation. The results are dependent on the experimental design you have done. So, remember that. So, I also wrote it. There are 2 different processes. One is the simulation process, and the second is getting results. It would help if you had some ideas about the data you want to collect in your system. The more you understand more you design the experiment in such a way better the outcome of the experiment.

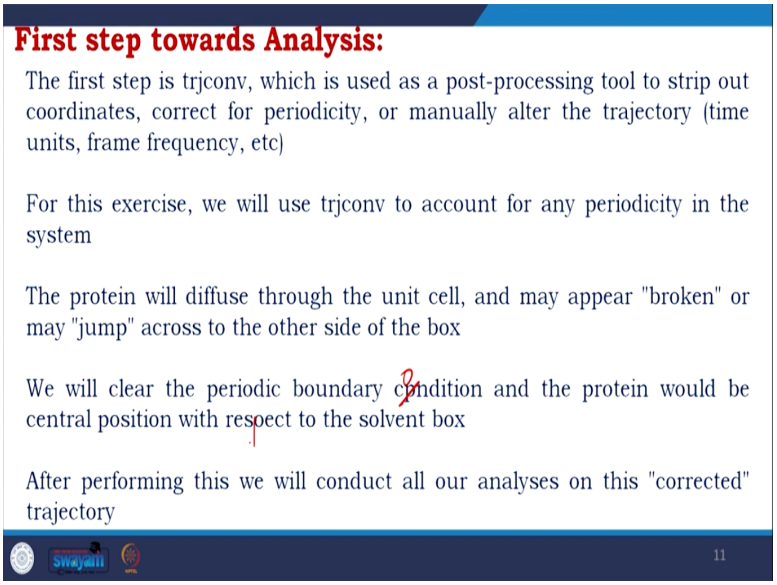
(Refer Slide Time: 09:55)



The first step is visualizing trajectories. As I told you, to have a force field, you put a force on your sample, and it would go up and down, so you will create a trajectory. VMD is our favorite software for molecular dynamics visualization. If you remember, I talked about visualization software. VMD is a visualization software used for looking at proteins and many other purposes, but it was developed for molecular dynamics.

That is why the name VMD came from visualizing molecular dynamics. It provides a variety of easy-to-use GUI tools for trajectory analysis. To watch a gromacs trajectory in VMD. First, you load the dot gro file and then select load data into molecule and load that dot xtc or dot trr. I will prefer dot xtc cause of the reduced size of the file. And when you load them into the dot gro structure, you prefer to load the structure from the common line. You issue VMD common gro underscore file under score gro xtc dot xtc, and this only loads the xtc file into the gro structure automatically. You will see the outcome of the trajectory.

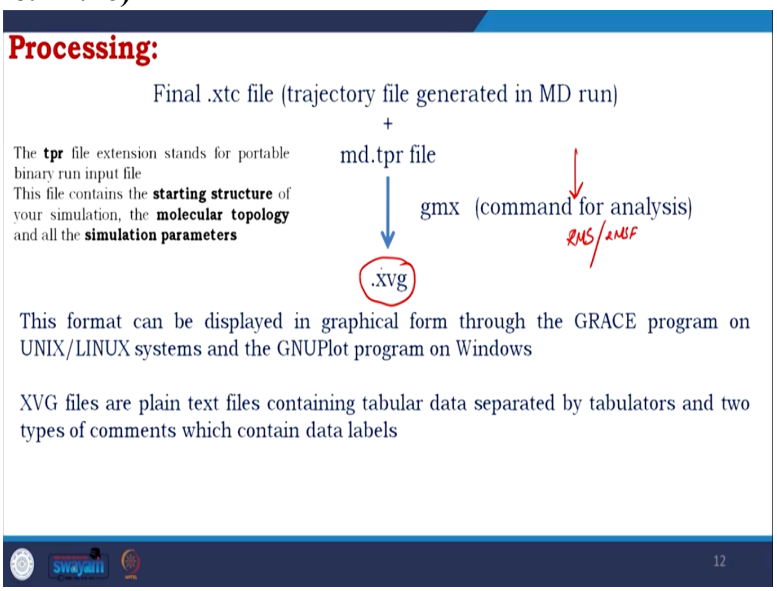
(Refer Slide Time: 11:27)



But before you go for visualizing and analyzing, there is one step you should always perform. The step is called trjconv, trjconv, which is used as a post-processing tool to strip out the coordinates all the coordinates correct for periodicity or manually alter the trajectory time units, frame frequency, etc. We will use trjconv to account for any periodicity in the system.

At the time of the run, the protein might diffuse to the unit cell or appear broken or jump across to the other side of the box. All those things could be corrected, we will clear the periodic boundary condition, and the protein would be central position concerning the solvent box. After performing these, we will get the corrected trajectory, and after all the analysis I am talking about next, we will use this corrected trajectory. Because, here, the mistakes happen all corrected.

(Refer Slide Time: 12:46)

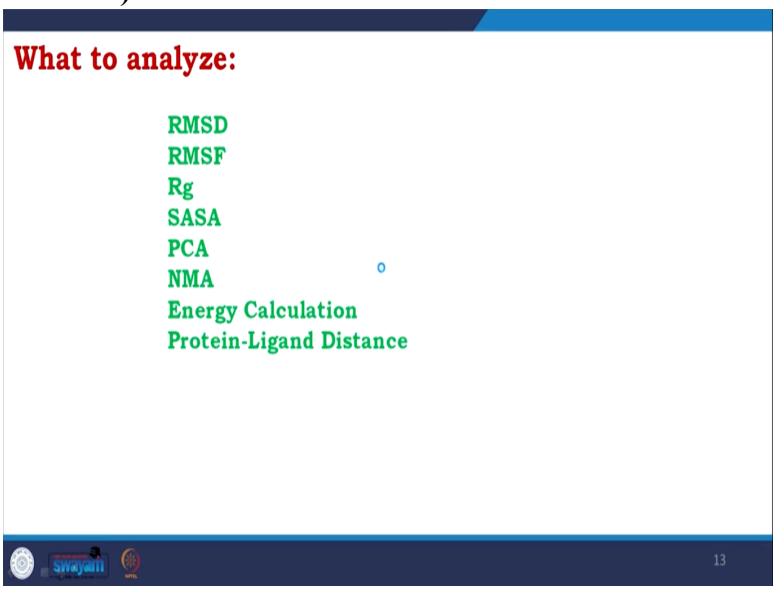


So how do you process you? As I told you to get the final dot xtc, the trajectory file, and you have the md dot tpr file. I have talked about this earlier. But the tpr file extension stands for the portable binary run input file. This file contains the starting structure of your simulation, the molecule topology, and the simulation parameters. Now you will give command gmx common will give you dot xvg depending on what you are putting here.

You could put RMS, and you could put rmsF like that. Now you get a dot xvg file. So from your dot xtc and dot tpr file, you get the dot xvg file. This dot xvg file format could be displayed in graphical form through the GRACE program. GRACE is a popular UNIX program that helps you to get the graphical representation, whereas if you are not in UNIX or LINUX system, you are using Windows, then you use the GNUPlot program.

XVG files are plain text files containing tabular data separated by tabulators and 2 types of comments which contain the data labels. Now, because it is a plain text file, you could understand that you could manually edit the file directly. But considering when you run the entire simulation, the file would be huge. It would not be a realistic idea to edit it manually. You should take the help of the system automation programming modules. So, I told that you have to design your experiment.

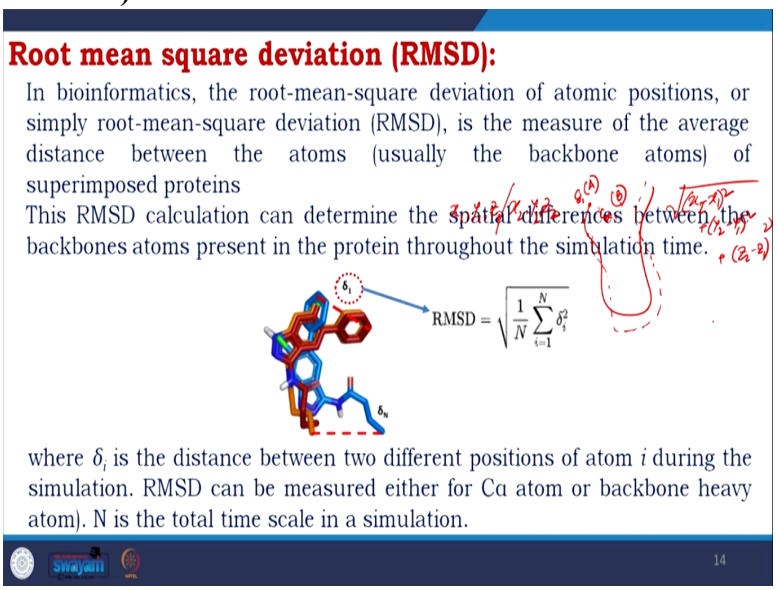
(Refer Slide Time: 15:03)



But there are some analysis modules people generally follow. Yes, there are. So, root mean square deviation, where I will talk about details, root mean square fluctuation radius of gyration solvent accessible surface area principal component analysis, normal mode analysis energy calculation protein-ligand distance I wrote it protein-ligand distance but it could be a

protein with DNA protein is protein any differences you are looking up at the simulation run. You could get information from your simulation run from intramolecular distances and dihedral angles. We will discuss a few in detail for your learning.

(Refer Slide Time: 16:07)



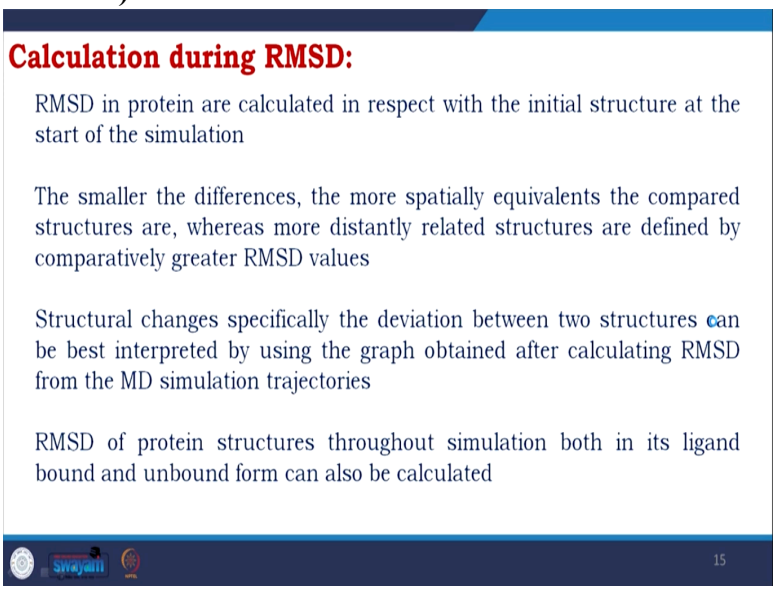
In Bioinformatics, the root-mean-square deviation of atomic positions or simply root mean square deviation is the measure of the average distance between the atoms up superimposed proteins. It means that suppose you have a protein, you have the protein. This is a protein, and the protein now moves here. How could you measure this? What do you do?

You take this carbon C alpha. So, let us say this is C alpha 1 from the A position and C alpha 1 from the B position. Each of them has their coordinates. So, x 1 y 1 z 1 x 2 y 2 z 2, the distance, if you remember, would be a root of x 2 - x 1 square + y 2 - y 1 square + z 2 - z 1 square RMSD is the average of all those distances the all the C alpha present in the entire protein. So, let us see how it goes.

But RMSD calculation can determine the spatial differences between backbone atoms in the protein throughout the simulation time. If you look at this as the compression and you get the delta 1 RMSD is root over 1 by n summation i = 1 to n delta i square where delta is the distance between 2 different positions of atom i during the simulation RMSD can be measured either for c alpha atom or heavy backbone atom. N is the total time scale in a simulation.

So, looking at RMSD is very critical because, when you run the simulation, you want to see the displacement, you want to see the movement, and the best and easy way of determining this is to get the root mean squared deviation value.

(Refer Slide Time: 19:11)



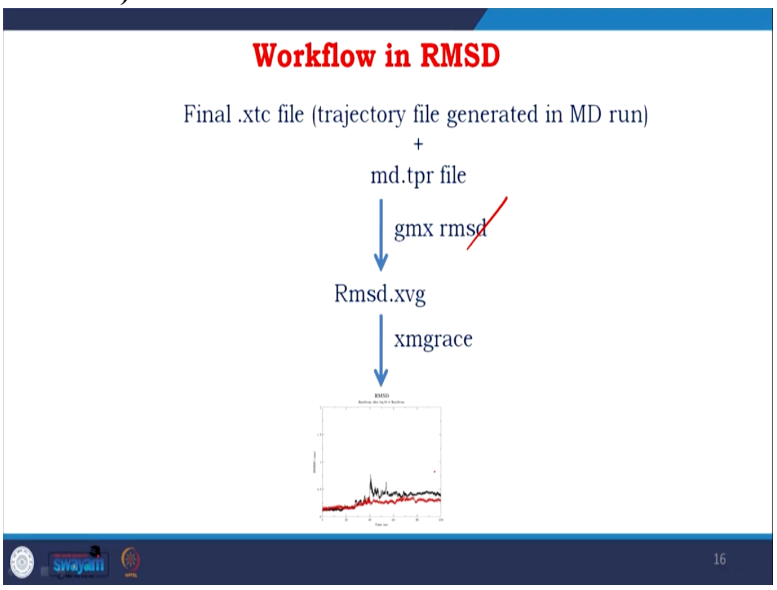
RMSD in protein is calculated concerning that initial structure at the start of the simulation. So, remember, when we start, we have a starting structure starting PDB. Now, you get the snapshots, and you compare the deviation. The smaller the differences, the more spatially equivalents the compared structures are, whereas comparatively greater RMSD values define more distantly related structures.

When comparing 2 different structures, in the case of simulation, you want to see that you have an initial structure that is debated throughout the simulation. Let us say you put heat, so what is the effect of the heat? Let us say you increase the pH. What is the effect of the pH on the deformity if the structure is affected? If the non-covalent bonds are broken all these things as a primary measure, you could use RMSD.

Structural changes, specifically the deviation between 2 structures, can best be interpreted using the graph obtained after calculating RMSD from the MD simulation trajectories. What type of graph? Remember, I told you that when you have the simulation trajectory, you apply the gromacs command and get the dot xvg file. Now in the dot xvg file, when you do this program or GED plot, you will get the graph you will see next.

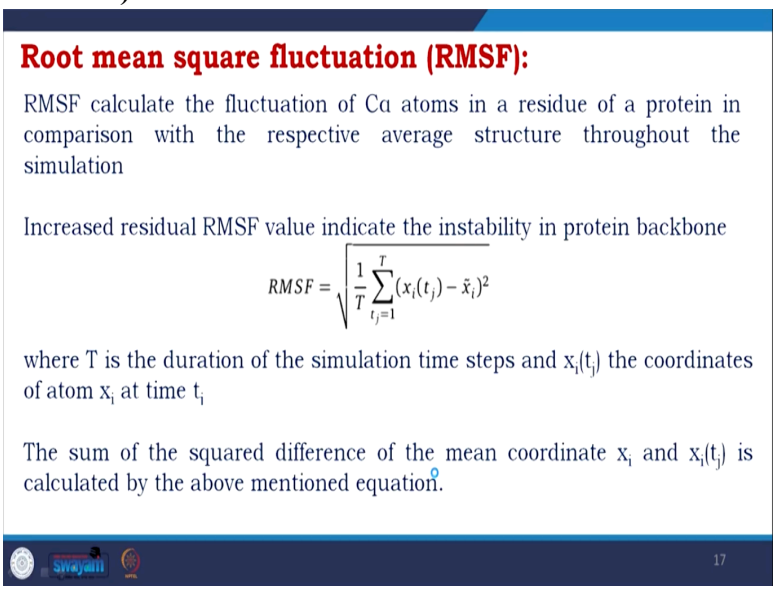
RMSD of protein structures throughout the simulation, both in its ligand-bound and unbound form, could also be calculated like you have an eco structure and a ligand-bound structure is the ligand-bound structure makes it stabilized the eco structure is more flexible. All this information you could get primarily through RMSD.

(Refer Slide Time: 21:18)



How do you work on RMSD? As I told you to, take the dot xtc file, you have the dot tpr file, and you apply gromacs RMSD or RMS, and you get generally used 2 types RMSD I think more popularly these at gmx dot rmsd you get Rmsd dot xvg you apply xmgrace, and you get the graph where you are putting rmsd with time and see the changes between the initial structure and the different other structure.

(Refer Slide Time: 22:01)

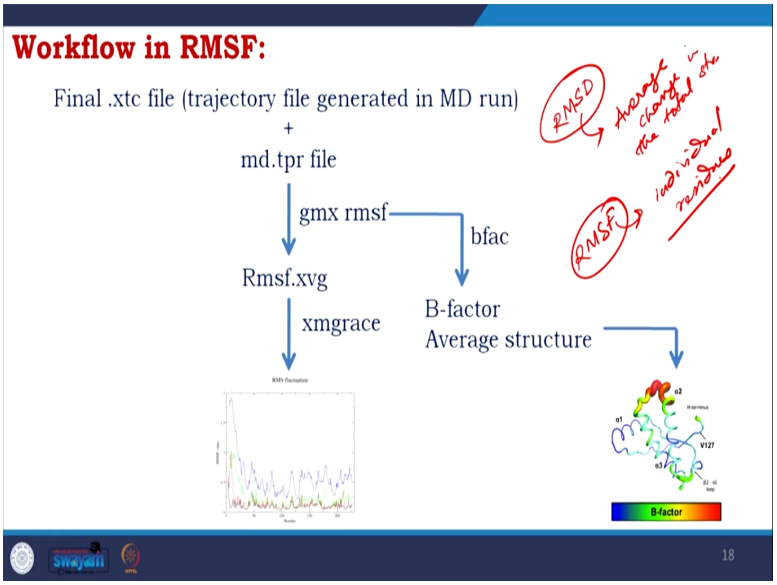


Coming to RMSF, which is root mean square fluctuation, RMSF calculates the fluctuation of C alpha atoms in a residue of a protein in comparison with the respective average structure

throughout the simulation. Increased residual RMSF value indicates instability in the protein backbone. So, if you see a high value of RMSF, the protein is either unstable or flexible.

RMSF is calculated as the square root of 1 / T summation t j = 1 to T x i t j - x i whole square where t is the duration of the simulation time steps and x i t j the coordinates of atom x i at time t the sum of the squared difference of the mean coordinate x i and x i t j is calculated by the equation as mentioned above which we could see.

(Refer Slide Time: 23:02)



So, what is the workflow similar to here? You have the dot xtc file again. In that dot tpr file, you apply gm x rmsf, you get Rmsf dot xvg, then apply xmgrace, and you see that graph again Rmsf concerning the residue number. So, if you see, we will see that individual residues here and you will see their fluctuation in different structures, some people might wonder, what is the difference is RMSD is also talking about deviation RMSF is also talking about fluctuation.

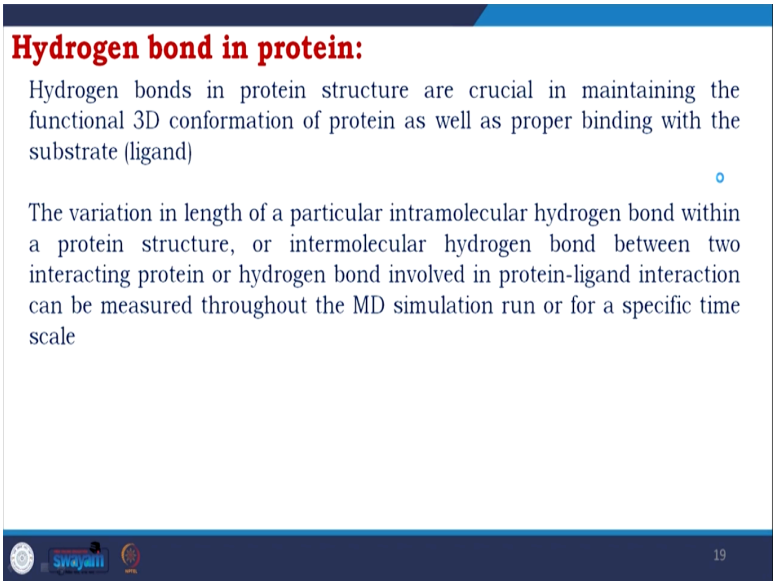
Which is the deviation. So, what is the huge difference? When going for RMSD, you are looking for an average change in the total structure. But when you are performing RMSF, looking at individual residues, RMSF information is extremely critical for mechanistic derivation, which I am talking about. I am talking about if you are looking at the behavior of a protein in its epo form and substrate-bound form.

By analyzing the particular residue's RMSF value. You could differ yet how it behaves that it is an epo orm, and how differently it begins with binding to different ligands. So, you get a

mechanistic view of how they will ligand bind, how much stabilization it could provide to the structure, what type of interaction is happening, and all those interestingly, if you remember, in the PDB file.

We have something called the B factor. You could also get the B factor of the average structure from the simulation and compare it from there with that experimentally available structure. So, this is also something that allows you to compare.

(Refer Slide Time: 25:47)



We are coming to hydrogen bonds in a protein. We have talked multiple times, and I do not think people learning here at this point understand how important non-covalent bonds are, especially hydrogen bonds, which are like energetically less in contribution, but more in number. So, together the effect is huge. So, in simulation, we want to see the breaking and making of hydrogen bonds.

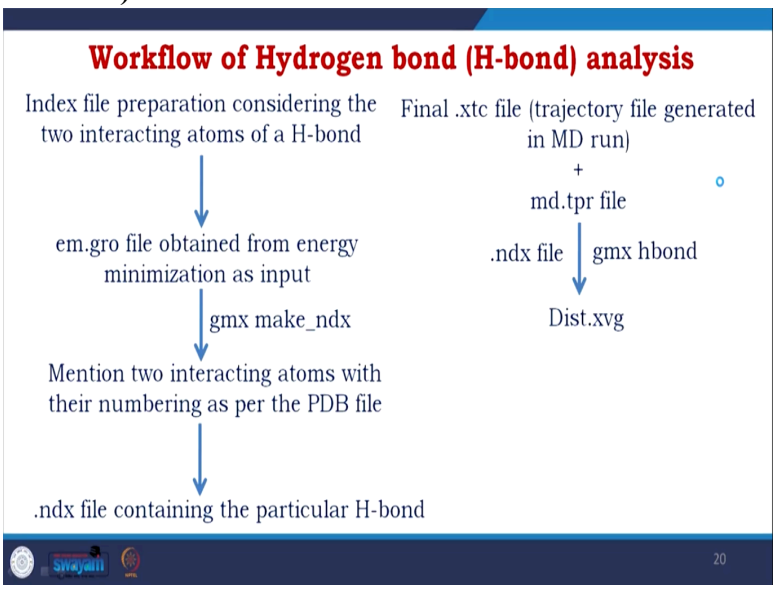
To understand how the protein switch is distorted, not changing conformation like this. So, I do not need to say that, but hydrogen bonding protein structures are crucial in maintaining the functional 3D conformation of the protein and proper binding with the substrate. The variation in the length of a particular intramolecular hydrogen bond within a protein structure or intermolecular hydrogen bond.

Between 2 interacting proteins or hydrogen bonds involving proteins, like an interaction, could be measured throughout the MD simulation run or for a specific timescale. Now, it is very important to understand, you know, that is what a dynamic is giving you. If I say I hold

this pen, it will not give you the specific information. This pen is hold tightly or loosely by me.

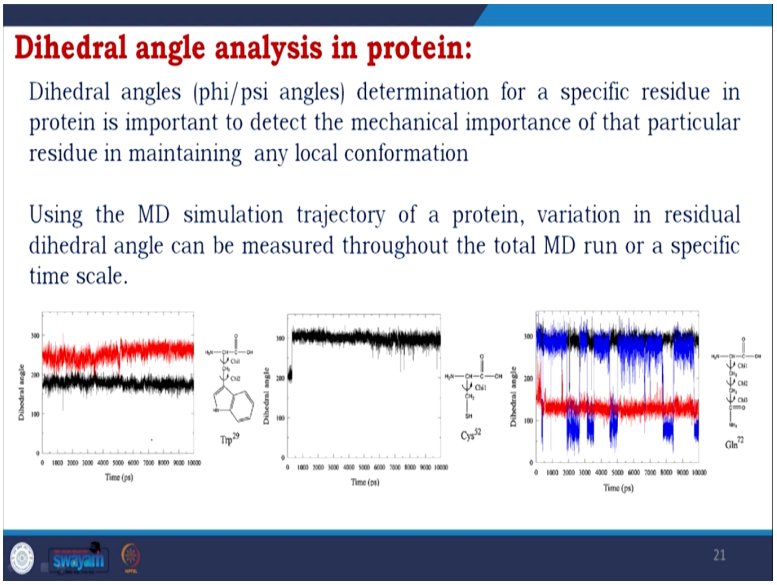
But if you put the whole system under simulation so that a force could be given and to calculate the hydrogen bond distances throughout the run, if you find that the distance of hydrogen bond always kept a good hydrogen bonding allowed distance or not, depending on that, you could comment a substrate or inhibitor is bound to a protein strongly or tightly this information you could not get from the static structures. So, this is a representation again, whereby the hydrogen bond length and time scale are given.

(Refer Slide Time: 28:23)



Workflow, index file preparation considering the 2 interacting atoms of a hydrogen bond here you have to talk about which bond you want to use em dot gro file obtained from energy minimization as input, you have to put the common gmx make underscore ndx to index file mention to interacting atoms with their numbering as per of the PDB file and dot ndx file containing the particular hydrogen bond. So, the final dot xdc file is the trajectory file, and the md do tpr tpl file gmx hydrogen bond, and with the dot ndx file generated here, you get that Dist dot xvg you apply Xmgrace you get it to the 2D plot.

(Refer Slide Time: 29:13)

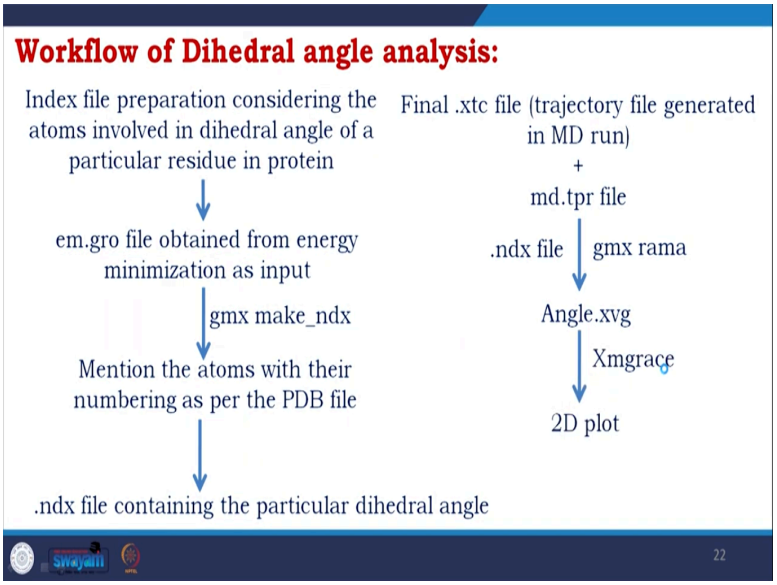


Coming to the dihedral angle analysis in protein, again, students who are following this course, it is not surprising to you that whenever we are looking for protein movement, it is about the dihedral angles because initially, like mainly the backbone, the phi psi then the chis, they are determining when they are changing, you could take the changing spectrum and comment on the dynamic nature of the protein.

So dihedral angles phi psi angle determination for a specific residue in a protein is important to detect the mechanical importance of that particular residue in maintaining any local conformation. Using the MD simulation trajectory of a protein variation in residual dihedral angle can be measured throughout the total MD run or a specific time scale. You see here, and I talked about, you see here, 2 chi, so, in the crypto fun.

So, you get this representation of the dihedral angle with the time change in picoseconds. When you consider 16, there is 1 chi; when you consider good, there are 3 chis, and how they fluctuate. Depending on that, you could have seen that change in the dynamic nature of the protein.

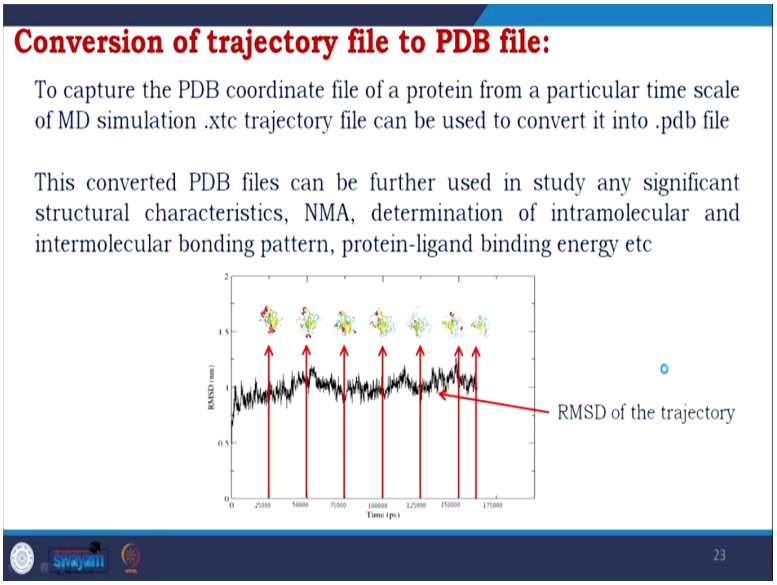
(Refer Slide Time: 30:48)



The workflow here also needs to prepare the index file. So index file preparation considers the atoms involved in the dihedral angle of a particular residue in the protein because you have to mention the particular residues for a dihedral hydrogen bond. That is why you need to do the index filing. em dot gro file obtained from energy minimization, and you will use it as input gmx make underscore ndx.

And the atoms with their numbering as per the PDB file, you get the dot ndx file containing the particular dihedral angle. Now it is the same thing, you have the final dot xtc file, you get the dot tpr file, you apply the gmx gmx rama, and you add the dot ndx file containing the information about the dihedral angle, you get angle dot xvg you apply xmgrace you get the 2D plot. So you need we are talking about trajectory, but you need the PDB files. How do you convert the trajectory file, the dot xtc, or the dot trr file to coordinate file PDB file?

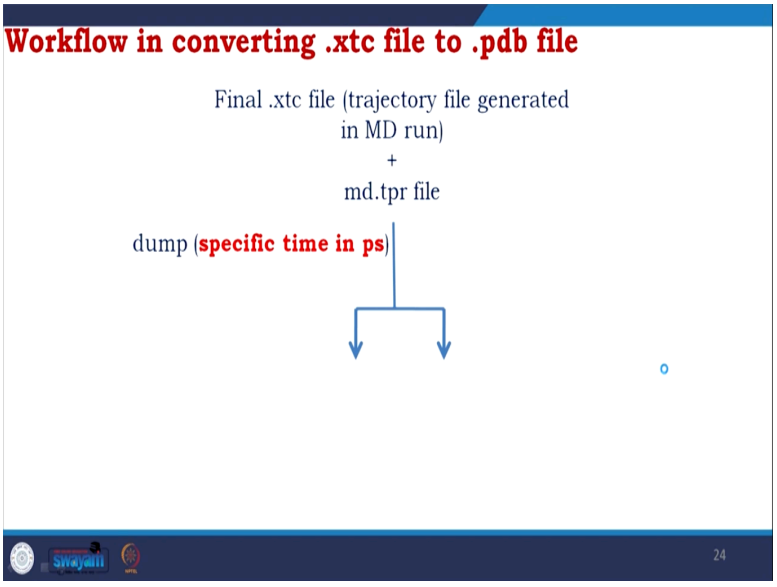
(Refer Slide Time: 32:06)



To capture the pdb coordinate file of a protein for a particular time scale of MD simulation dot xtc trajectory file can be used to convert it into a dot pdb file. These converted pdb files can be further used in the study of any significant structural characteristics like normal mode analysis, determination of intramolecular and intermolecular bonding patterns, protein-ligand binding energy, etc.

So, if you see, this is the RMSD versus time plot. Now, if you select these trajectories, get different time points of the trajectory, and get the snapshots. So, as I told this, the RMSD of the trajectory and you identify the snapshots from the pdb file.

(Refer Slide Time: 33:06)

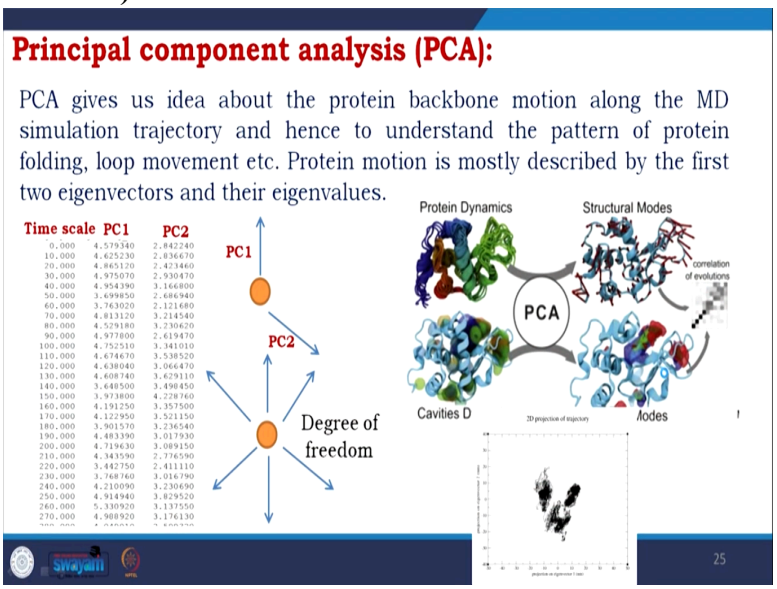


What you will do, you get the Final dot xtc file and dot tpr file, then you do dump specific time in a picosecond, which is chosen here in this file, and you apply gmx trjconv for total system or gmx make dot ndx for selection when you want to select you could have the whole

system you could have a specific point as I told in case of hydrogen bond in case of dihedral if you want everything you go with the dot xtc.

Suppose you want particular you up to make an index. Similarly, here for the total system for a particular one, you will convert it to the dot pdb file, which is the coordinate file. From there, you could get binding energy calculations. You could do a normal mode analysis. You could go to the Pymol PMD Chi Mira code, and you could do the visualization in case of binding energy calculation. You could also take the energy file dot edr file, but now the advanced programs are already picking up the information.

(Refer Slide Time: 34:37)



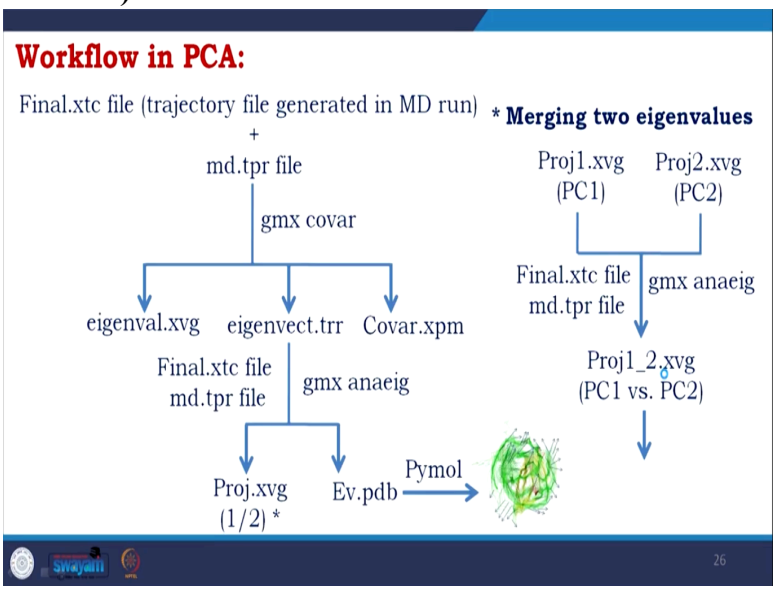
I am coming to principal component analysis or PCA. PCA gives us an idea about the protein backbone motion along that MD simulation trajectory and hence to understand the pattern of protein folding look movement etc. Protein motion is mostly described by the first 2 Eigen vectors and Eigen values. If you compare RMSD and PCA, they are calculating the change.

But there is a huge difference because, in RMSD, there is only value here in principal component analysis. You have the Eigenvector, so there is a vector, there is value, so there is a directionality. What is the principal component in linear algebra principal component is used where you see the comparison between 2, and you make or choose a principal component where you see the maximum deviation.

And then, you choose PC2, the second principal component, which is unrelated to principal component one, and then calculate the covariance. So, as I told PC1 and PC2, you get the

degrees of freedom and that 2D projection showing how the protein has deviated. So far, starting protein dynamics to find structural modes to find cavities where you get the binding modes, correlation of evolution PCA is used.

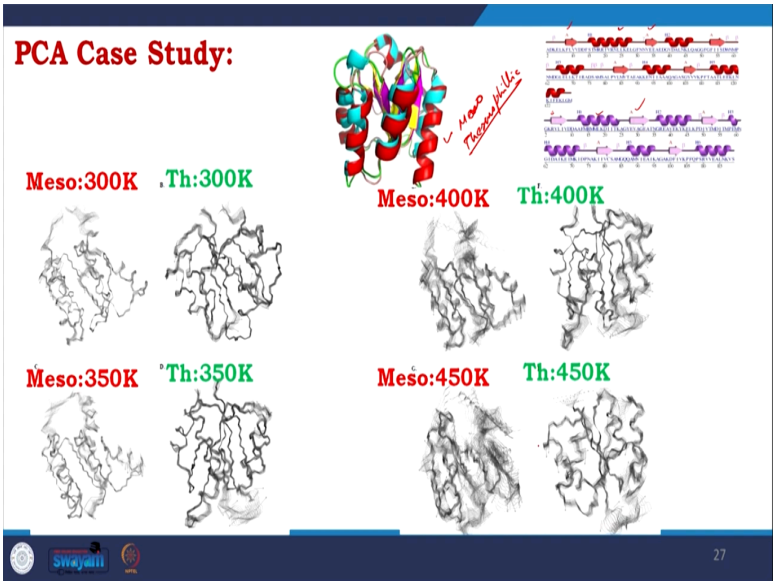
(Refer Slide Time: 37:01)



How does PCA workflow happen again? Having the Final dot xtc and md dot tpr files would be best. You perform gmx cover you get eigenvalue dot xvg eigenvector dot trr and Covar dot xpm. Now, use the Final dot xtc md dot tpr file with gmx, anaeig you get Proj dot xvg and Ev dot pdb. This is coming from the first principal component or the second one. It would be best if you had another one to show in Pymol. You see that changes, but then you merge the 2 Eigenvalues.

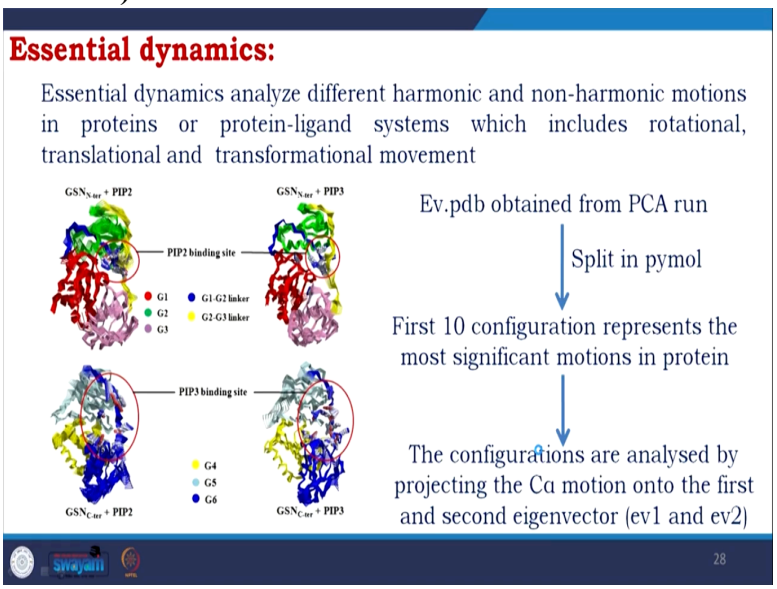
You get up towards Proj1 dot xvg from PC1, you get Proj2 dot xvg from PC2, and then with the help of Final dot xtc and md dot tpr file again by applying gmx anaeig you get the Proj underscore 2 dot xvg with xmgrace now, you get the final graphical representation.

(Refer Slide Time: 38:26)



We see a study these are 2 proteins mesophilic protein and thermophilic protein if you look at them, they merge quite well, if you look at the secondary structures, the secondary structures are quite similar. So, this is difficult to understand why sub one is thermostable and the other is not. But, when you measure through analysis a principal component you see how 300 350 400, and 450 the changes are showing more in mesophilic showing that this has more movement in comparison to the thermo stable protein.

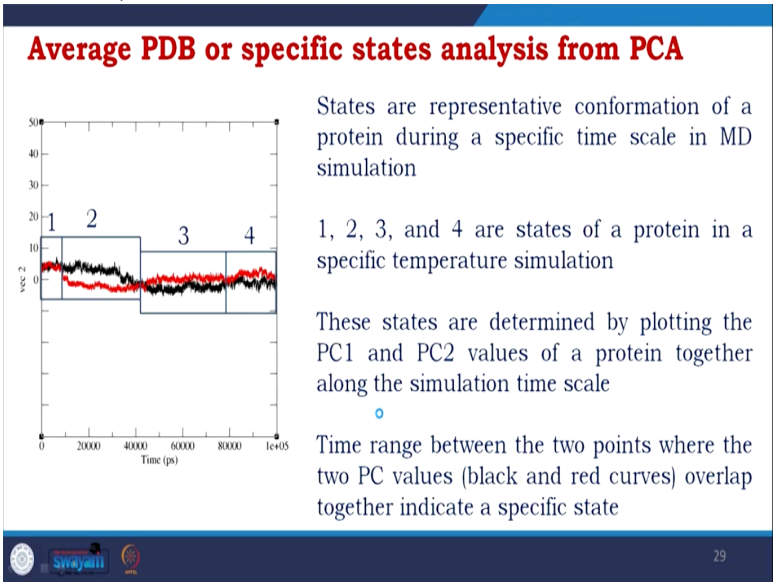
(Refer Slide Time: 39:34)



Coming to essential dynamics, essential dynamics analyze different harmonic and non-harmonic motions in proteins or protein-ligand systems which include rotational, translational, and transformational movements. As you see here, it is showing the binding of protein with ligand and you see the relative movements here, if you look at the red circle portion, you see the movement.

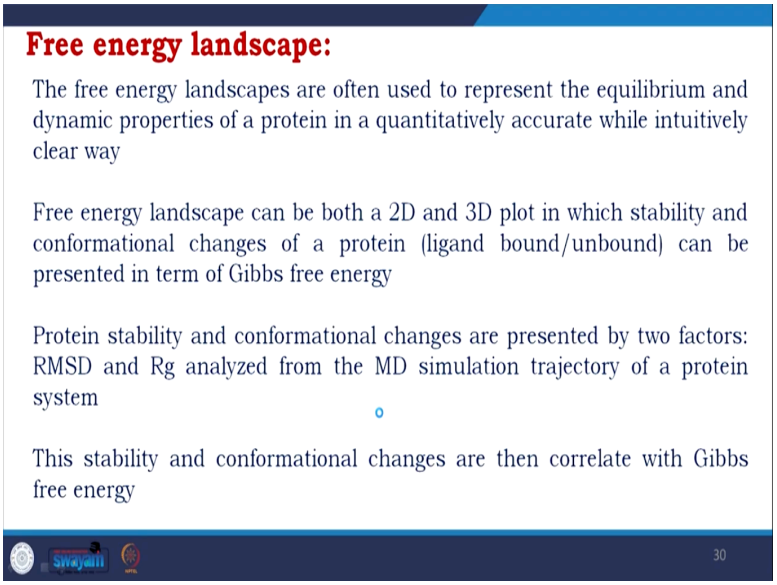
So, Ev dot pdb obtained from PCA run you split those files in Pymol then you take the first 10 configurations which represent the most significant motion in protein and then the configurations are analyzed by projecting the C alpha motion into the first and second Eigenvectors which are represented by ev1 dot pdb and ev2 dot pdb. If you look here, you see the movements you see that changes.

(Refer Slide Time: 40:53)



So, the average PDB or specific steps analysis from PCA is also possible the states are, the representative conformation of a protein during a specific time scale in MD simulation. Here you see 1 2 3 4 the steps of a protein in a specific temperature simulation, you see that the vector 2 and time are plotted here. These states are determined by plotting the PC1 principal component 1 and principal component 2 values of a protein together along with the simulation time scale. The time range between the 2 points where the 2 principal component values overlap together indicates the presence of a specific state.

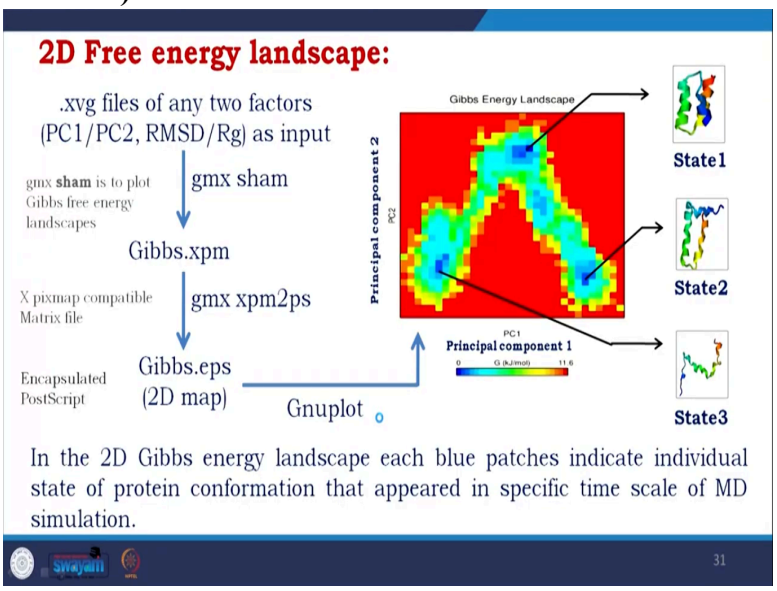
(Refer Slide Time: 41:42)



Coming to free energy landscape free energy landscapes are often used to represent the equilibrium and dynamic properties of a protein in a quantitatively accurate and intuitively clear way. Free energy landscape can be both 2D or 3D plots in which stability and conformational changes of a protein can be presented in terms of Gibbs free energy. So free energy landscape is a plot of the principal component RMSD RMSA or any changes along with the Gibbs free energy.

Protein stability and conformational changes are presented by 2 factors, RMSD and Rg analyzed from the MD simulation trajectory of a protein system. This stability and conformational changes are then correlated with the Gibbs free energy.

(Refer Slide Time: 42:34)



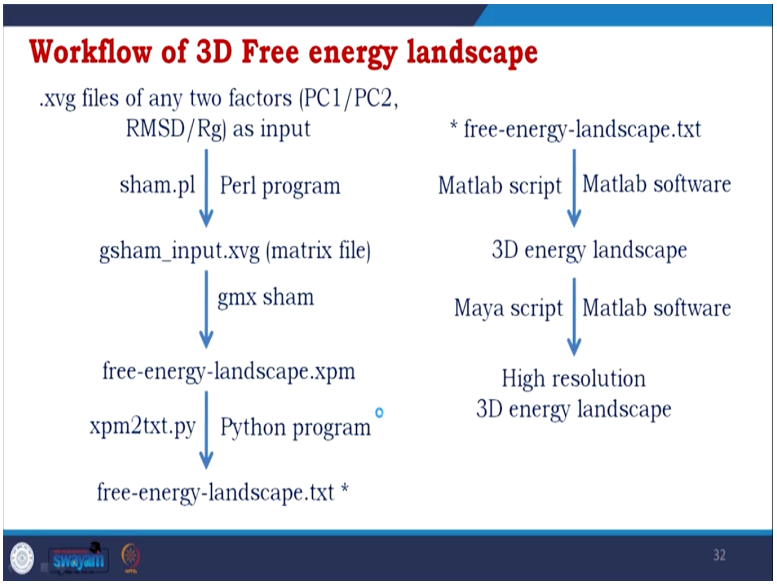
So if you see that 2D free energy landscape, you use the dot xvg file to have any 2 factors. As I told PC1, PC2, RMSD, and Rg as input then you run gmx sham and you get Gibbs dot xpm

gmx sham to plot Gibbs free energy landscapes. So, get it xpm, then you apply gmx xpm dot ps you got Gibbs dot eps, which is giving you a 2D map, you put it through gnuplot, and you get the plot here.

Where you have PC1 and PC2 with respect to Gibbs energy. So, from the energy here, you could have identified different states, which is wonderful for further analysis. If you look at where you see the blue color, you know, that stabilizes the state of a protein belonging there and you isolate from there, the PDB file and represent states. In the 2D Gibbs energy landscape, each blue patches indicate the individual state of protein conformation that appeared in a specific time scale-up MD simulation.

So by identifying them, you could talk about the process where you correlate the movement of the protein with the functional implication of what the protein or the state of that protein might have a handle. Coming to 2 other things, the xpm is x fix map compatible matrix file, and eps is encapsulated postscript file.

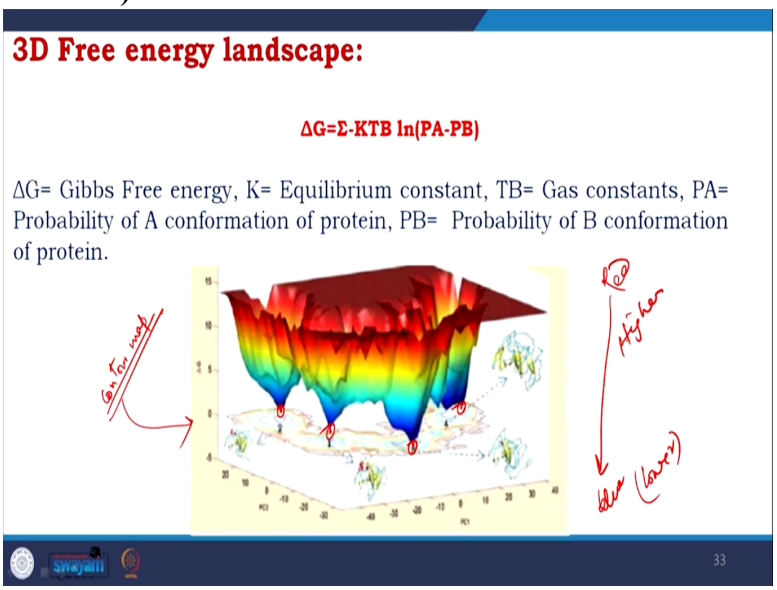
(Refer Slide Time: 44:37)



Coming to 3D free landscape dot xvg file of any 2 factors again, PC1 PC2 RMSD Rg, as we talked about, are used as input sham gsham underscore input dot xvg it is a matrix file, again, so you get the matrix file instead of one file, you apply gms sham, you got free energy landscape dot xpm, you apply xpm2xt dot py, the Python program, you get free energy landscape dot txt, that free energy landscape dot txt file.

Will be going through the Matlab script in the Matlab software, and you get a 3D energy landscape. And then you go through the Maya script again, in Matlab software, and you get a high-resolution 3D energy landscape.

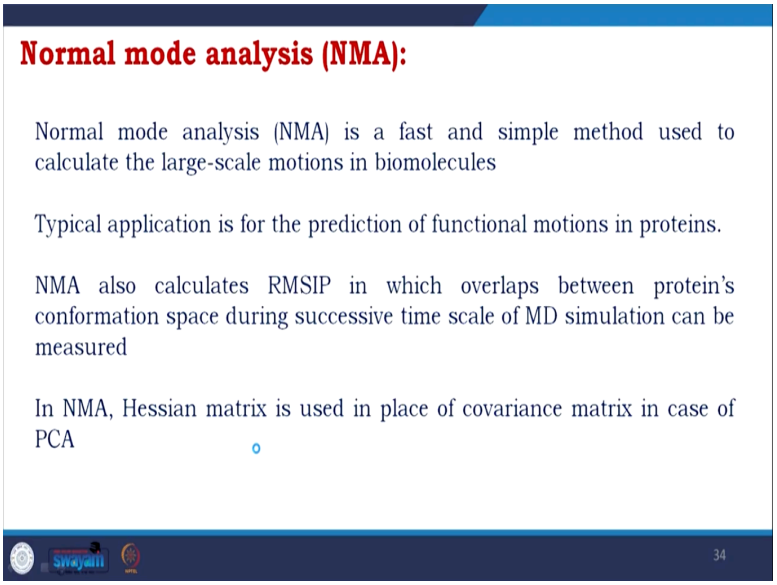
(Refer Slide Time: 45:33)



So let us take a look at the 3D free energy landscape where it is formulated by del G the free energy equal to the summation of -KTB ln PA -PB, where del G is Gibbs free energy case equilibrium constant TB is gas constant PA probability of A confirmation of protein and PB probability of B confirmation of protein. If you see the energy here, it is red to blue. It is higher to lower.

At the end, you see the peaks it is very sharp. The map you see at the lower is called a contour map. As I told it is a 3D map, so in one dimension, it is the principal component in the other dimension is principal component 2, and in the third dimension, it is free energy very interesting me from those peak points, you could isolate state of the protein. So these are the states you could have identified, and you could again correlate them with the functional implication.

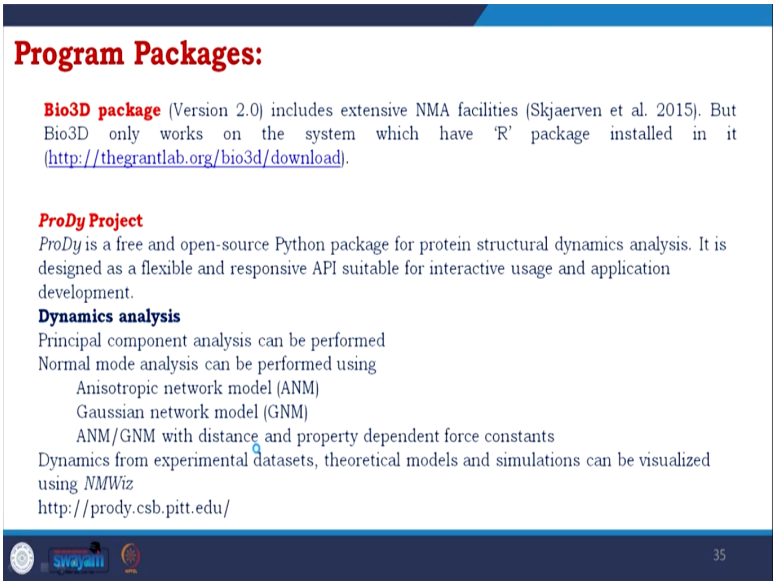
(Refer Slide Time: 47:18)



Coming to normal mode analysis is a fast and simple method used to calculate the large-scale motion in biomolecules. Here you do not need to do the simulation. If you have simulation data good if you do not have simulation data, but 2 confirmations still you could perform normal mode analysis. A typical application is for the prediction of the functional motion of proteins.

Normal mode analysis also calculates RMSIP in which overlaps between protein's conformation space during the successive time scale of MD simulation could be measured. In NMA Hessian matrix is used in place of the covariant matrix in the case of PCA.

(Refer Slide Time: 48:08)

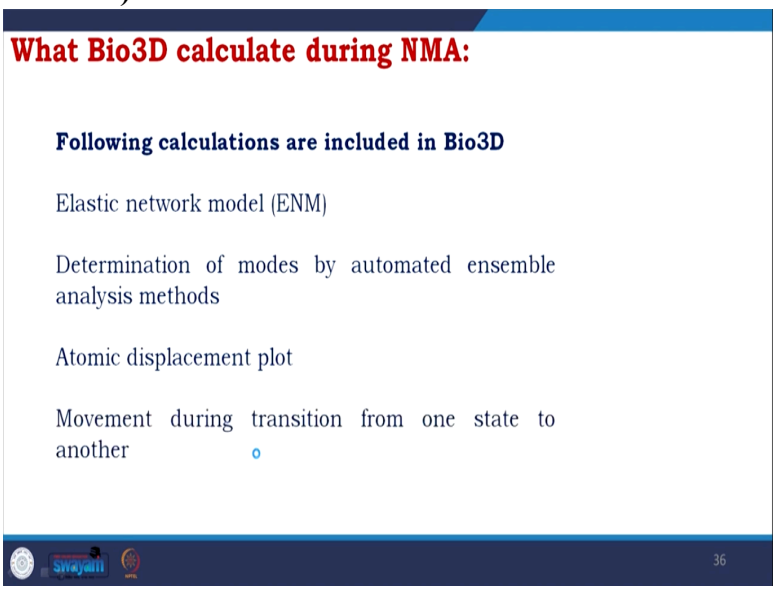


There are many program packages, and I would suggest 2, one little talk about is the bio 3D package, which includes extensive NMA facilities, but it only works on the system which are R package installed on the computer. Prody again is a free and open-source Python package

for protein structural dynamics analysis. It is designed as a flexible and responsive API suitable for interactive usage and application development in terms of dynamics.

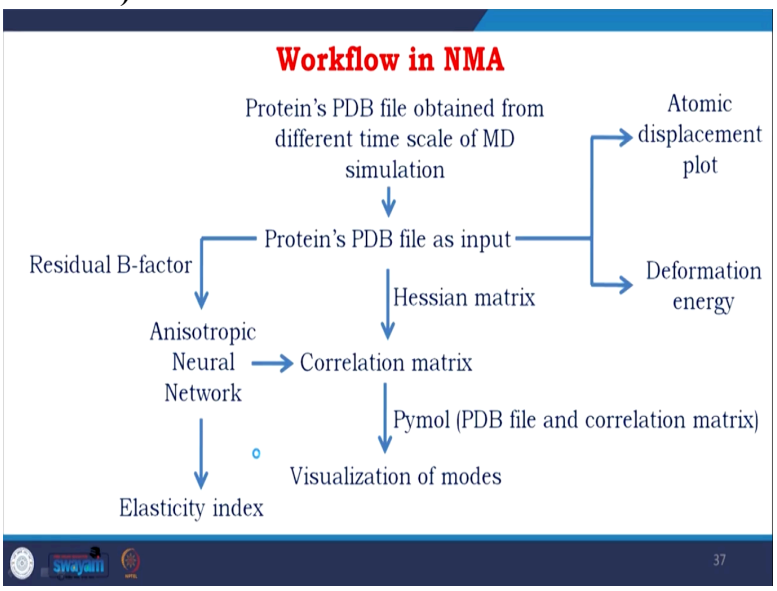
They do, they analyze principal components, and they do normal mode analysis, where an isotropic network model or ANM Gaussian network model GNM and ANM GNM with distance and property dependent for constants, dynamics from experimental data sets, theoretical models, and simulations can be visualized using NMWiz set wizard option. So, I have given the links and I would analyze a case study with a bio 3D package.

(Refer Slide Time: 49:21)



So, what bio3D does in terms of calculating NMA is develop the elastic network model ENM determination of modes by automated ensemble analysis method atomic displacement plot movement during the transition from one state to another.

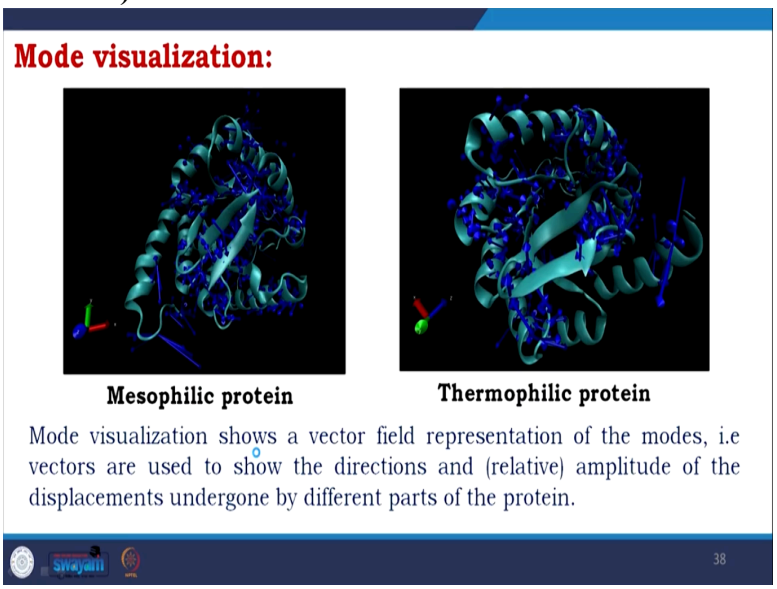
(Refer Slide Time: 49:46)



So, how the workflow happened proteins PDB file obtained from the different time scale of MD simulation this is taken as input and then you calculate residual B factor which gives you an isotropic neural network and ultimately that elasticity index which will give you the flexibility, on the other hand, the PDB is used by applying hessian matrix and together with an isotropic neural network.

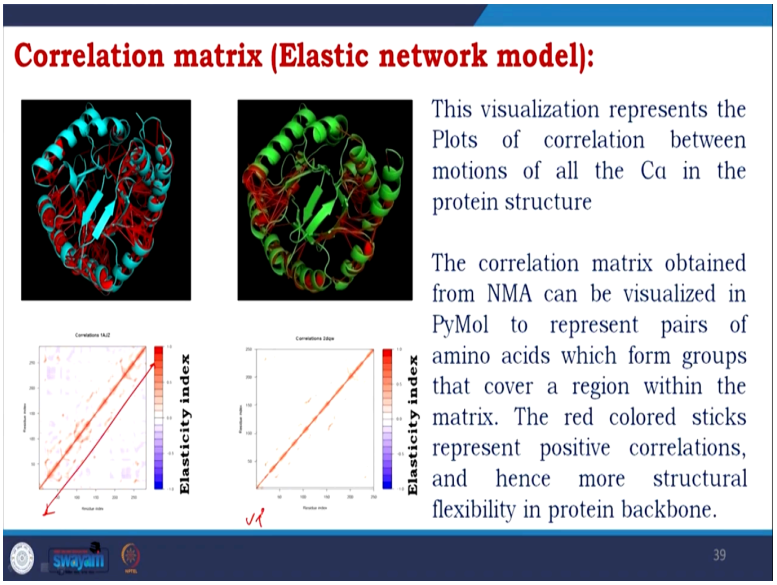
You get a correlation matrix where you see Pymol see the modes of movement you use that PDB file and the correlation matrix. Again, you could have directly obtained an atomic displacement plot, where you see the flexibility of the protein and deformation energy.

(Refer Slide Time: 50:46)



So, these are mode visualization this is a mesophilic and thermophilic protein and you see the blue arrows you see how the modes are there. So, the mode visualization options show a vector field representation of the modes that are vectors are used to so, that directions and relative amplitude of the displacement undergone by different parts of the protein. So when you look at it carefully, by looking at the arrow direction. You get the direction where the protein wants to move with the amplitude should have the arrow you understand where the movement is more or less.

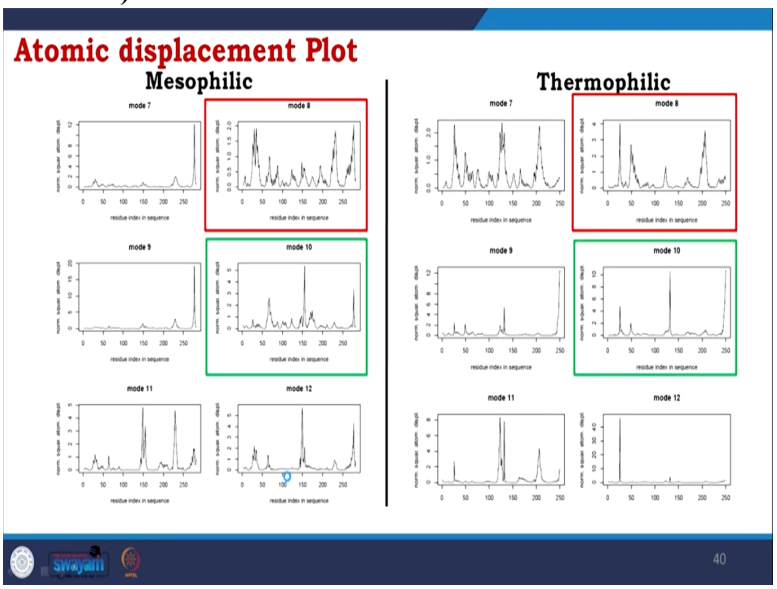
(Refer Slide Time: 51:37)



The correlation matrix comes from the elastic network model these visualizations represent the plots of correlation between the motion of all the C alpha and the protein structure. So, again you see the movement of the correlation matrix obtained from NMA can be visualized in Pymol to the present pairs of amino acids which form groups that cover a region within the matrix. If you look at these 2 you see the movements.

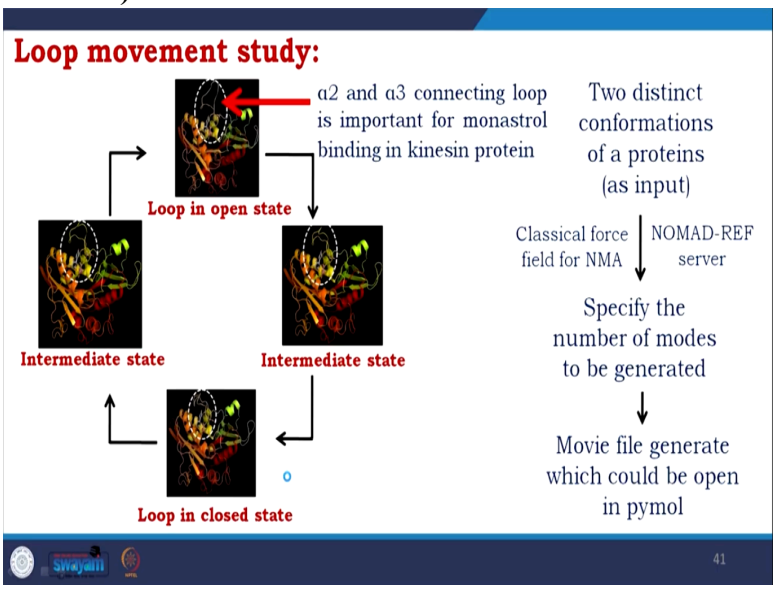
And now, if you look at the graphical representation, you see that in the case of the mesophilic protein, you get more of these red spots, it represents a positive correlation and hence more structural flexibility. So, with more red dots coming here, if you see, you could easily understand that the mesophilic protein we use here is more flexible in its backbone than the thermophilic protein.

(Refer Slide Time: 52:51)



Again atomic displacements plot, you see, you will easily understand that more displacement happened in the mesophilic in all the modes, we are showing B mode 8 B mode 10 B mode 12, and you will see more displacement in the mesophilic. So, when you will use your protein by comparing these you understand which protein is more flexible than another.

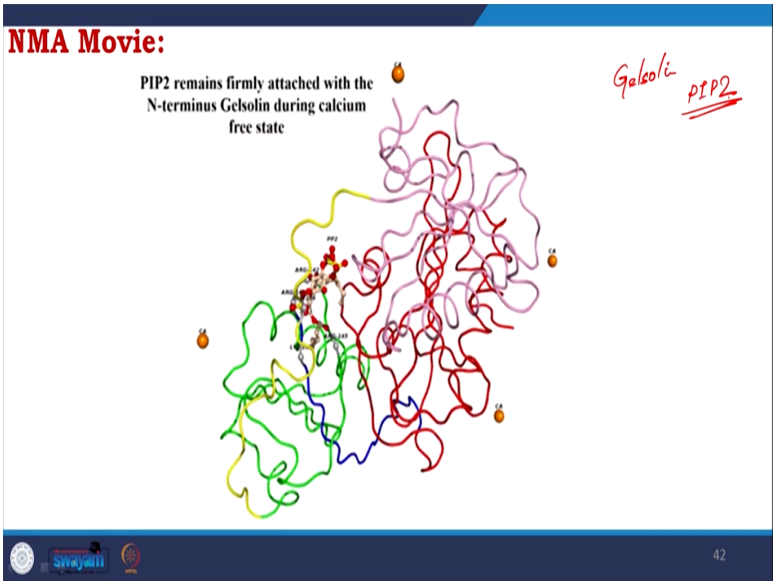
(Refer Slide Time: 53:23)



Also, you could study look movements of this if your loop has some intermediate states or closed states, open state you could study them here the alpha helix 2 and 3 connecting loop is important for monastrol binding in a kinesin protein kinesin proteins are very important proteins that are used for the study where you could develop inhibitors and you stop the function of this protein in case of cancer affected patients.

So, it is an anti-cancer drug monastrol. So, here as I told you 2 distinct conformations of a protein are used as the input you use the classical force field for NMA and NOMAD reference server, you specify the number of modes to be generated and you will generate a movie file which could open in Pymol. Though I am talking about the movement of the kinasin protein along with the monastrol.

(Refer Slide Time: 54:37)

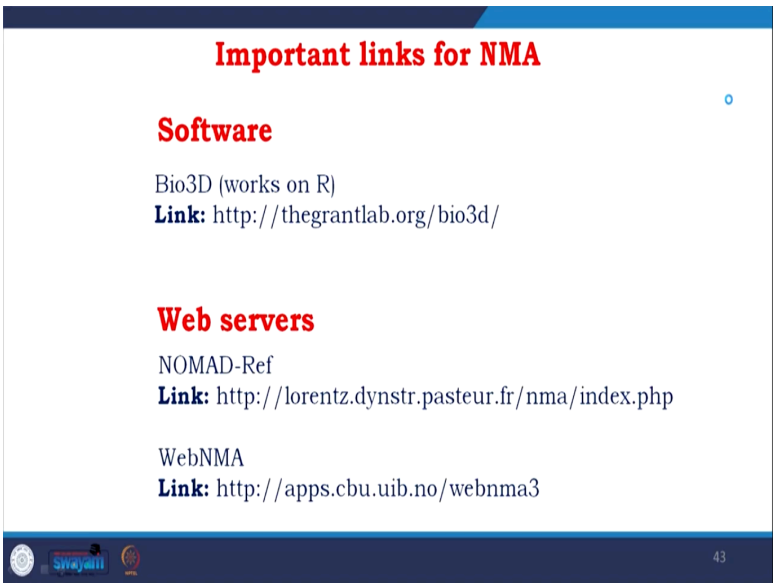


I did not use that protein. I use a protein called gelsolin and the 3 colors you see here are representing 3 domains of gel solid here gelsolin is interacting with PIP2, you see gelsolin interacts with PIP2 and binds it, but when calcium comes there is a huge change in the conformation and there is the opening up of the domains, you see the huge movement if you carefully look at the blue portion of the loop.

You will see how extensively it moved and that makes PIP2 come out of the system. So, the opening has very significant biological relevance in that open state it could bind to acting, especially cytoskeletal actin, and sometimes this results in cardiac muscle contraction, hence, heart attack. So, if you could use a PIP2 derivative as a drug, you could stop this domain movement and without being opened or without coming into this state you are looking at it will not be capable of interacting with the acting.

Hence, there is no cytoskeletal muscle contraction is there is no heart attack. So, a PIP2 derivative with high affinity binding stopping the domain movements would use as a drug that stops the heart attack of the patient. So, it stops the cardiac-acting gelsolin interaction, hence contraction. So, this is an example of an NMA movie.

(Refer Slide Time: 57:16)



These are a few important links up NMA you as I already have given the Bio3D link, the web servers Nomad and WebNMA are also popular. So with that, we finish today's class, and we talked about the process of data analysis and the different methodologies we use to analyze the data. Thank you very much for listening. Thank you.