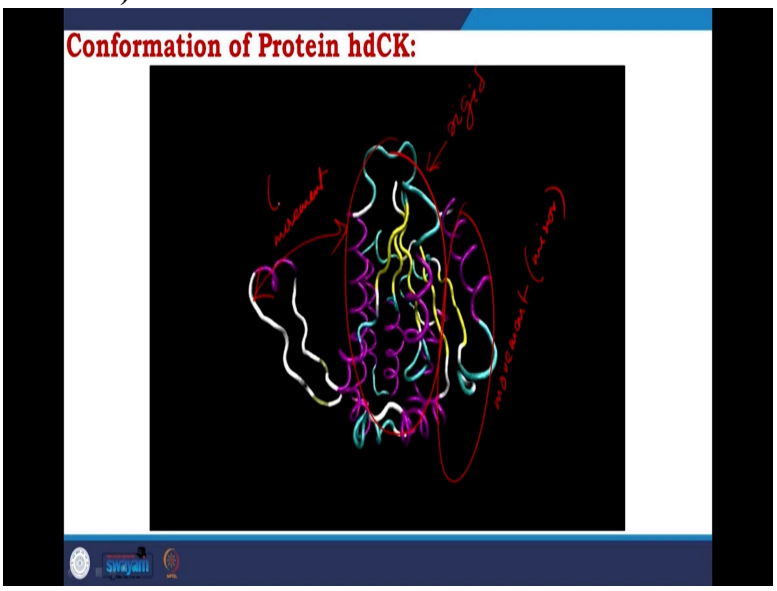
Structural Biology
Prof. Saugata Hazra
Department of Biotechnology
Indian Institute of Technology - Roorkee

Lecture - 50
Application of Molecular Dynamic Simulation

Hi everyone, again, welcome to the structural biology course. Today is the last class of the module MD simulation. In the previous 4 classes, we have discussed general features, attributes, force fields, and conceptions about MD simulation. Is how it is going? In the second and third classes, I discussed the process in detail. In the 4th class, I discussed data analysis.

Today in the last class, I will talk about 2 case studies, which will help you to understand how the knowledge you have acquired could be used in helping your research. The first case study on the application for molecular dynamics simulation is very short. I put that story because I wanted to tell you it is not always we perform MD simulation and publish the MD simulation data. It could even help design your experiment. So, the protein you are looking at here in the movie is the protein I studied in my Ph.D. This protein is called human deoxycytidine kinase.

(Refer Slide Time: 02:07)



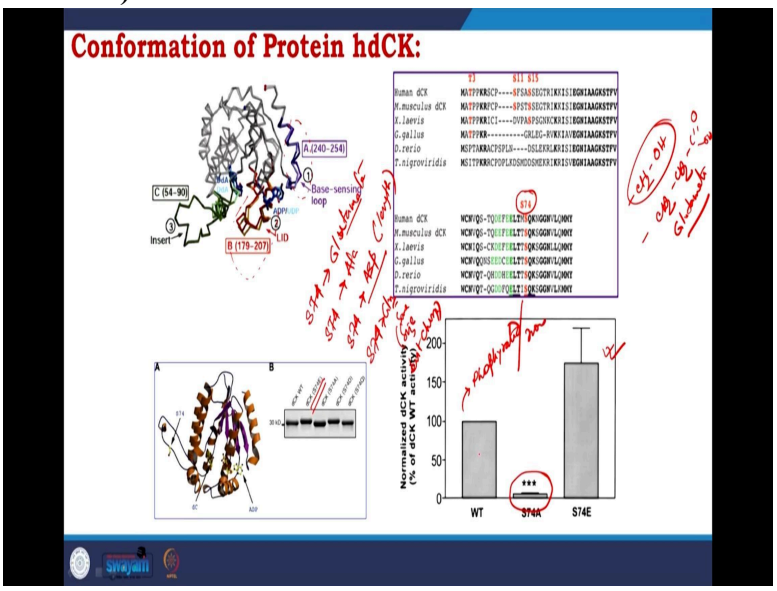
Human deoxycytidine kinase is a very important protein because it is one of the major targets in cancer chemotherapeutics. So, people study this enzyme to try to understand the design of new drugs and all those things. Specially nucleotide analogs are mostly successful drugs. And people design many of them. Your name is cytabine, gem cytabine. And mostly, the L

nucleoside analogs are used as anti-cancer drugs. Why study this protein? I was not someone who used to use or is an MD simulation expert neither was my lab.

But being in Chicago at the time of 2004, 5,6,7 where a revolution took Klaus Holton in UI UC Benard Ross in Chicago. I also wanted to learn with my initial understanding. I have run this protein a 5 nanosecond run. What are you all looking at? Many of you who have now acquired knowledge about MD simulation probably think that 5 nanoseconds are not a considerable amount of run. But that time with the 2 core computer it takes a huge time.

But what I am trying to say is if you look at this carefully, you will understand that this 5-nanosecond run and the movie I have created from these helped me understand that the central core of my protein has certain rigidity. So, this part is rigid. There is a big movement of this loop, which also has movement. So, this is a minor movement, and this is a major movement. How was this information helpful in my research?

(Refer Slide Time: 05:09)



When I was working, this convinced me because I saw the movements. So, there would be other conformation of the protein. So, I keep trying different crystallization conditions. And if you see that, we see a clear movement here, and we also see a movement here. So, we published a modified form of this protein by doing cystography. But the inspiration I got came from my initial MD simulation data. Now I will talk about the LID region. If you go back again, you will see the LID region is the most moving region.

Especially the tip region, if you see, they undergo a huge change, show the time we did not have programs to check. I thought there might be a chance of post-translational modification, which probably changes the conformation. At that time, we do not have the availability of the programs as servers. So I wrote a program and compared the sequence of different deoxycytidine kinases, and I found that at position 74, there is a serine that has a high chance of being phosphorylated.

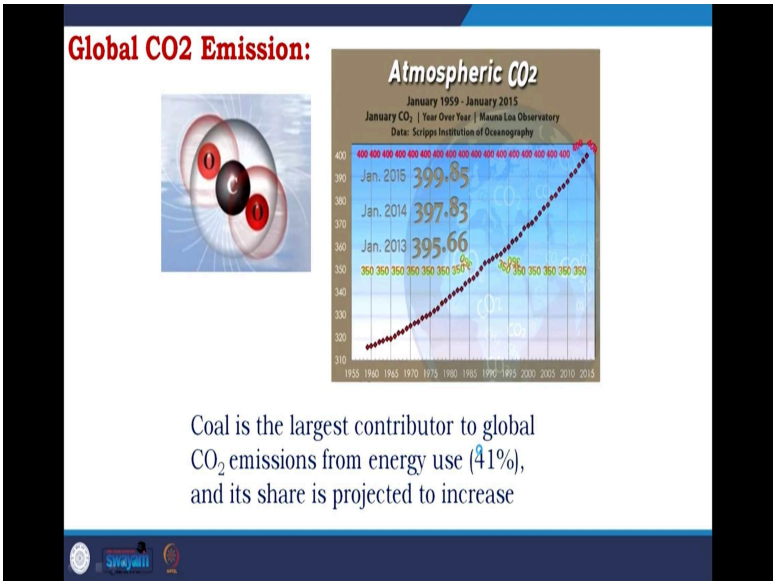
I am getting that idea along with the movement I have seen in my protein structure. I make a phospho mimic. Students who do not know about phospho mimic, phospho mimic is, you know, nowadays, you see something there is a probability of phosphorylation. You buy a kit and phosphorylate. That time in the cystography lab in 2005, 6 we did not have that expertise or courage. But on the other hand, being structural biologists, we did our modeling.

And if you do modeling, you could find that where you have a serine, if you replace it with glutamate, it mimics the serine in its phosphorylated state. So, we make a few mutations to make our logic clear. First, as I told we make S74 to glutamate. We change S74 to alanine, so we make glutamate. We make alanine to see what is the null effect. We made S74 to aspartate because we wanted to put that it is not the charge, only the length. After all, glutamate and aspartate have the same charge, but the length is less.

So we wanted to compare and make glutamine which will say size, same size but charge. So we mean it would only work if the length and the charge fit and when we make an S74 e mutant. We get high activity if you compare in comparison to a S74A; It is high. But the wild type was phosphorylated when we put it in the cell. So, there is a mixture of phosphorylated and nonphosphorylated, which is why the differences exist.

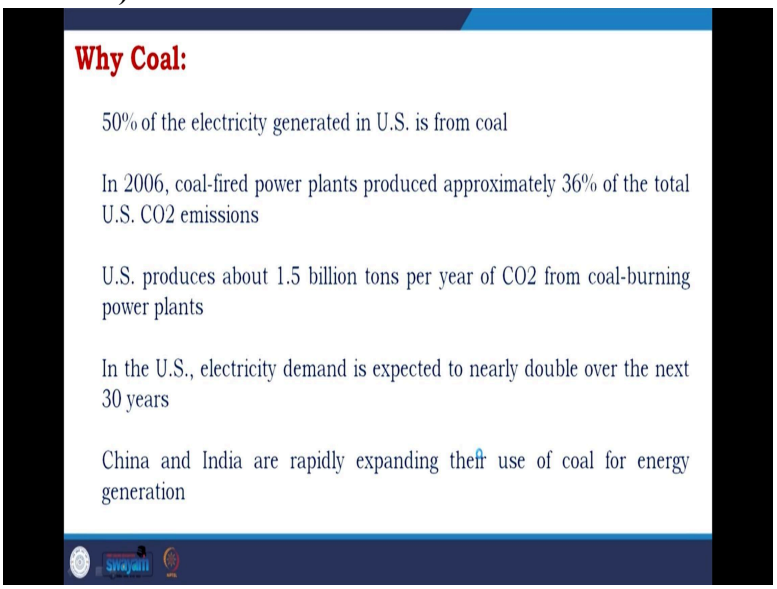
So, by making a simulation, I have understood and designed my experiment successfully, which is one of the very interesting applications. Now, coming to another one where I have done in- silico work to make a thermostable carbonic anhydrase design. So, this is the application of molecular dynamics simulation in screening thermostable carbonic anhydrase for industrial usage.

(Refer Slide Time: 11:02)



First, let me tell you about the project. You all know about global CO2 emissions. If you see this, you will see how it has increased rapidly. Coal is the largest contributor to global CO2 emissions from energy use, which is 41%, and its share is projected to increase daily.

(Refer Slide Time: 11:30)

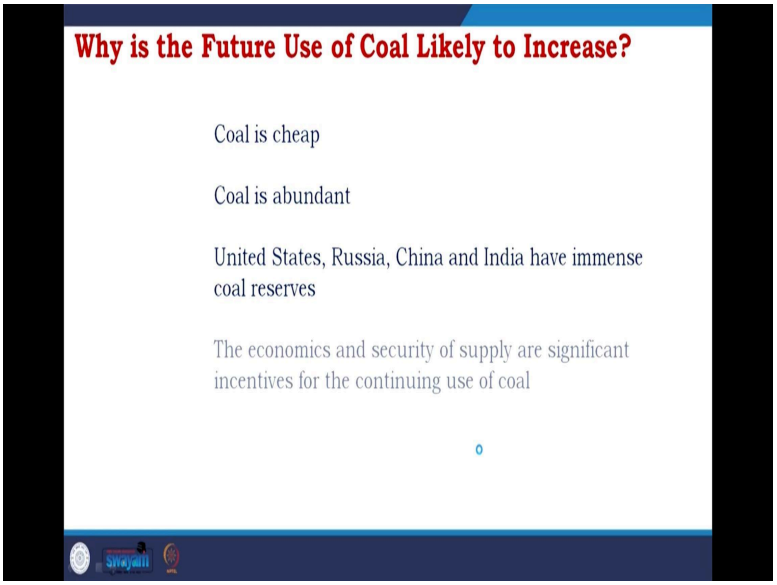


So, who is unaware of why coal is used? Let me tell you 50% of the electricity generated in the US is from coal. In 2006, coal-fired power plants produced approximately 36% of US CO2 emissions. The US produces about 1.5 billion tons per year of CO2 from coal-burning power plants. In the US, electricity demand is expected to nearly double over the next 30 years, meaning more CO2 emissions.

Now, why do I talk about the US? Because they are maintained those rules and restrictions the best.

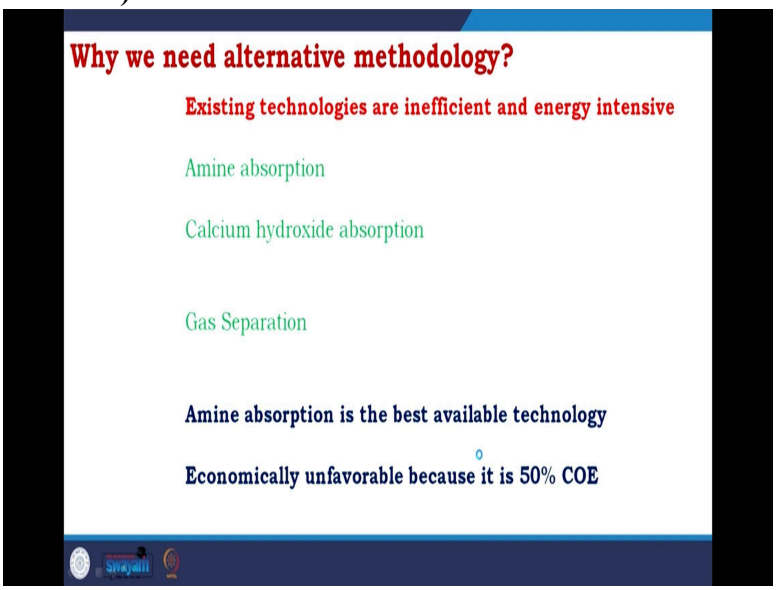
So, if there are statistics of the US for other countries, you could easily understand they are even worse than China and India, rapidly expanding their use of coal for energy generation.

(Refer Slide Time: 12:40)



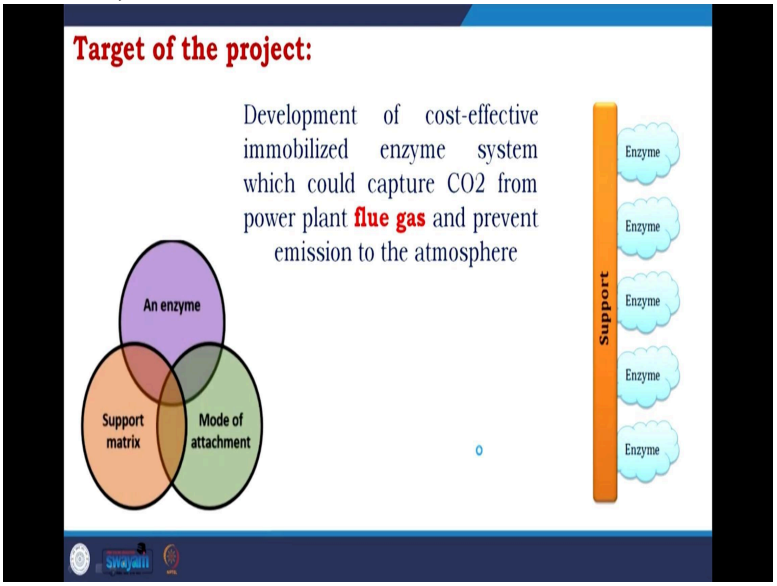
Why is coal the future use of coal likely to increase? 1 coal is cheap, 2 coal is abundant. United States, Russia, China, and India have immense coal reserves. The economics and security of supply are significant incentives for the continuing use of coal.

(Refer Slide Time: 13:05)



So, when CO2 gas is emitted, what is the existing technology? So, existing technologies are inefficient and energy-intensive. There is amine absorption, calcium hydroxide absorption, and gas separation among the 3 process amine absorption, calcium hydroxide absorption, and gas separation. Amine absorption is the best available technology. And for these, the best one is economically unfavorable because it takes 50% of the efficiency cost. So, you are generating power 50% of what is needed to absorb the emitting carbon dioxide.

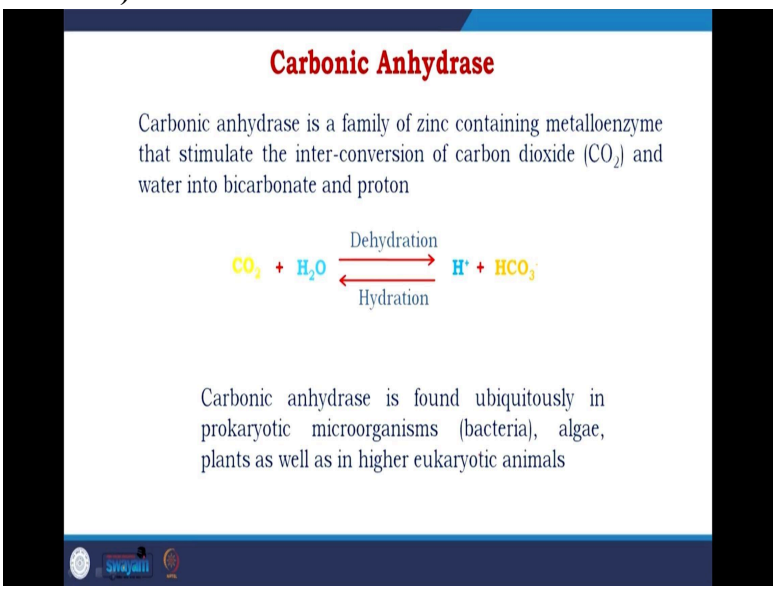
(Refer Slide Time: 14:11)



So, we have targeted a project to develop a cost-effective immobilized enzyme system. We want to develop a biological enzyme that could capture carbon dioxide from power plant flue gas and prevent its emission to the atmosphere. So, we are trying, you know, if flue gas is coming out, this is high temperature, and no biological enzyme can be there.

If it were there, it would be denatured, which means it is non-functional, but flue gas, when it comes out, you could cool it down, coming in around 90 or 100 degrees, and then go through the screening system of this biological enzyme. So, what are trying to, we are trying to get an enzyme that could absorb CO2. We want to put it in a support matrix where it would be attached. So, this would be a device like that where there would be a support system, and there would be enzyme immobilized. This is called an enzyme-immobilized support device.

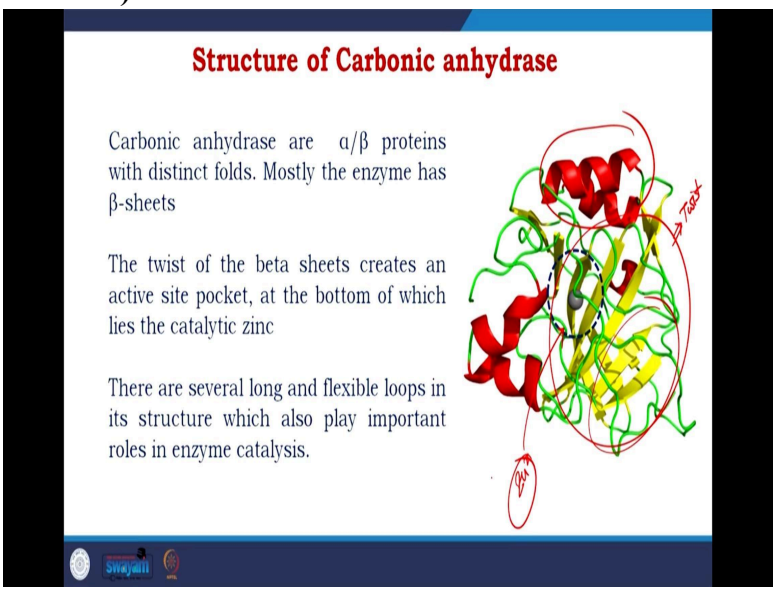
(Refer Slide Time: 15:30)



And our choice was carbonic anhydrase because carbonic anhydrase is a family of zinc-containing metalloenzymes that stimulates the interconversion of carbon dioxide and water into bicarbonate and proton. So, it would convert CO2 to bicarbonate, and further, you could convert it to methanol and other products. So, not only do you absorb the carbon dioxide, but you can also develop products out of it.

This is the biochemical reaction of carbon dioxide plus H 2 O through producing bicarbonate. Also, an interesting fact is that carbonic anhydrase is found ubiquitously in prokaryotic microorganisms, bacteria, algae, and plants and in higher eukaryotic systems. So, this could do, and this is available. These are the 2 points which make us interested to work on.

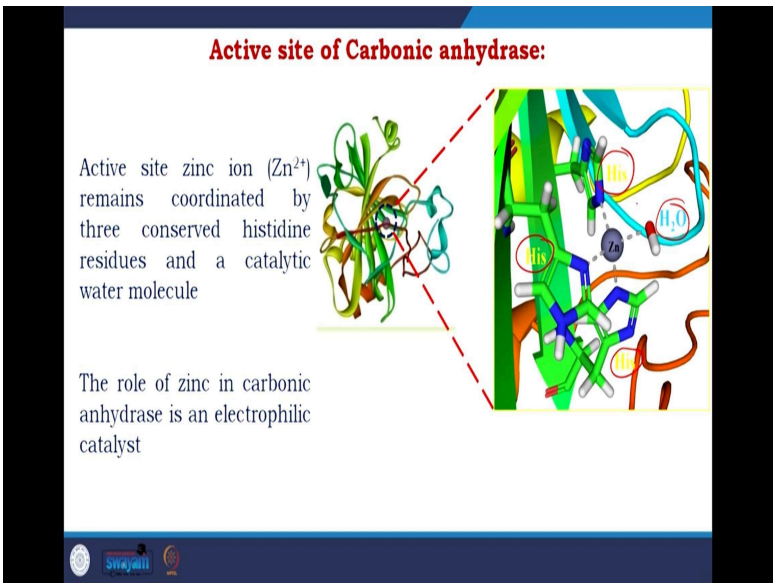
(Refer Slide Time: 16:32)



Before going into the details of the process, if you look at the structure of carbonic anhydrase. It is an alpha-beta protein with a distinct fold. Mostly if you look at so, there are alpha helices and beta sheets. There is a twist if you see it in this portion. You will find that there is a twist of the beta-sheet and these twists create the active site pocket if you see a zinc at the bottom of which lies that catalytic zinc.

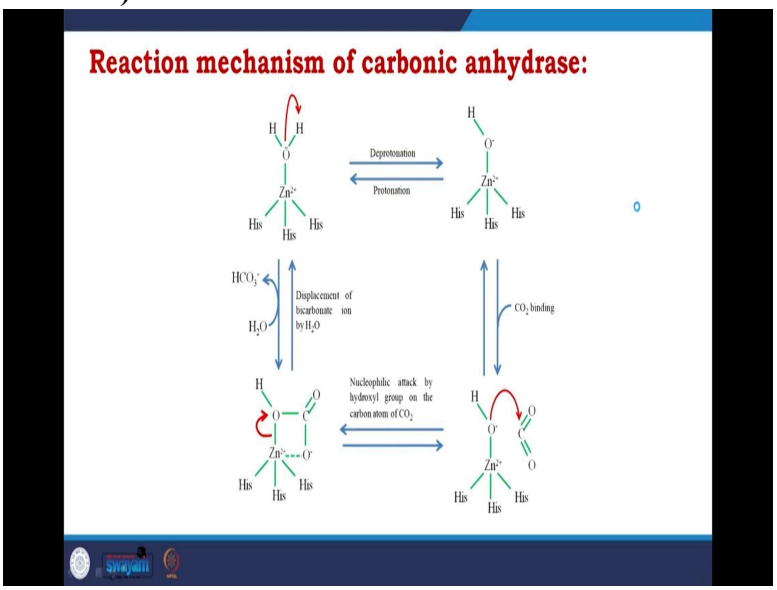
Several long and flexible loops in its structure also play an important role in enzyme catalysis. So, if you see these green ones, green ones are loops, you see that this enzyme is flexible, and these loops have very interesting or critical roles in catalysis. But, at the same time, these long loops would make your enzyme less adaptive to high temperatures.

(Refer Slide Time: 18:03)



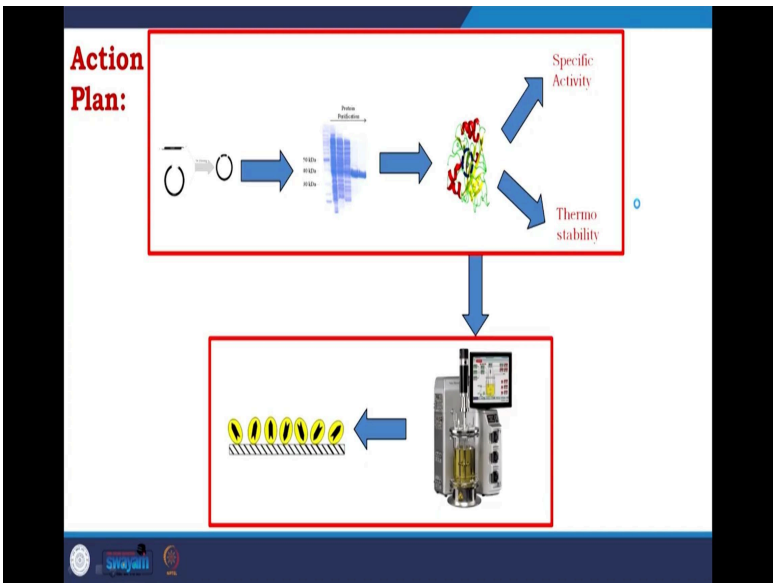
If you look at the active site of the enzyme carbonic anhydrase, you see if you look at here, it is zinc. This coordinates with histidine residues 1 histidine residue, 2 histidine residue, and 3 histidine residue, and a catalytic water molecule here. The role of zinc in carbonic anhydrase is an electrophilic catalyst. We will take a look at that.

(Refer Slide Time: 18:39)



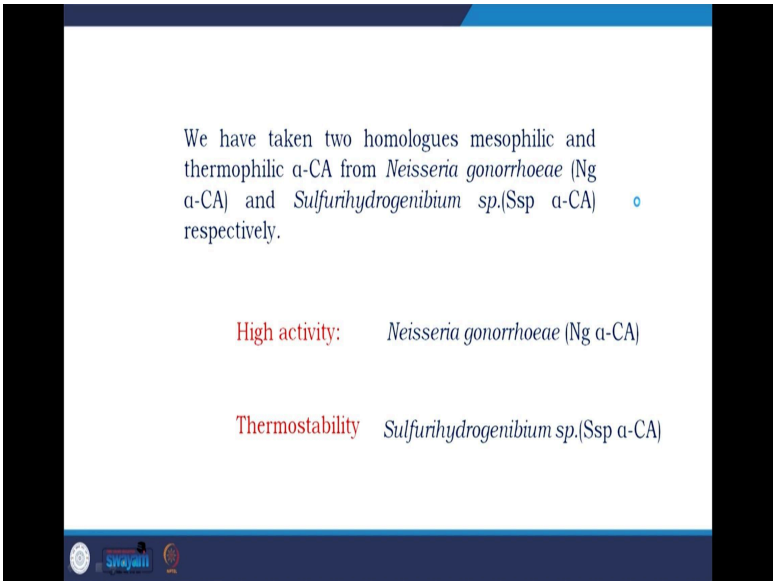
So, this is what is happening mechanistically inside the enzyme, the zinc binds to the water, and then it deprotonates. So, the oxygen becomes negative and binds to the carbon dioxide. Then, there would be a nucleophilic attack by the hydroxyl group on the carbon atom of the carbon dioxide and, finally, the displacement of the bicarbonate ion by an upcoming H 2 O. So, that is the reaction mechanism.

(Refer Slide Time: 19:21)



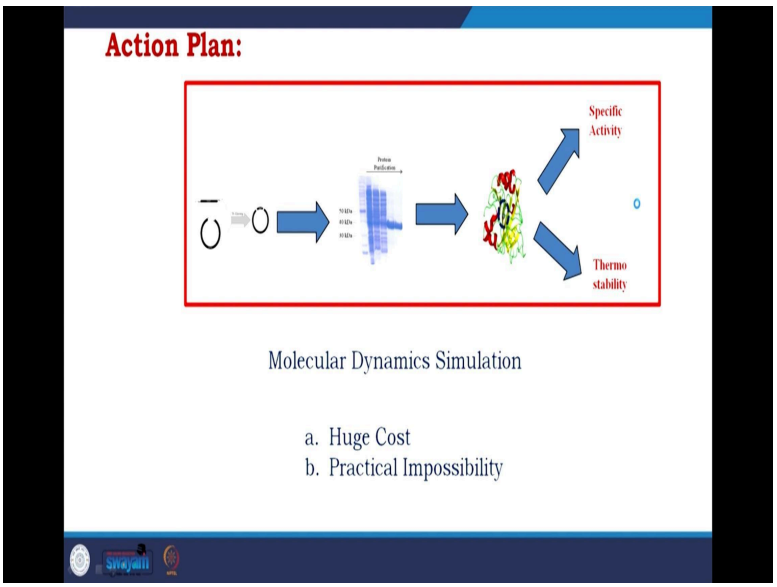
Our initial action plan was to clone probable carbonic anhydrase and then screen them based on specific activity and thermostability. Then when we get a good one, we could use a biochemical engineering approach to scale it up and make the enzyme in high amounts and then immobilize it in the device and use that as a screener or absorber on the exit side from where the flue gas is coming out. So, there are 2 phases one is protein biochemistry, and the second one is the biochemical engineering approach.

(Refer Slide Time: 20:20)



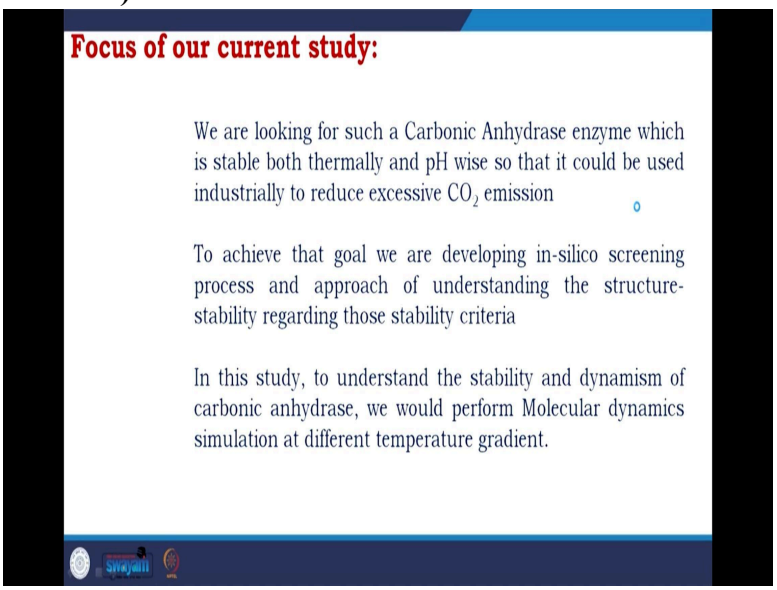
So, we want to screen carbonic anhydrase with high enzyme activity and thermostability.

(Refer Slide Time: 20:30)



But, the whole process was high cost demanding and practically impossible because you have to bring new enzymes, clone every one, and then test their specific activity thermostability. It demands a lot of money, a lot of manpower and kind of says that it is impossible. So, we decided to replace the whole thing with molecular dynamics simulation.

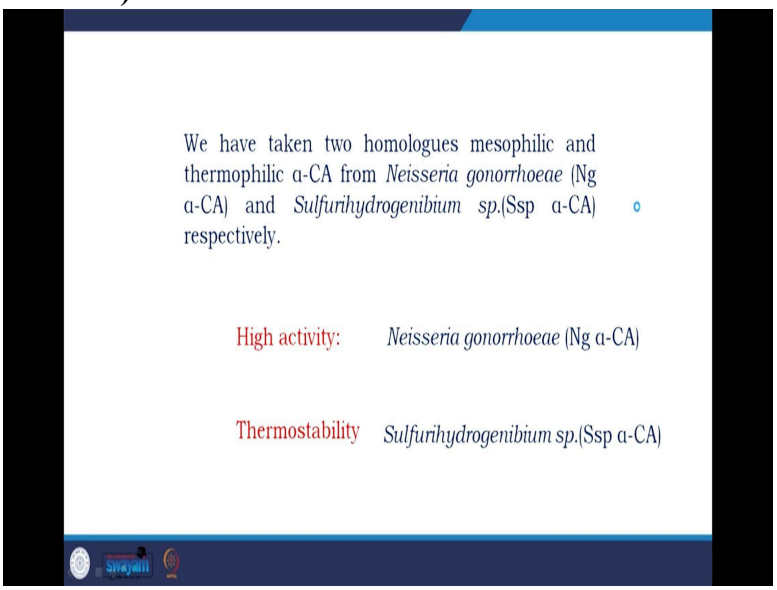
(Refer Slide Time: 21:02)



What do you want to do? We want to look for such a carbonic anhydrase enzyme that is stable both thermally and pH-wise so that it could be used industrially to reduce excessive carbon dioxide emission. To achieve that goal, we are developing in-silico screening instead of in vitro screening. Initially, we plan to alter the in-silico screening process and approach to understanding the structure stability regarding those stability criteria.

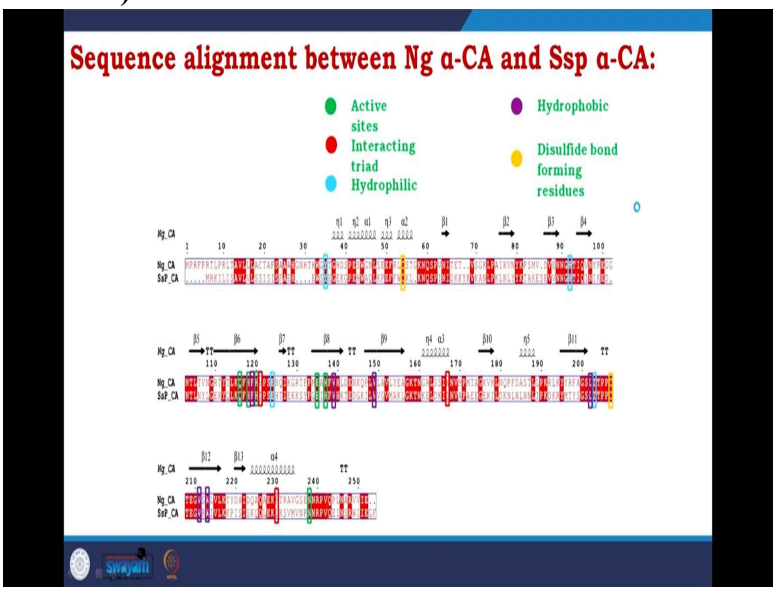
This study aims to understand the stability and dynamism of carbonic anhydrase. We decided to perform molecular dynamic simulations at different temperature gradients. So, let me explain the plan more elaborately.

(Refer Slide Time: 21:57)



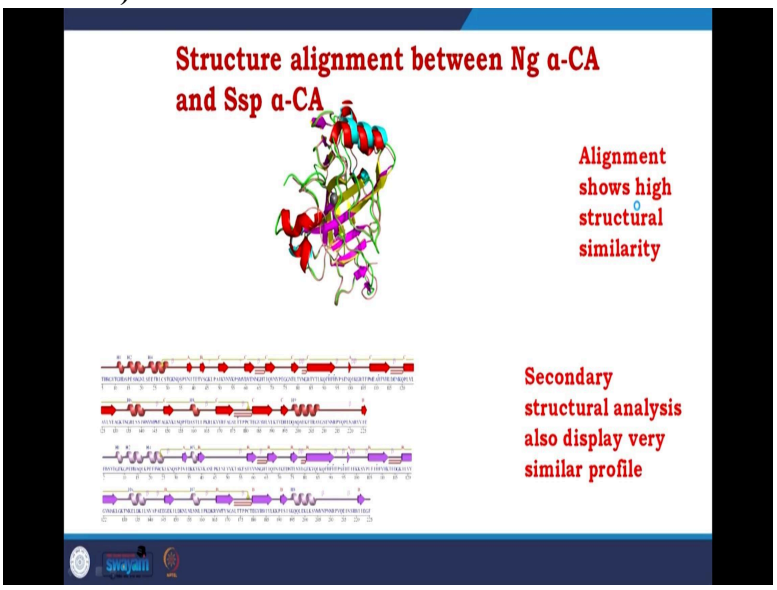
We already know that 1 alpha carbonic anhydrase found called Sulfurihydrogenibium species has very high stability. We know carbonic anhydrase from Neisseria gonorrhoeae with very high activity. So, as I told high activity, Neisseria gonorrhoeae alpha carbonic anhydrase, which we call Ng alpha CA, and thermostability Sulfurihydrogenibium species, which we call Ssp alpha CA, we decided we would understand the reasons for thermostability from Ssp alpha carbonic anhydrase. And those understanding, we will use it to make Neisseria gonorrhoeae alpha carbonic anhydrase engineered toward thermo stability.

(Refer Slide Time: 22:58)



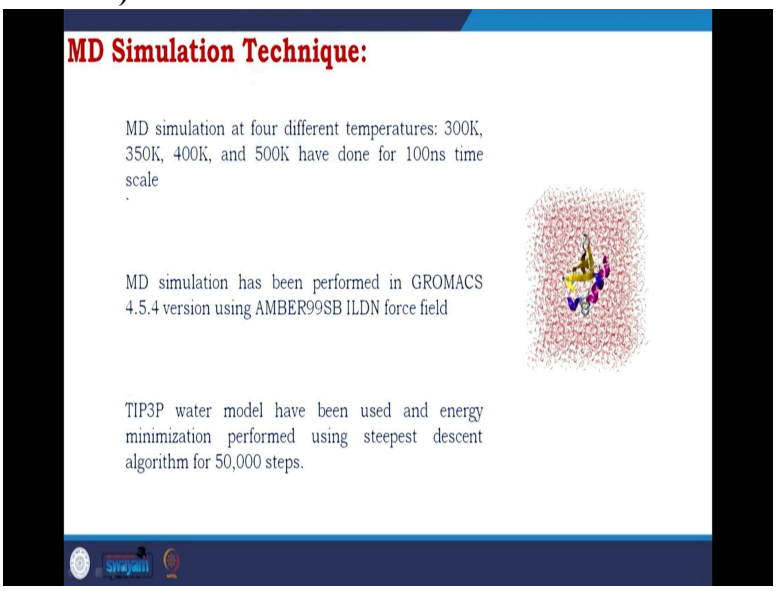
This compares sequence and secondary structures where we have taken information on the active site. You will see the green boxes interacting triad, the red boxes, and the hydrophilic residues, which are sky blue. The hydrophobic ones are purple, and the Disulfide bond forms orange residues. So, you get an idea about the protein.

(Refer Slide Time: 23:29)



These are 2 proteins; if I align 1 protein with another, you will agree that the alignment shows high structural similarity. So these 2 proteins are structurally similar, so if we get the knowledge from the Ssp alpha CA, it could be easy to apply to Ng alpha CA. Also, when we compare their secondary structures, we see a similar profile between the 2.

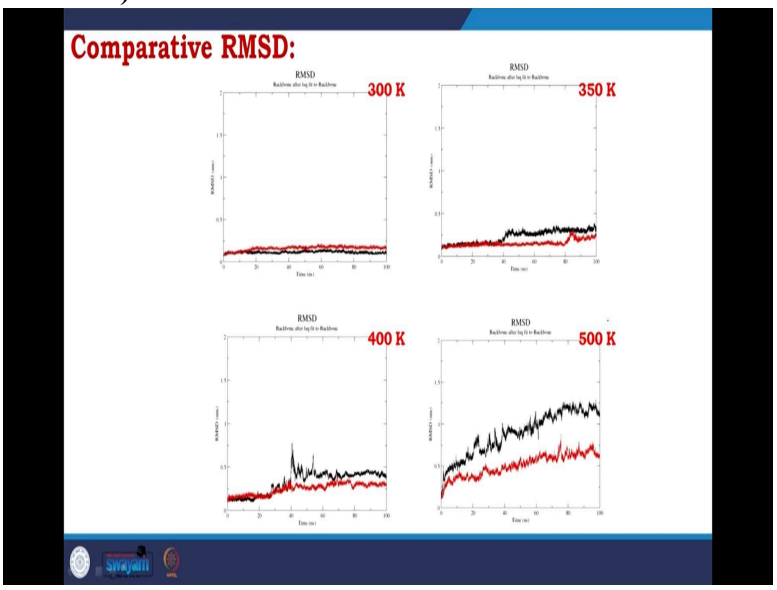
(Refer Slide Time: 24:11)



So, we go for MD simulation techniques. We perform MD simulation at 4 different temperatures, 300 Kelvin, 350 Kelvin, 400 Kelvin, and 500 Kelvin, and have done it for 100 nanosecond timescale. MD simulation has been performed in GROMACS 4.5.4 version using

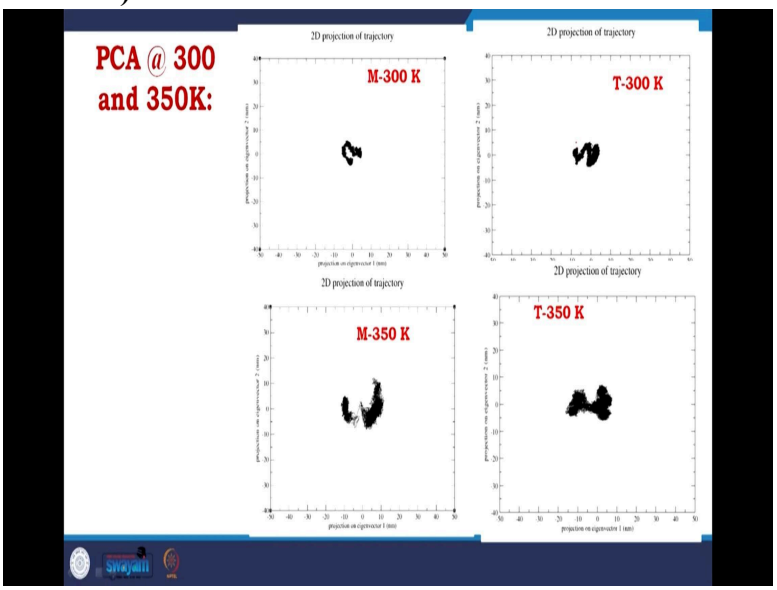
AMBER99SB ILDN force field. Generally, these we use for only protein. We used the TIP3P water model, and the energy minimization was performed using the steepest descent algorithm for 50,000 steps. And as we said, all the runs were run for 100 nanoseconds in the production run.

(Refer Slide Time: 25:08)



So, this is the analysis of the result, you see that query compares the root mean squared deviation, and we see that from 300 to 300, interestingly, we see more flexibility on the thermostable protein. Whereas, when it shifted from 300 to 350 further, we see that gradually there are more RMSD values in the case of mesophilic protein, which was expected.

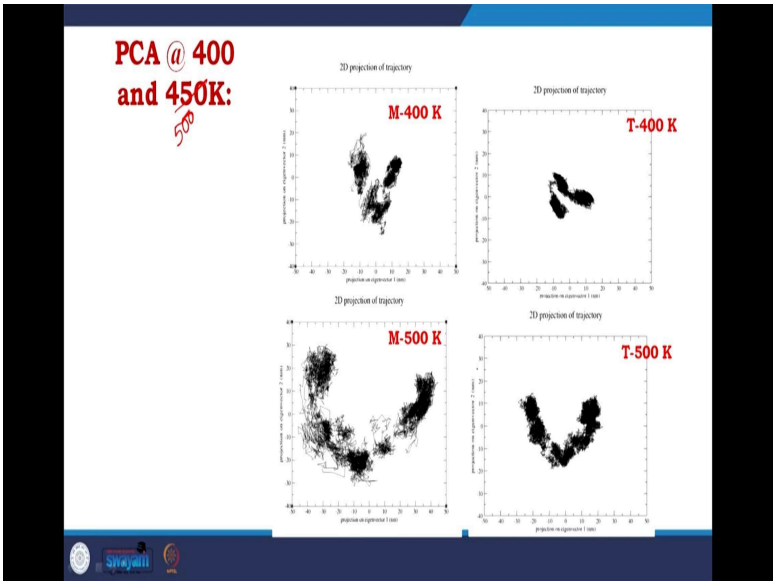
(Refer Slide Time: 25:55)



We have also checked the principle component analysis PCA. And if you see, we have performed 300 Kelvin. In 300 Kelvin, again, you see the mesophilic protein. The Ng is smaller than the thermophilic one. But when it is changed from 300 to 350, the PC 1 and 2

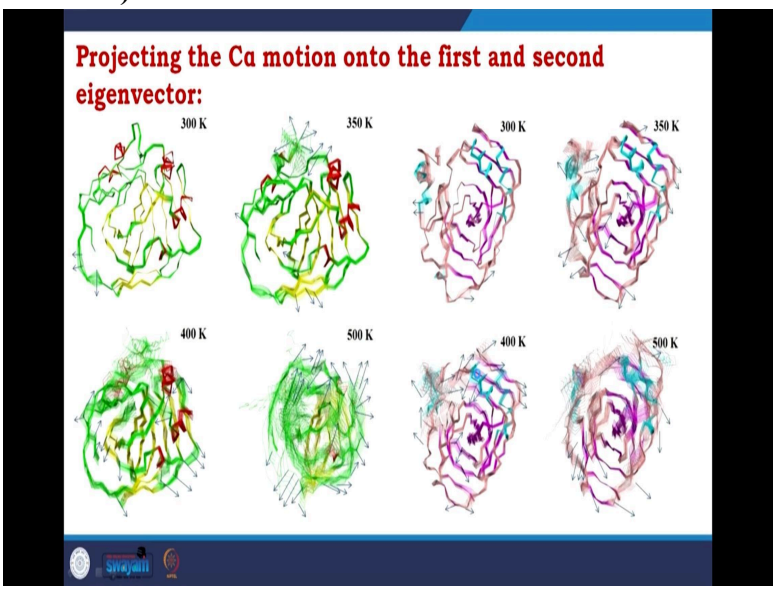
graph the projection on eigenvector 1 and 2 had increased significantly whereas, the relatively less increase we have seen in the carbophilic one.

(Refer Slide Time: 26:36)



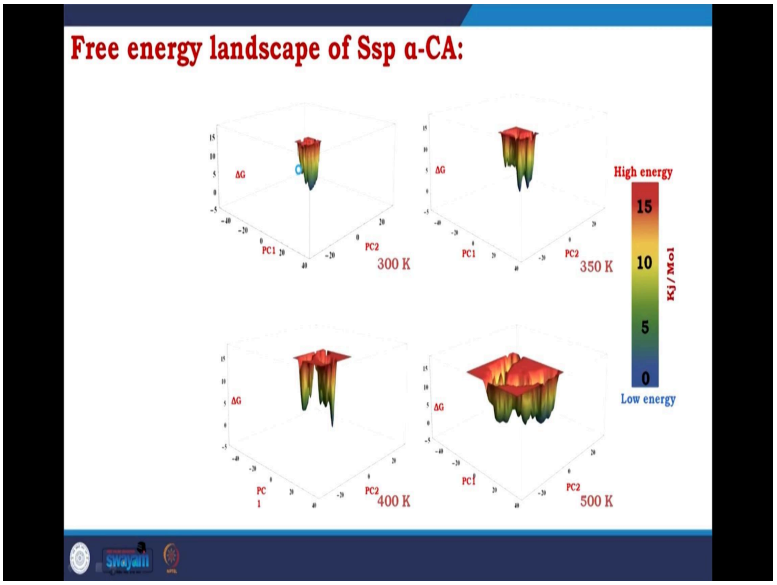
When we come to 500, you will see it opens up radically at 500 Kelvin.

(Refer Slide Time: 26:52)



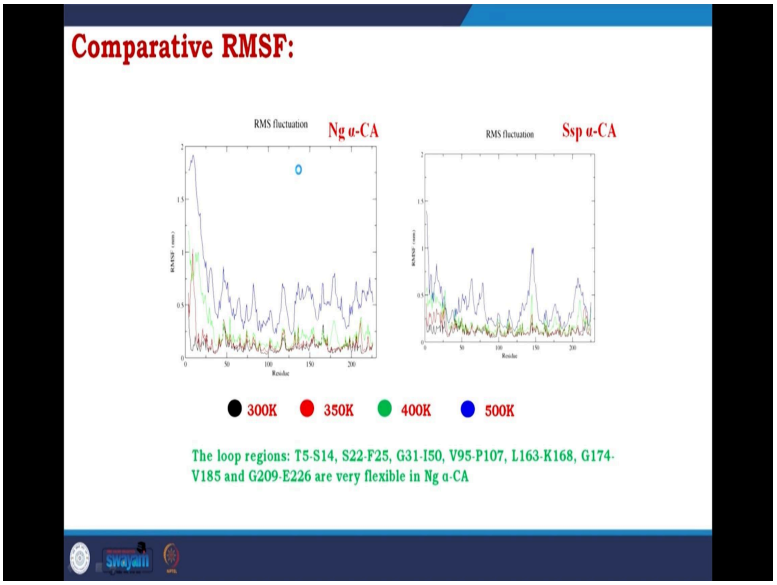
We will also look at the C alpha motion onto the first and second eigenvector. And here also we see a similar trend.

(Refer Slide Time: 27:06)



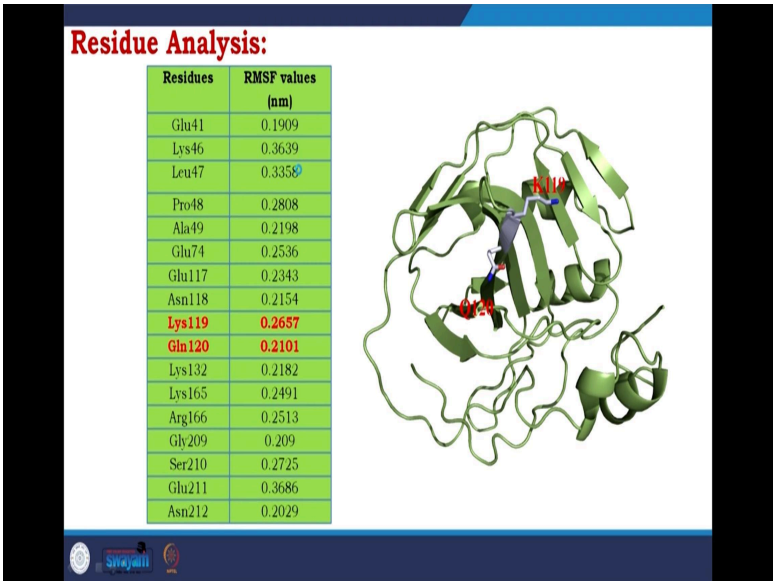
We look at the free energy landscape of Ng alpha CA. As I told these are the free energy landscape representation where PC 1 and PC 2 are in 2 dimensions and free energy is in the third dimension, we see that the graph is expanded, and at the initial stage, the UR 1 square is extremely served that become broadened in higher temperature. A similar trend is absorbed in the thermostable protein. But in a much-reduced scale in comparison to the mesophilic protein.

(Refer Slide Time: 27:57)



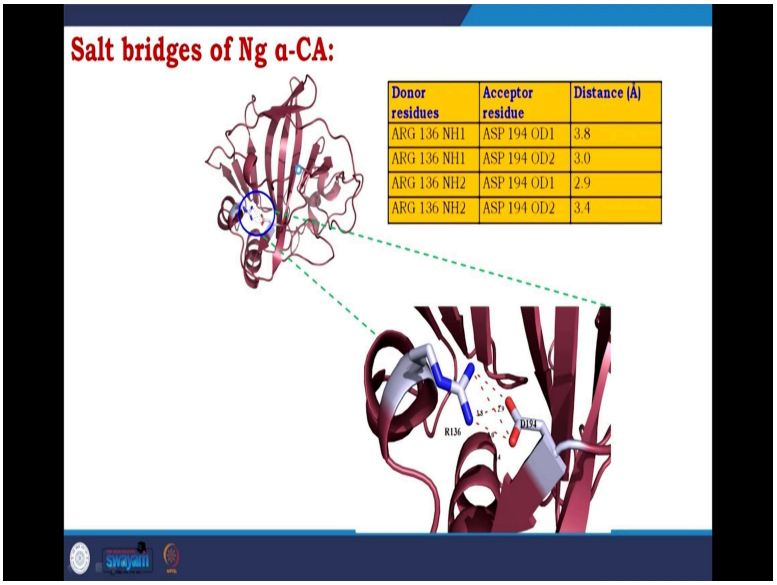
When we look at the comparative RMSF we see that the flexibilities are high in the case of Ng alpha CA compared to Ssp alpha CA. In the loop region, the threonine 5 to serine 14, serine 22 to phenylalanine 25, glycine 31 to isoleucine 50, valine 95 to proline 107, leucine 163 to lysine 168, glycine 174 to valine 185, glycine 209 to glutamate 226 are very flexible in Ng alpha CA, which is in the loop region.

(Refer Slide Time: 28:53)



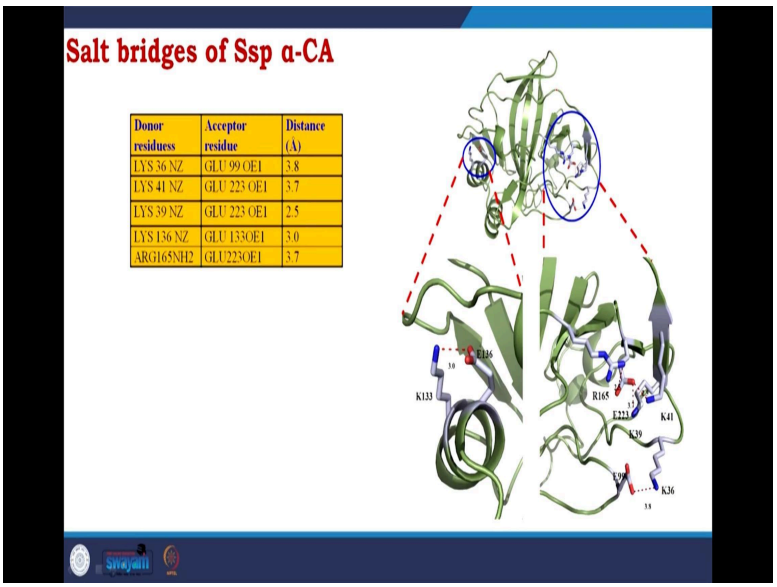
We have done individual residue analysis in terms of the fluctuation values.

(Refer Slide Time: 29:02)



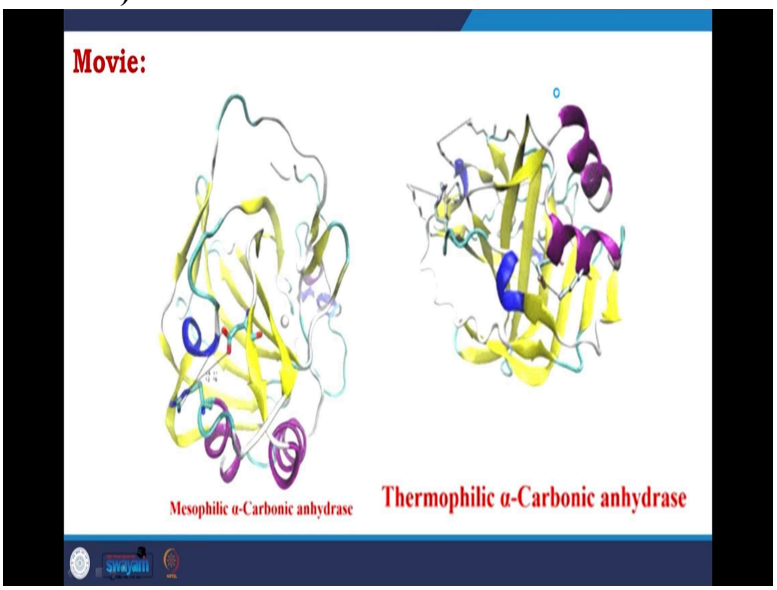
We will also check the salt bridges; salt bridges are non-covalent interactions, but because of the development of real charge, the interactions are high energy when you compare between the non-covalent interactions. So, in the case of Ng alpha CA, we see that the interactions happen mostly with 2 amino acids, the pair arginine in 136 and aspartate 194. So, we have observed 1 salt bridge formation development in the case of Ng alpha CA.

(Refer Slide Time: 29:45)



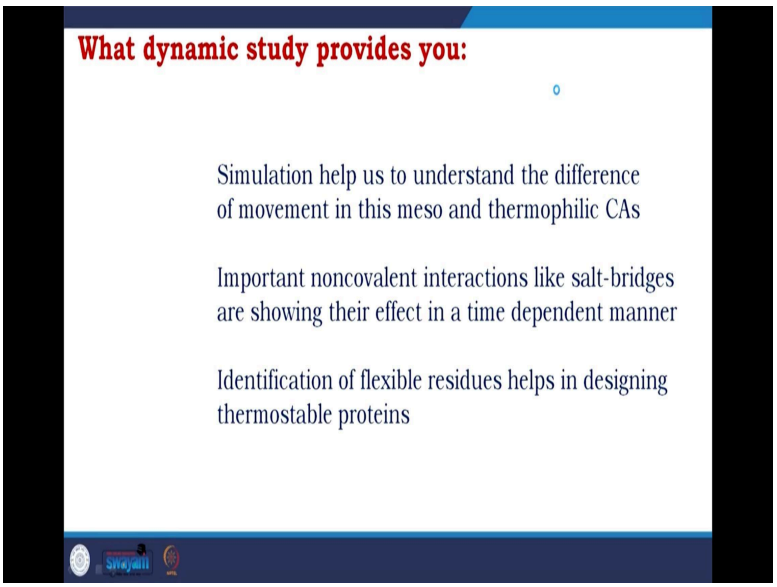
In the case of Ssp alpha CA, we see the salt bridge formation between lysine 36 and glutamate 99, lysine 41 and glutamate 223, lysine 39 and glutamate 223, lysine 136 and glutamate 133, and arginine 165 and glutamate 223. More importantly, if you see that they are not restricted to one place, they are spread all over the structure.

(Refer Slide Time: 30:21)



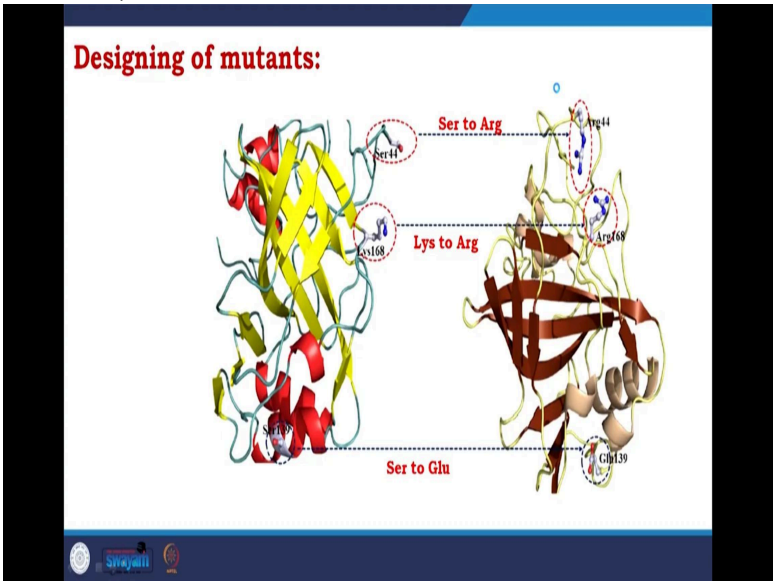
This is the representing movie; we have found by getting that MD simulation trajectory. If you carefully look at the movies, you will see that in the case of mesophilic protein, this is taken at 350 Kelvin. You will get a more pronounced effect at 400 and 500. But here also you could see that the mesophilic protein expanded whereas the thermophilic protein is more into maintaining its structure that, clearly shows the difference in structural integrity between the thermophilic and mesophilic protein.

(Refer Slide Time: 31:04)



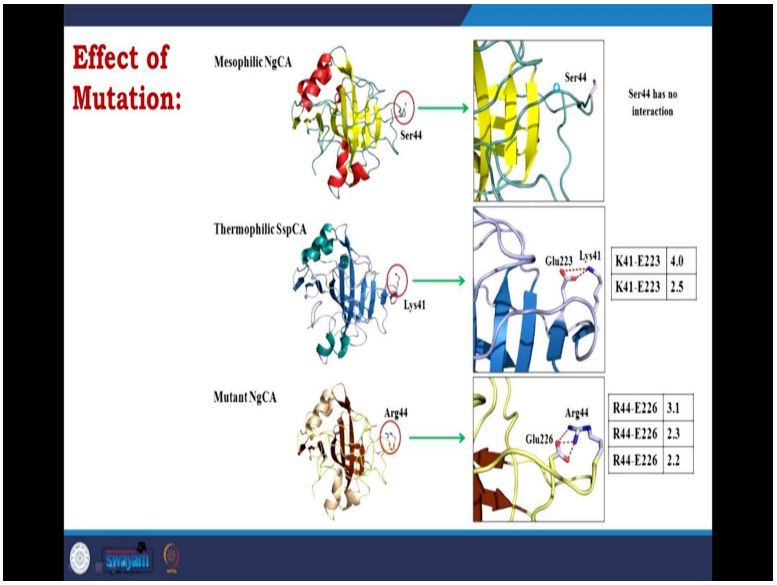
So, simulations help us understand the movement difference in these mesophilic and thermophilic carbonic anhydrases. Important non-covalent interactions like salt bridges show their effect in a time-dependent manner. Identification of flexible residues helps in designing thermostable proteins.

(Refer Slide Time: 31:27)



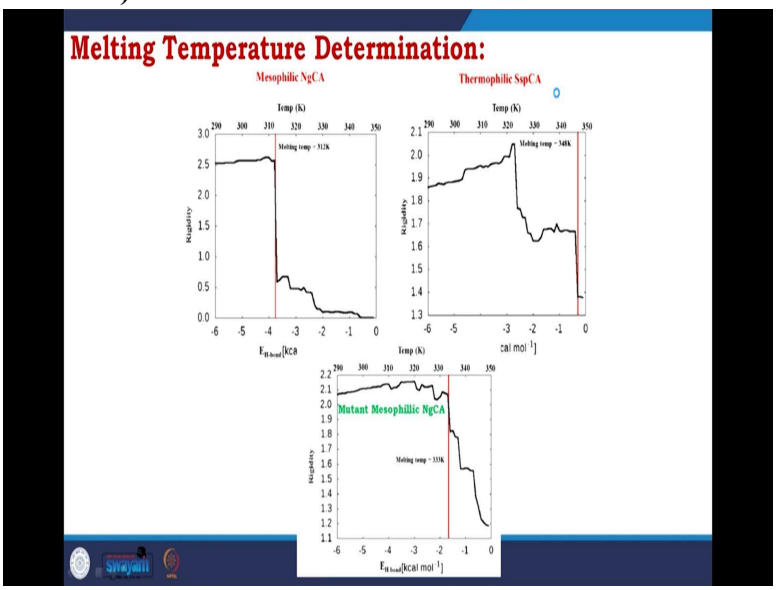
So, with all those, we started designing, and the first idea was to incorporate some salt bridges. So, 1 strategy was to take the serine 44 residue and change it to arginine, lysine 168. This is very important. If you see lysine and arginine in both are positively charged. But looking at that interaction, it seems that if we change lysine 168 to arginine. Because of its double sphere head interaction, much more stable interaction is possible. And also change serine 139 to glutamate 139.

(Refer Slide Time: 32:21)



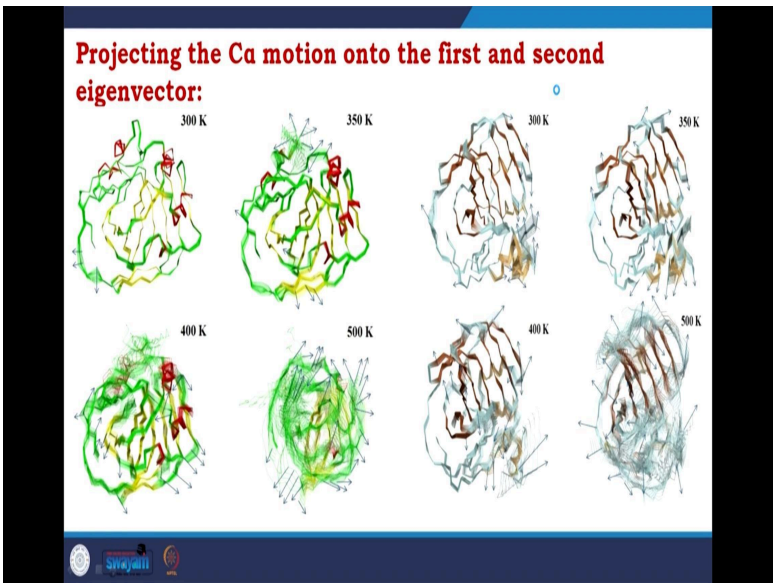
So, before going into the simulation story, I want to show you the effect of mutation on the position of serine 44. Serine 44 has low interaction in the mesophilic protein in the thermophilic protein. The equivalent position is lysine 41, which shows interaction with glutamate 223. In the mutant version of the mesophilic Ng CA, where we mutated the serine 44 to arginine 44, we observed 3 interactions of arginine 44 with the glutamate 226.

(Refer Slide Time: 33:08)



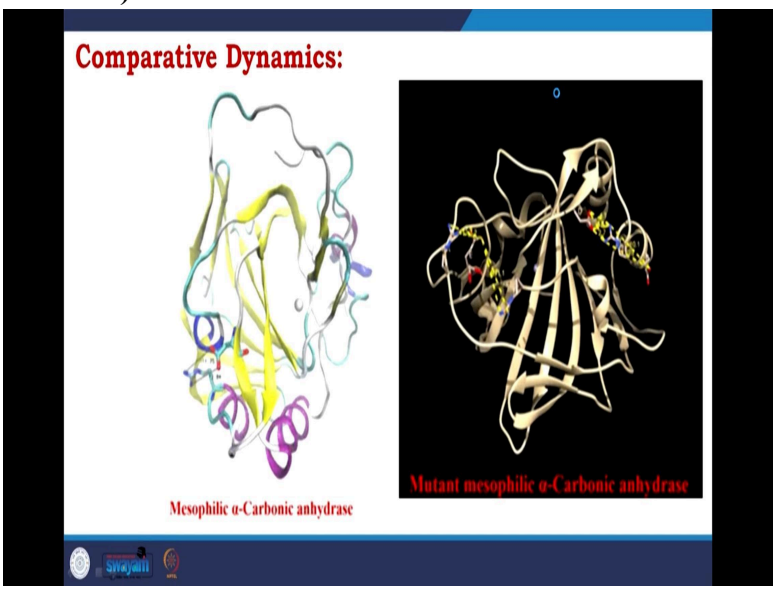
Based on our analysis, we have also determined the melting temperature. The melting temperatures show that for mesophilic protein, the melting temperature is 312 Kelvin. This is shifted to 348, which is 348 Kelvin, 36 Kelvin higher than the mesophilic protein. Only the introduction of 3 mutations change this melting temperature from 312 to 333, an increase of 21 Kelvin.

(Refer Slide Time: 33:50)



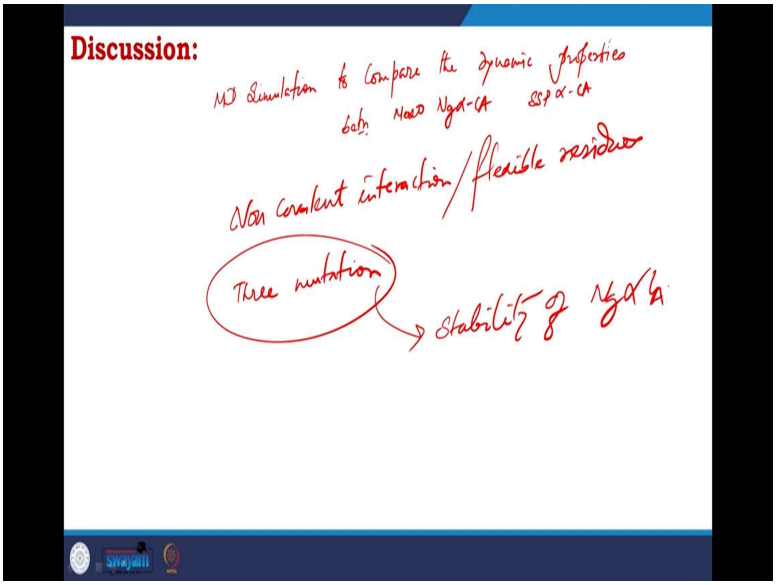
Comparing the MD simulation data of the mutant protein shows enhanced movement compared to the same protein in its non-mutated form.

(Refer Slide Time: 34:10)



Also, this is the movie comparing the mesophilic protein and the same mesophilic protein with the mutant. Again, taking a careful look, you will understand that all those loose motions present in the mesophilic protein are now absent because of the development of those 3 mutations.

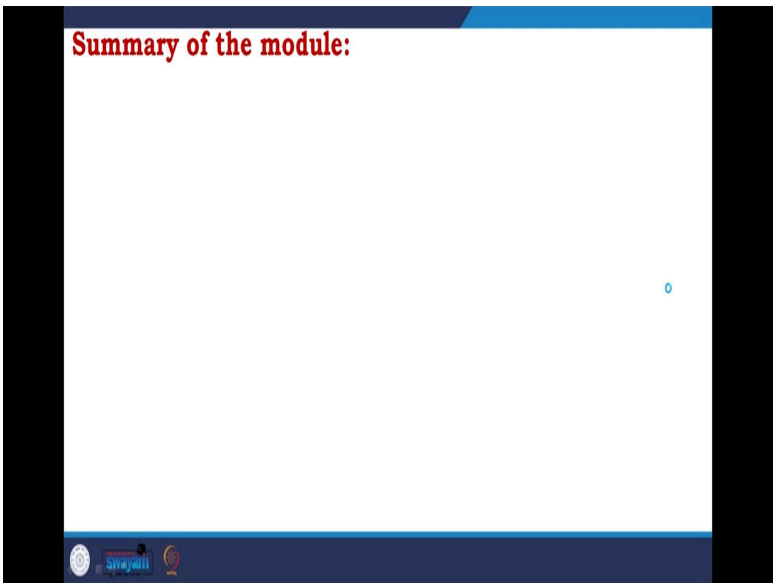
(Refer Slide Time: 34:43)



So, in summary, how did we use MD simulation to compare the dynamic properties between the mesophilic Ng alpha CA with the thermophilic Ssp alpha CA that we looked at? We looked at mostly non-covalent interactions looked at flexible residues, and based on our understanding, we made 3 mutations. These 3 mutations change the stability of the Ng alpha CA and if you look at the MD simulation method allows you to look at your protein which is impossible with any other techniques.

So, that is one of the beauties here; you could look at the dynamic property. You get into another angle to analyze your protein. You could look at the intrinsic flexibility in your protein and all this knowledge, especially if you have ligands not shown here in this case study. You have ligand how it is bound and stabilizing, and how you could alter all these would be critical towards enzyme engineering, drug designing, and whatnot.

(Refer Slide Time: 37:04)



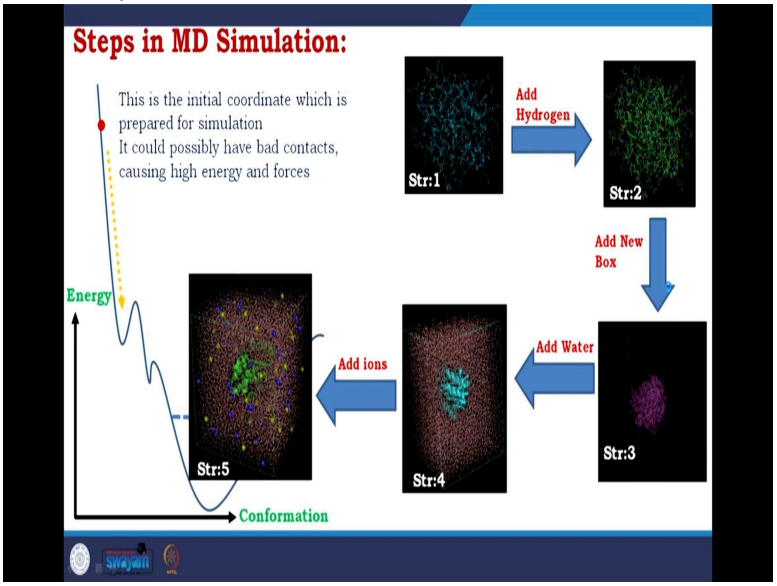
So, we are on the edge of finishing this module, and I agree this is the introductory course. So, the things we have discussed would not make you an expert in MD simulation. But I have tried my best to discuss the basics of the process and the components. I start with very basic questions. Why and what is MD? What is simulation? How does simulation mimic experiments? What type of experimental system it mimics, and how it proceeds further? How the timescales are there? What type of physics is used?

How is the force field? How could you influence a protein by developing a force field competent? What are the components? How could they be used through an algorithm like balded velocity balded and lift the fog and all those algorithms by allowing Newton's law of physics? And then how, from a sequence, you start your journey to get a structure, repair those structures, modify those structures come to the simulation process.

They knew again to make things like you add hydrogens, you add a box, you add ions, and you make a system experimental like mimicking of an experiment. Then you go for minimization energy minimization to make the bad contacts better. And then go for equilibration with the ensemble conditions and, finally, the MD simulation run. I have also discussed several analysis methods which will be helpful for you to perform, and then I have discussed case studies.

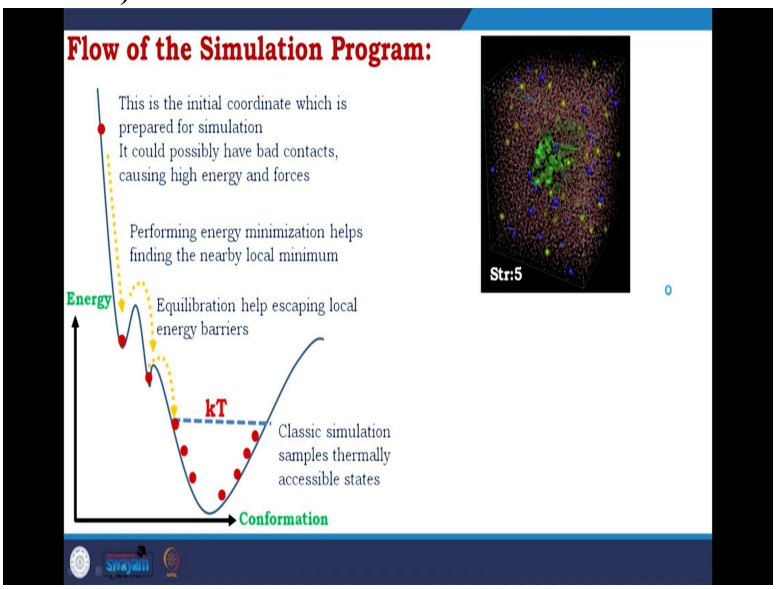
I will finish up just by mentioning that in this introductory course, in this limited opportunity, what I have discussed, we call classic molecular dynamics. But the field of MD simulation in protein is not restricted to classic molecular dynamics. Let us discuss that.

(Refer Slide Time: 40:00)



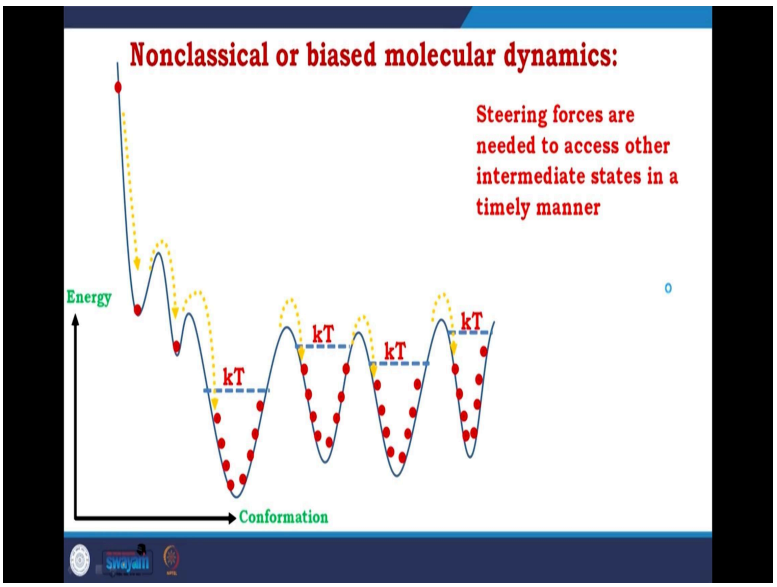
Do you remember this I talked about? We have the initial coordinate, which we just prepared for simulation, then we add hydrogen, then a new box, then we solve it by adding water, then we add ions, and we get the final structure from there. Should we possibly have bad contacts? So, we have to do minimization.

(Refer Slide Time: 40:25)



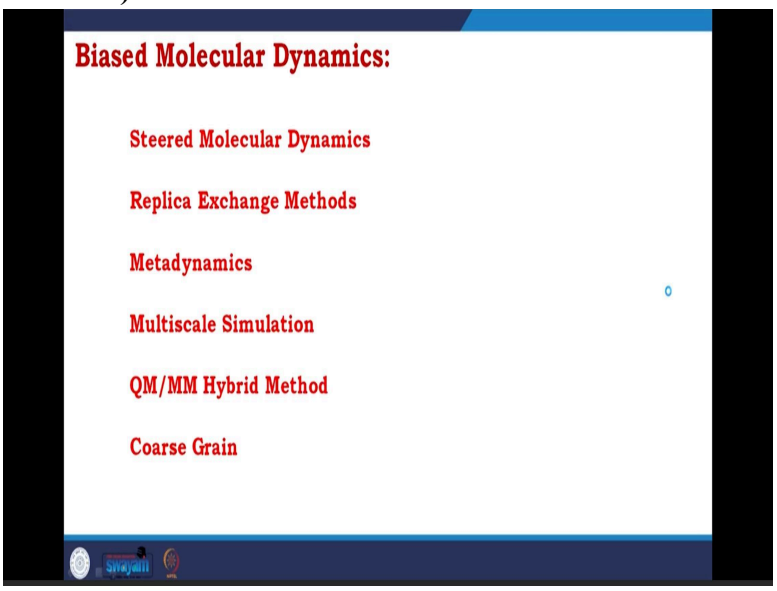
So, we start with the arginine and go for minimization energy minimization, which helps find the nearby local minimum. But then, from the local minima, equilibrium helps us escape the local energy barriers, which is a small barrier, then the classic MD run helps us to come to the global ones, and it is a sample of thermally accessible states.

(Refer Slide Time: 41:08)



Now, from there, it potentially looks like this is the end of the scope of the simulation, but it is not true. Whatever we do in MD simulation, we compare the thermal states. But we could go beyond that by applying the steering forces needed to access other intermediate states promptly. So, if you want to study confirmations in different states, if you see these are energy barriers.

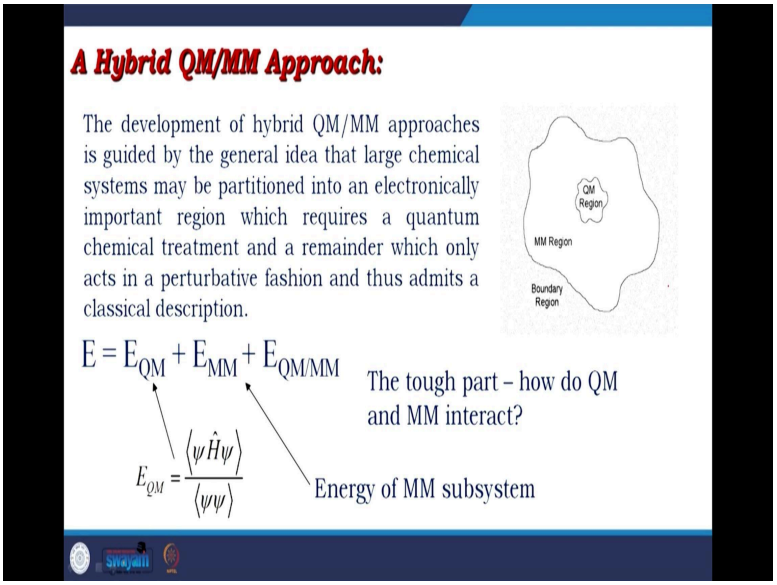
(Refer Slide Time: 41:58)



We could have to perform something called biased or targeted molecular dynamics. Nowadays, very popular methods are coming, as I told in the scope of this course. I cannot go beyond, but I will talk about them. So, interested people could learn steered molecular dynamics, which is critical, especially to understanding the interaction between macromolecules and small molecules.

Replica exchange method: by resampling samples from different temperatures, metadynamics, and multiscale simulation as we think that if we start from 1 point, it is not easy to go to the next conformer. So, what do you do? We run a simulation, or we get different conformers and apply simulations to all of them as multiscale simulation, quantum mechanical, molecular mechanical hybrid method, and coarse grain if you do not need them on the atomic level. If you want to look at the movements of bigger structures, a whole loop, a whole structure, then the coarse grain is your method.

(Refer Slide Time: 43:36)



I would finish by talking a bit about one of my favorite methods, the hybrid QMMM approach. As you already understood, by applying Newtonian physics, you cannot perform bond breaking and making when you are studying an enzyme. It is critical to look at its conformation during the chemical reaction. But when studying the breaking and making of bonds, you must make quantum mechanical calculations.

Is it possible to perform the quantum mechanical calculation of the entire protein? No, because if you think, even if you consider a small protein with 200 amino acids, it would have several electrons in such a high number with their cross-interaction that even the best supercomputer the world has would take innumerable amounts of time. To solve this problem, scientists have devised a hybrid method called the hybrid quantum mechanical molecular mechanical approach.

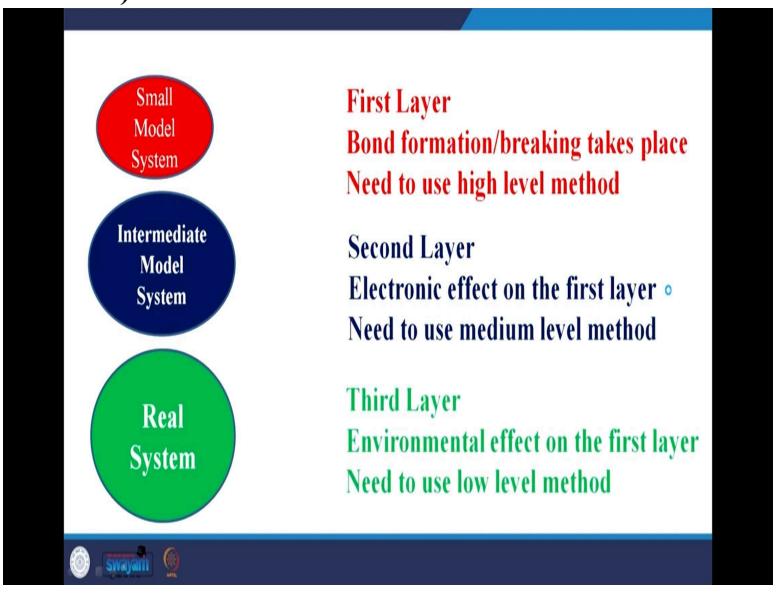
The development of this approach is guided by the general idea that large chemical systems may be partitioned into an electronically important region, which requires a quantum

chemical treatment and a reminder that only acts in a perturbative fashion and thus admits a classical description. So, if you look here, the whole system is treated with the molecular mechanical approach, and the small portion is treated with a quantum mechanical approach.

Let us say now you are talking about an enzyme. The amino acid side chains of the active site and the substrate would be in the quantum mechanical region. The rest of the protein shows movement in the molecular mechanical region. So if you calculate the energy, the energy now has 3 components energy of the quantum mechanical region, the energy of the molecular mechanical region, and the energy of the quantum mechanical molecular mechanical region.

The energy of the quantum mechanical region will be calculated using the Hamiltonian operator, and you already know how to calculate the energy of the molecular mechanical subsystem. The top part is how to calculate the E QMMM portion because you have to consider the interaction between the QM and MM region. There are several approaches, considering a link atom considering cost orbital. And all of them, again, not going into, but I am just showing you the possibility and the successful implementation of a calculation of a biochemical reaction.

(Refer Slide Time: 47:17)



So, there would be 3 layers and 1 small model system. In the first layer, bond formation breaking takes place, requiring a high-level method of quantum calculation. The intermediate model system's second layer, the electronic effect on the first layer, needs to use a medium-level method, and the real system, the third layer environmental effect on the first layer, needs to use a low-level method.

So, like these, we have to go into detail about all the processes, and knowing about them and understanding them would help us with more options to develop a more realistic model and the calculation of biological macromolecules. Thank you very much. I am finishing this module here. Thanks for listening. Thank you.