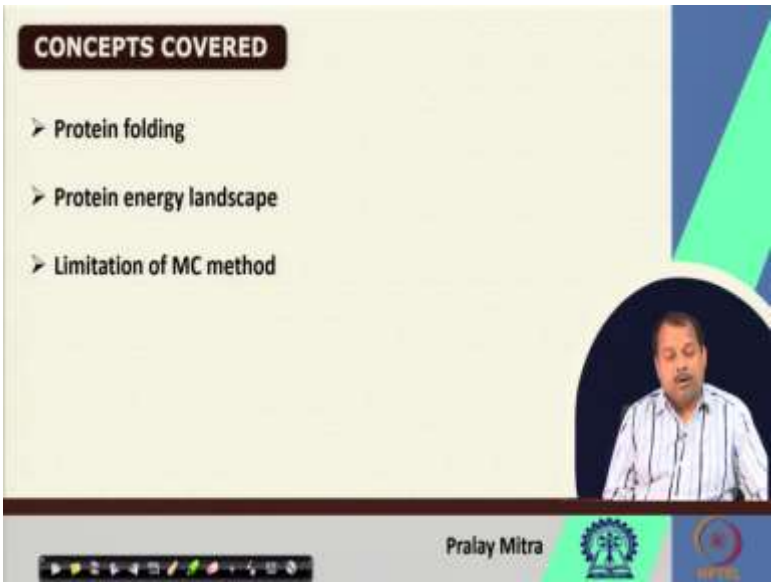
Algorithms for Protein Modelling and Engineering
Professor Pralay Mitra
Department of Computer Science and Engineering
Indian Institute of Technology, Kharagpur
Lecture: 22
Protein Energy Landscape (Contd.), Limitation of MC

Welcome back. So, we started to discuss regarding the protein energy landscape. And at the beginning started with the protein folding problem I mentioned the forces which are responsible probably for the protein folding and during the process even in vivo, there is a possibility that the protein may assume some wrong folate state.

Wrong in the sense that corresponding to each protein sequence it is supposed to assume one particular structure and there is a one to one correspondence between the structure and the function. But it may happen that particular structure is not assume because of some reason during the protein folding process in cell and if it is not then what will happen that it may lead to some misfolded state which may be say may create some disease or some allergy, it may be toxic to our body or our immune system can able to identify them and throw it away.

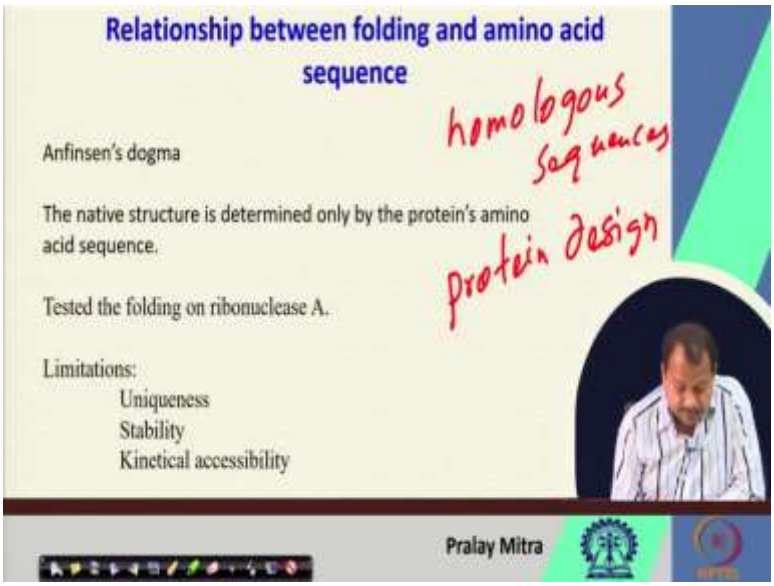
Whatever it may be when one protein sequence assumes to take one particular folate state then we are definitely missing its functionality in our body. So, whatever function it is. Now, that thing we concluded on the last slide regarding two different folding models that exist one is the belief is that diffusion coallision model another is the nucleus and condensation model.

(Refer Slide Time: 01:51)



Now, here, we will continue with the protein energy landscape. And we will start discussing regarding what is the limitation of the Monte Carlo technique in the context of this protein energy landscape.

(Refer Slide Time: 2:05)



So, the relationship between folding and the amino acid sequence was observed long back by Anfinsen's. So, it is now known Anfinsen's dogma which says that the native structure is determined only by the proteins amino acid sequence and he tested his hypothesis on the folding of ribonuclease A and got the Nobel Prize for this. However, it is long back However, some limitations of his observations are noted what he mentioned if there is only the amino acid sequence which will tell.

What will be the native structure then as if there is no effect of the environment say regarding the features or the driving forces that we mentioned that what is the solvent whether it is a water or lipid by layer based upon that one there will be some changes in the folding state what will be the pH what will be the concentration of the salt in the solvent what will be the same temperature.

So, based upon that one, the folder structure may change. So, one thing one example you can assume that the protein which say is assuming some stable state at room temperature which you can consider as the 27 degrees Celsius, if I keep on increasing the temperature say if it goes beyond 100 degree 200 degree weather it will assume the same structure no it is not possible. So, there are some limitations that it is not unique.

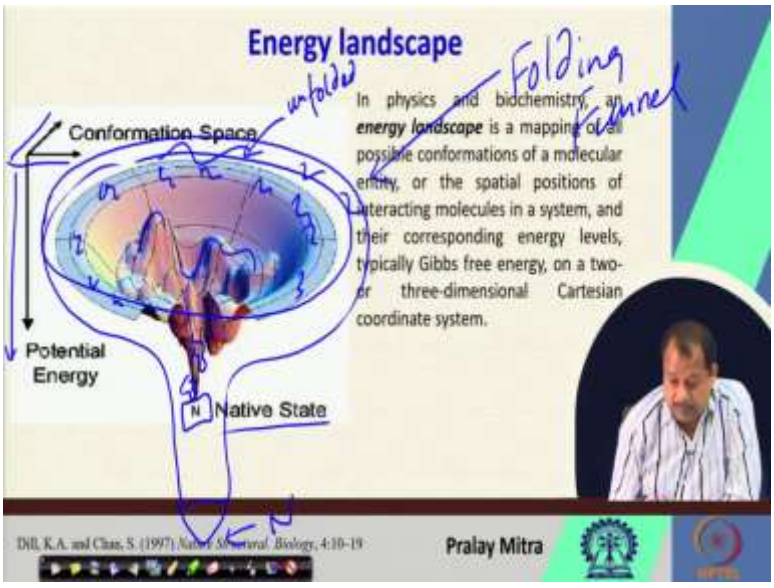
So, one sequence can assume one structure if the environment changes by environment I used to mention that solvent say pH of the solvent, then temperature then the concentration of the salt in that solvent if it changes then there will be some changes in the structure and accordingly there may be some changes on the function. So, it is not unique if I take the opposite way, I mean that.

If I consider one structure whether corresponding to that structure there is only one sequence exists no we defined protein design problem where I mentioned that corresponding to one protein structure there may exist more than one sequences. If more than one sequences exist, then you can consider that corresponding to multiple sequences, you can map to one particular structure.

Again, there exist multiple sequences who are very similar, but may assume to the same structure, but they are not same, those sequences are generally called as the homologous another is the protein design problem, which says that corresponding to one protein structure there can be multiple protein sequences who are different.

So, the stability may change again based upon the environment and kinetical accessibility. So, those things were not considered by that Anfinsen's. So, that is why some limitations we observed nevertheless, it was long back and one of the classical work in this particular area

(Refer Slide Time: 05:53)



kandil in 1997 published one very good paper in nature structural biology. So, taking the (())(06:06) from there, he mentioned or give one concept that in physics and biochemistry and

energy landscape is a mapping of all possible confirmations of a molecular entity or the special positions of interacting molecules in a system and their corresponding energy levels typically.

Gibbs free energy on a two or three dimensional Cartesian coordinate system and also he gave this diagram. So, this diagram I taken from his paper, where you can see that the conformation space was mentioned here, this is my conformation space and potential energy is in another dimension.

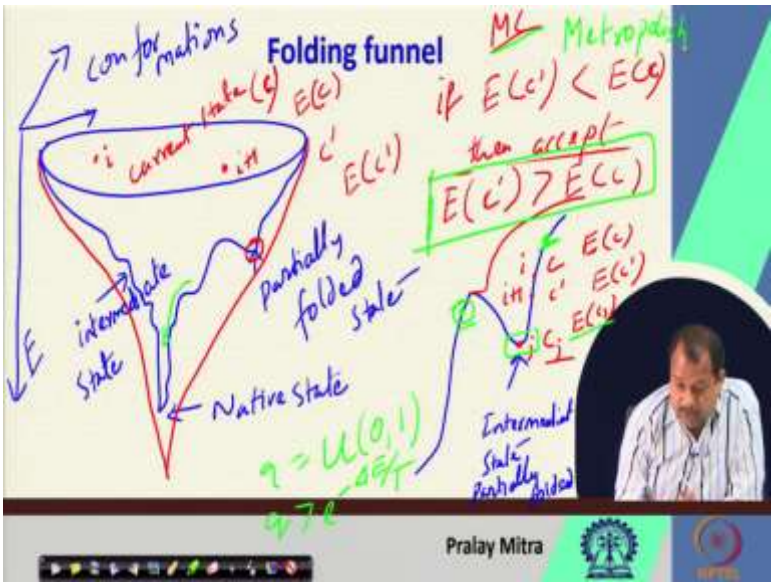
So, we can assume that corresponding to one protein sequence you have generated all the possibilities and say God is there corresponding to each possibility God computed its energy function exactly what it is then if you plot them, then you will see something like this something like this means that if the potential energy decreases this way.

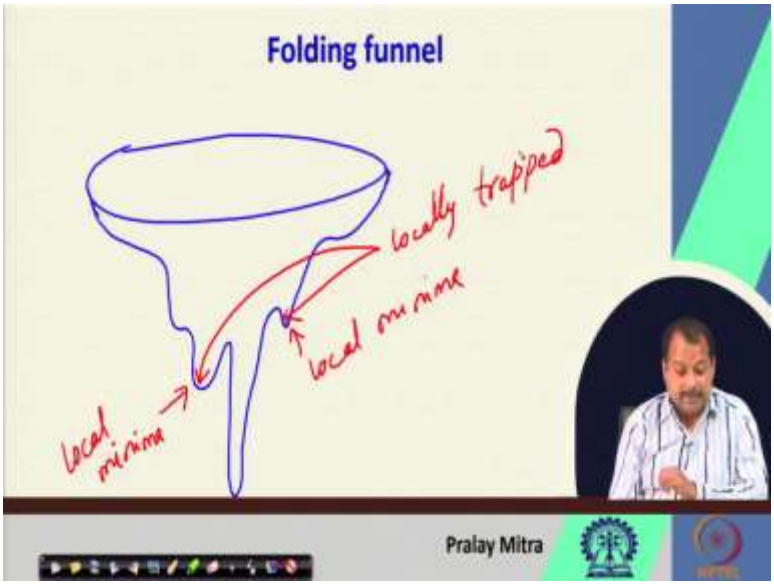
Then it says that, I am going to the native state and that way you will see that all the surface if I say this is my surface or conformation space. So, here you will see most of the unfolded or loosely folded or just folding started this kind of situation structure. So, here now, if you go inside, then it is taking more or less the correct shape or structure and finally, he will go to this end which is my native state.

Now, this way if you go and again I am assuming all the partially folded or intermediate say starting from the protein sequence which is unfolded to the native state when I am moving I am generating all the possibilities when I am generating all the possibilities definitely they are following Ramachandran plot, which means that they are following that phi psi angle of the Ramachandran plot.

If I assume that one then I will see there are some reasons which are forbidden means I cannot generate the confirmation that way some hilly regions are there inside but overall it is looking like a funnel it is looking like a funnel. So, that is why it is also called as the folding funnel where these are my partially folded states are all are and finally when I will reach to this tip that is my end that is my native state. So, this is the hypothesis given by at the time.

(Refer Slide Time: 09:36)





Next, if I take that particular energy landscape and project on a 2-D then you agree with me that it will be looking like something is here this is my potential energy and this are conformations perfect. This is my native state. This you can consider as partially folded state or you can consider this as intermediate.

So, this is my folding funnel. So, this intermediate state and this is my native state. Now, the first problem you should note will say we are running Monte Carlo simulation technique and why

Monte Carlo simulation was not enough, but I have to go for Metropolis criteria. So, that I am going to demonstrate fast.

So, let us assume that you are started with some situation and it is your current situation that is my i, i instance if you remember the Monte Carlo algorithm that we have discussed corresponding to this i there this i I iteration and current state is C. So, corresponding to this c there is energy function E say E of C.

Now, it is a Monte Carlo simulation, so, randomly I am generating another move that move let us assume is here. So, this move say i plus 1 and for that my conformation is C prime. So, energy will be E C prime. Now, theoretically if E C prime less than E C then accept that should be the situation theoretically it is true if the surface is this folding funnel is something like this which means it is gradually decreasing.

So, that whenever you are generating some C prime then the C prime will be either over C or below C then it is fine if it is the situation this rate contour is my the folding funnel then I can go by this and that is my Monte Carlo simulation and I am perfectly fine with it. Unfortunately, it is not the case.

The case is here there is an intermediate state or partially folded state if this is my partially folded state then what will be the situation. Let me blow it up and draw here. So, I am having some state here and then it is going down. So, this I am calling this region as partially folded or say intermediate state.

If it is then what is my problem you see say this is my i, if this is my i then I will get C here E of C here. Now, I generate something here say i plus 1 C prime E c prime. Now, if it is E C prime then say that way I will reach to this position. Let us assume I have reached there and say that is my j. So, Cj and E Cj is my energy after this one.

If I make another random move, and the random move is always in this region then always the next move will have an energy function which is higher compared to Cj that way I will falsely assume that I have reached to the native state or optimal state or the minimum state falsely I will assume that one but that is not correct here.

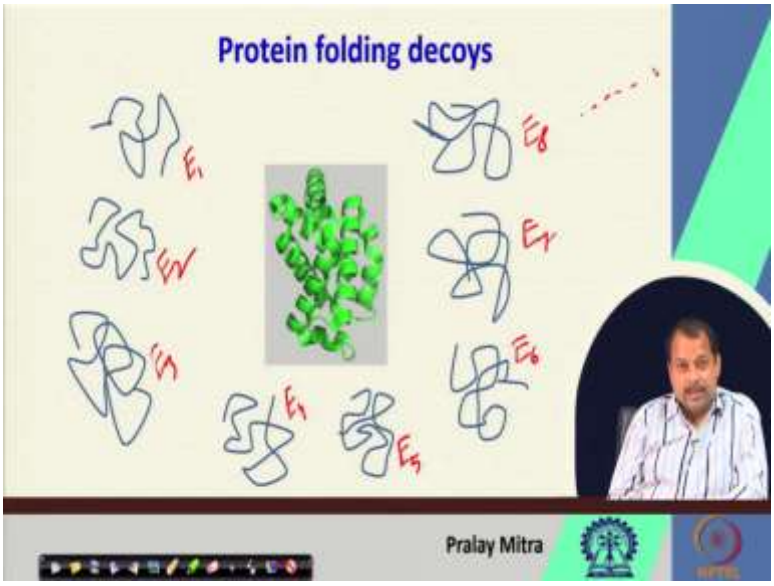
Because of this situation because of this region I am kind of trapped and falsely I am assuming that I have reached to the minimum However, it is not. So, what will be my strategy here that when I will reach to some region then always not always I will accept when it is say this, but I will also accept will say is C prime is greater than E C sometimes I will accept this one also.

Sometimes means, how many times for that I have generated one random number q equals to U 0 1 if q is greater than it q e minus delta E by T where T is the Boltzmann temperature and delta is the difference between these two energy then I will accept otherwise I will not accept that way I am ensuring that even if I will reach at this position then I may take some region here as accepted.

And if I accept that one then my energy is not E Cj now, it will be here next where I will have another move then there is a probability that I will pick somebody here then it is with the minimum energy this is minimum then this one So, I will pick this one that will keep on moving and I will go at this region is this clear. So, that is the reason why it is not MC only but Metro Polish criteria is also required for my purpose.

Now, if I erase everything. And say draw that folding funnel only then this region is called local minima. This is also called as the local minima and if I reached here or say here then that situation I will call us locally trapped. So, I am trapped locally for that I have to go for Metro Polish criteria. But sometimes that will not be enough also that I will come later

(Refer Slide Time: 19:08)

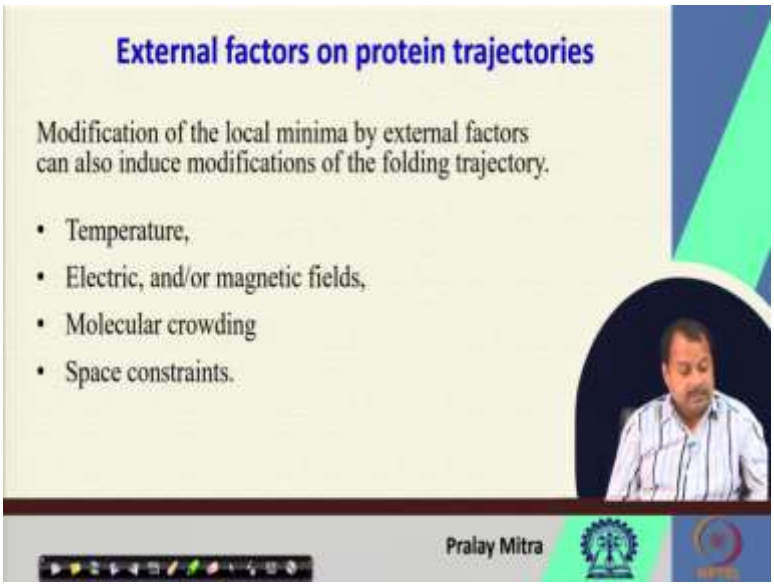


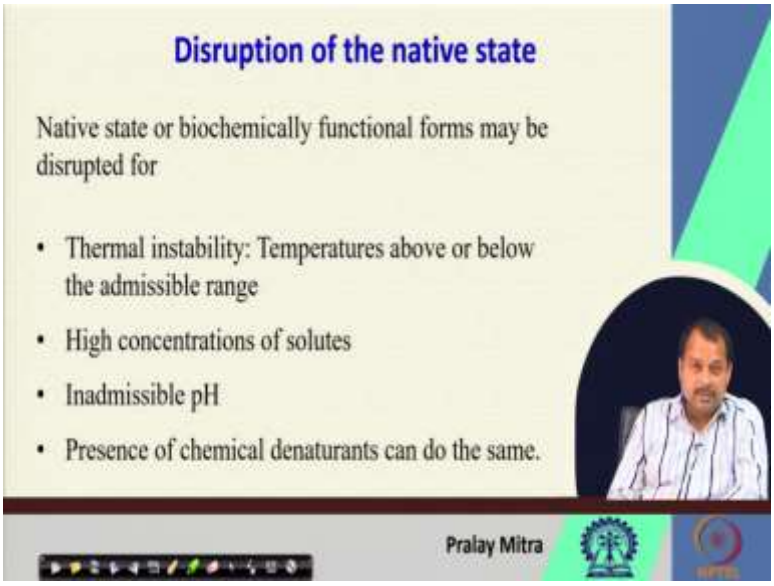
When we are working with a compressional technique, then I have a number of intermediate state in computation I can call them as partially folded state or say the coils. So, such thing we already discussed in the context of protein complex modeling in case of protein folding also there can be some decoys because like these because like these. Now these decoys will have one individual energy say if without any loss of generality, I can assume that.

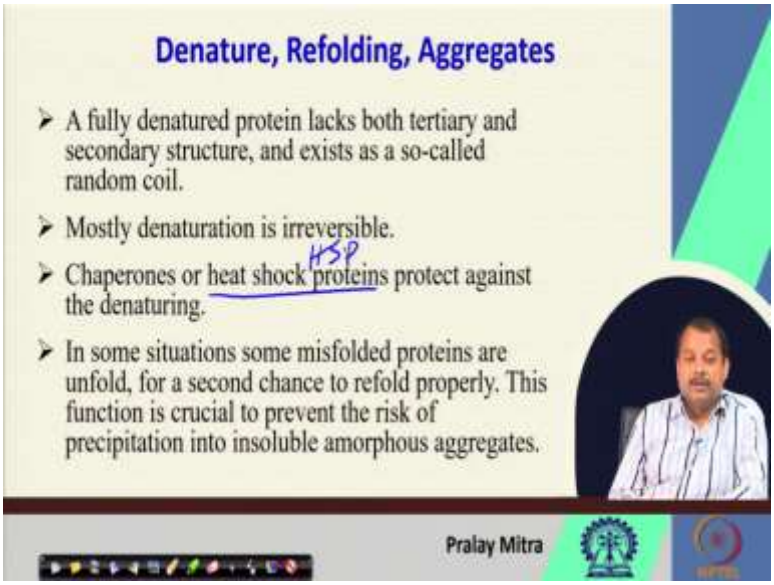
So this is a E1, E2, E3, E4, E5, E6, E7, E8 and dot dot dot dot. I do not know in the energy landscape or folding funnel where they decides and also I mentioned about one situation where it will go to some amyloid fibrils and I mentioned that if it reaches there, then it will not able to come back. So, that kind of situation is something when my folding funnel will be like this.

If I assume some folding funnel like this, then this can be one native state and this will be amoloyd fibrils. The amoloyd fibrils are looking like this mostly they are with the beta seed they have patterned like beta seed and when you reach there then you can see it seems the energy state is very less energy state is very less even compared to the native state then from here it is not possible to come back here. So, we have to be careful about that one also.

(Refer Slide Time: 22:18)







Now, the external factors on protein trajectory. So, taking some of the examples or cases from the biological knowledge that we discussed on the last lecture, we can say modification of local minima by external forces can also induce modification of the folding trajectories temperature, electric and magnetic fields, molecular crowding, space constraints, those can happen.

Now, the disruption of the native state. So, in this context, it is better to discuss that point also. So, the native state or biologically functional forms may be disrupted for thermal instability that I mentioned that say one protein is folded and having some function in the room temperature. Now, I am taking that protein and keep on heating I mean, I keep on increasing its temperature then after say 100 or 200 degree etc, then some of its born because of the thermal agitation will break and then it will start unfold.

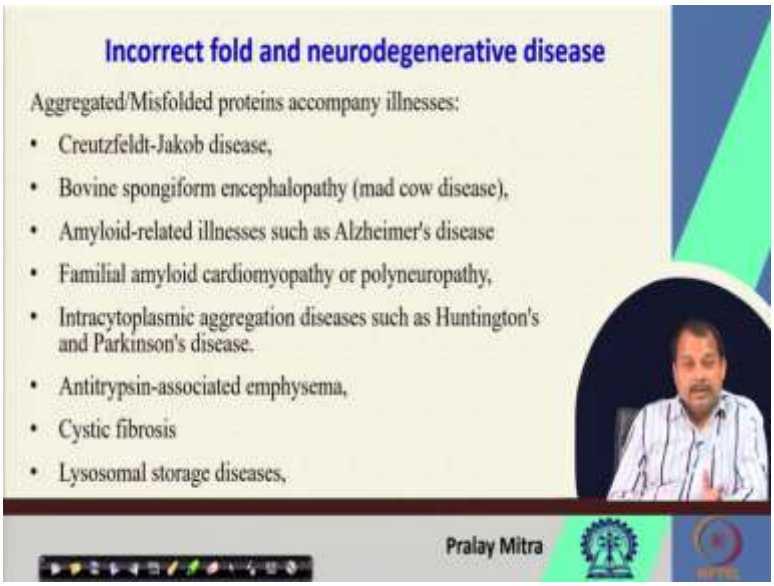
So, because of that one some unfolding may happen high concentration of the solute if I keep on changing the concentration of the solute, then also the it may unfold or the native state disrupt inadmissible pH. So, because of say high concentration of say charged high concentration of say acidic if on the surface there are some basic amino acids.

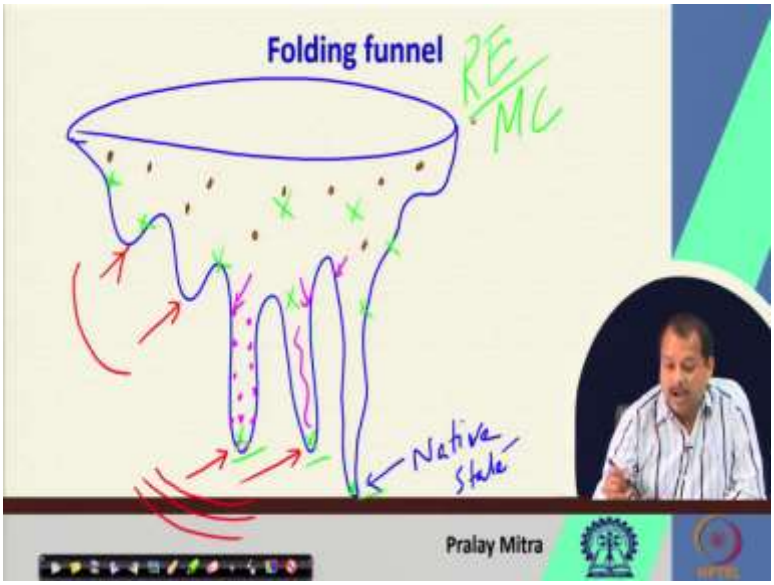
Then they will try to go for the ionic interaction with the salt on the environment compared to its own interaction for the protein. So, that may disrupt our native state presence of chemical denaturants can do the same regarding this chemical denaturants. So, most commonly used one is the urea. So, if you add urea to protein mostly it will denature or it will disrupt our native state, but you have to remember that not all disruption is reversible in nature.

So, mostly if the native state is disrupted, then it will not fold or it will not take the previous and native state. Denature, refolding and aggregates. So, a fully denatured protein lacks both tertiary and secondary structure and exist is so called random coil. Mostly denaturation is irreversible that I have mention just now. Chaperones or hits of proteins in salt this is a very famous protein this is called as the HSP proteins protect against that denaturing.

So, how interesting I mentioned on the last lecture also in some situations, some misfolded proteins in some situations, some misfolded proteins are unfold for a second chance to refold properly this function is crucial to prevent the risk of precipitation into insoluble amorphous aggregates. But, if given the second chance to refold properly also fails then perhaps it will go to some precipitation or insoluble amorphous aggregates which may lead to some disease condition mostly that neurodegenerative diseases kind of thing.

(Refer Slide Time: 26:11)





Incorrect fold and neurodegenerative disease aggregated are misfolded proteins accompany illness. So, these are the least of the illness these are not only they are maybe several others. So, researchers are working on this one.

Say the, Jakob disease. Bovine spongiform encephalothy mad cow disease, then amyloid related illnesses such as alzheimer disease, familial amyloid cardiomyopathy or poly neuropathy, intra cytoplasmic aggregation disease such as Huntington's and Parkinson's disease at antitrypsin in associated emphysema, cystic fibrosis lysosomal storage diseases.

So, these kinds of situation may happen if a protein is not correctly folded or it assumes some incorrect folding state again, if it goes for incorrect folding, then it gets mostly another chance to open and refold but even if it fails, then it will lead to some incorrectly folded state and as I mentioned, it is not possible for the immune system to identify all the miss correct or miss folded or incorrectly folded state if it cannot identify that one then that will definitely lead to some disease condition something like this.

Now, regarding the folding funnel, I mentioned but few remains. So, let us assume that this is my native state make it simple now, do not incorporate the mechanical control or amyloid situation this is my native state. Now, even then also you will find that so many local minima are there. So, these two local minima may not be very few severe, but these are very severe because they are close to native state.

Once I will reach there then it will problem. So, only Monte Carlo simulation may not give me the correct solution or suggestion for a protein whose folding funnel may look like this. If that is the situation then what shall I do? And before that one perhaps you noted during the Monte Carlo simulation technique. So, if say let me pick the color if my situation is here, then I can randomly move that is fine. But say it is Monte Carlo simulation.

So, I am starting with some random state. But if my random state starts from here sipping color and it always generates states within this region only then again miss correct incorrectly you will lead to one situation which is called as the locally trapped situation.

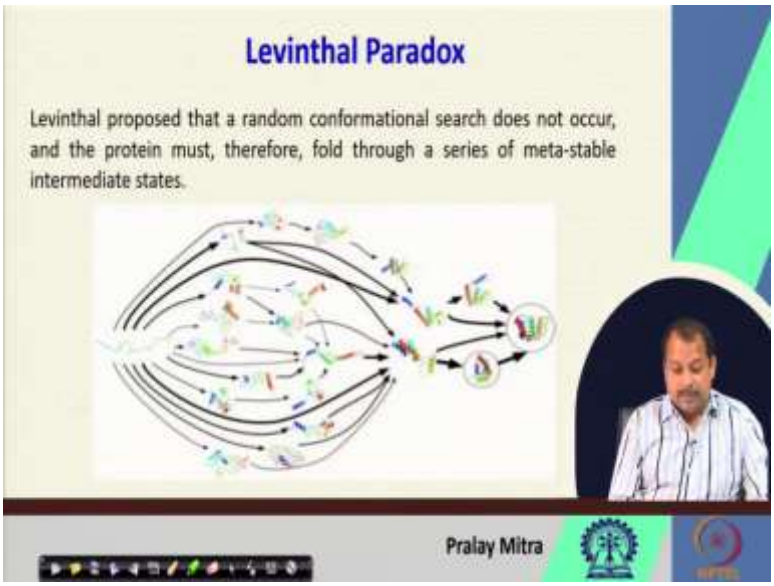
So, we will be trapped locally in this locality, you will not able to reach to the native state. If it is the situation, then what to do but problem probably you have identified that starts with the initial seed or random. So, it is here now it can be here also if it is here then also it will be the same thing or it can be here also if it is here then it will bingo.

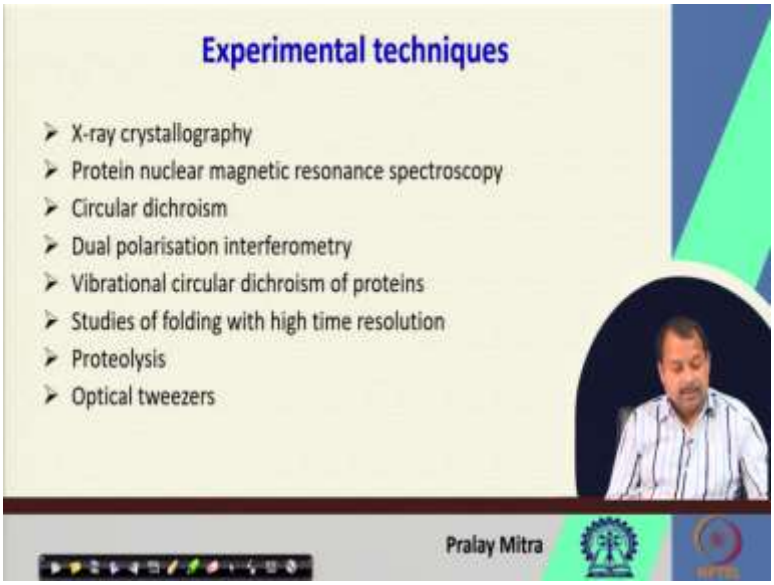
I will get that native state by minimizing that one the problem in this case is that you have the correct energy function assuming the perfect energy function then also what is your seed or from where you are starting based upon that to a lot of things may change. So, starting with one seed may not be enough if not then what.

It if I start with say multiple seed values with a hope that if all the seed values say random seed values are random in nature which means that there is no correlation between the initial seed value also then there is a probability that once a seed value let me take the green will be here another here another here another here another here another here another here another here that way if I go and during the process or iterations of simulation of the Monte Carlo simulation if it is then in during the simulation process if I combine the results or compare the results then I can able to understand.

So, who is here can reach up to here who is here can reach up to here who is here can reach up to here others can go somewhere else. Now comparing this this and this I can able to understand that what is the optimum energy? So, based upon that one, another extension of our Monte Carlo match with Metropolis criteria method will be modified that is called as the replica exchange or REMC that we will discuss in the next lecture.

(Refer Slide Time: 32:02)





So, now, let me conclude this with the experimental techniques, which are there to determine the structure of the protein it is x-raycrystallography protein nuclear magnetic resonance spectroscopy, circular dichroism. Dual polarisation interferometry, vibrational circular dichroism of protein, studies of folding with high time resolution. Proteolysis, optical tweezers, et cetera. Now, you should remember that these techniques are not required for our courts purpose.

But I mentioned this one so, that since you are working with the protein, so, you should know at least that what are the different techniques, experimental techniques are there or when you talk to a biologist or exponent release regarding the development of the algorithm or to understand their

problem, then you should know that these techniques exist, but we are not going to discuss these techniques for our course. Okay, that is it. Thank you very much.