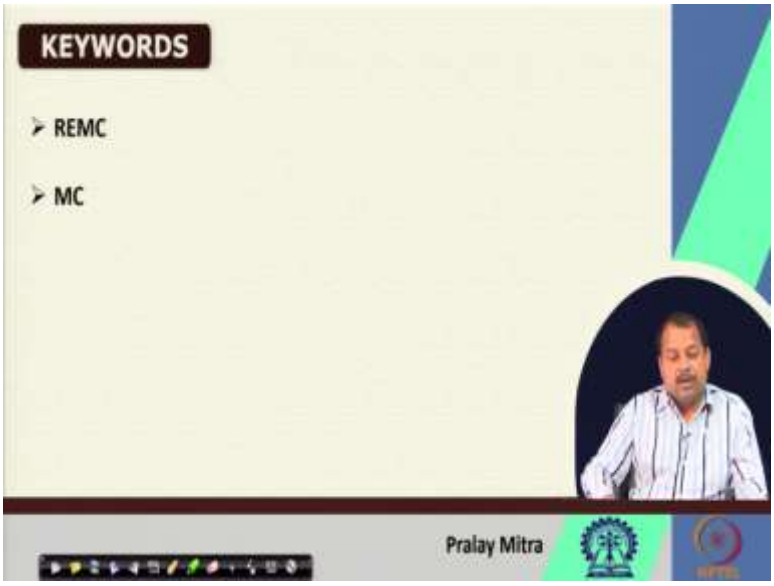
Algorithms for Protein Modelling and Engineering
Professor Pralay Mitra
Department of Computer Science and Engineering
Indian Institute of Technology, Kharagpur
Lecture: 23
Replica Exchange Monte Carlo (REMC)

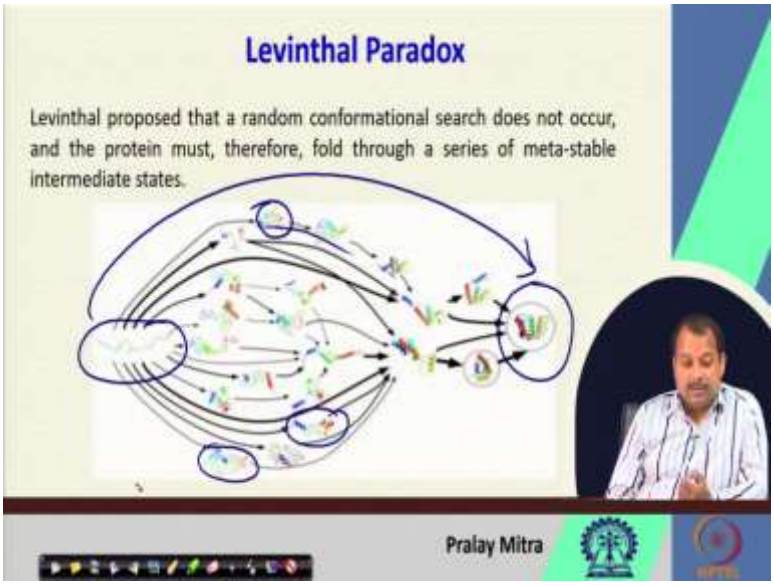
Welcome back. So, we already started to discuss about this protein energy landscape. And why we started the motivation I introduced on the last lecture that it indicates that in the context of the protein. So, when we are discussing the energy function protein folding funnel, then we noted that protein folding funnel suggested that Monte Carlo simulation method even with the metropolis criteria is not enough for protein folding problem, there are some problems and the problem is trapping locally which appears to be the minimum energy state to the Monte Carlo simulation, but not the minimum energy state.

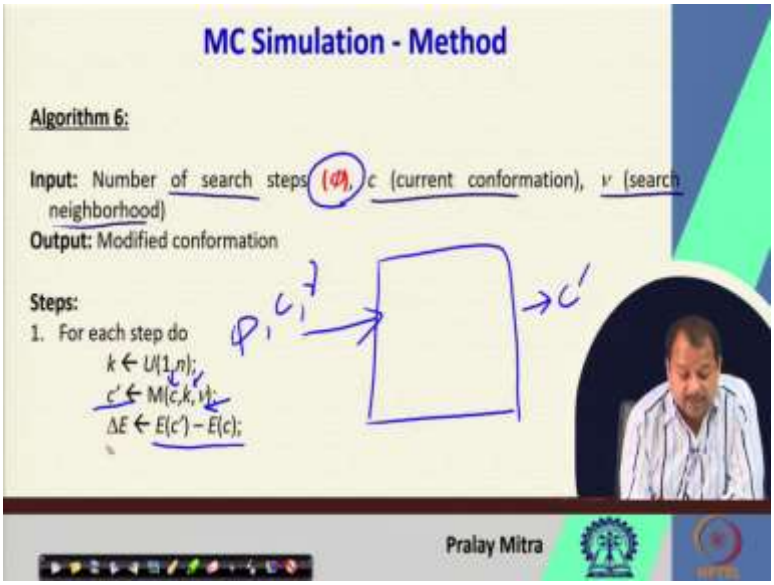
(Refer Slide Time: 01:11)

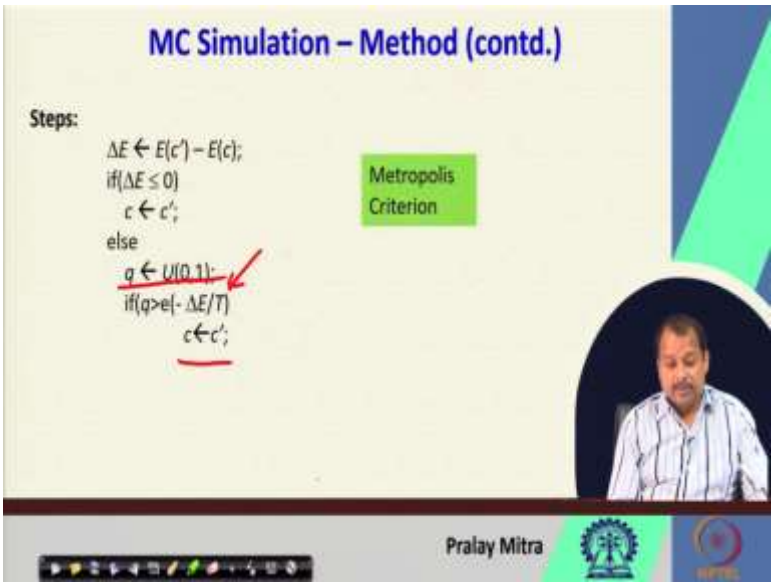


So, for that, we are planning to move on to one technique which is called as the replica exchange Monte Carlo technique. So, insert it is called as the REMC that we are going to discuss today.

(Refer Slide Time: 1:24)







Now, the basic motivation perhaps, for these Monte Carlo simulation technique or say replica exchange Monte Carlo simulation technique is that the existence of the Levinthal paradox. Which says that Levinthal proposed that a random confirmation search does not occur and the protein must therefore, fold through a series of meta stable intermediate state.

Borrowing this figure I can say that starting from this random cold situation and finally reaching to this native state it passes through a number of intermediate state. So, when it assumes this state then from here it cannot be reached here. So, that can also be acted as a motivation and based upon that one we are designing this replica exchange Monte Carlo but before that just let me recap that algorithm 6 that is our Monte Carlo simulation method.

Where number of steps input was number of search steps, how many number of steps I will search, then current conformation then search neighborhood those three was input at the time, the number of search steps was there, but now, for integrating this MC simulation along with the replica exchange part.

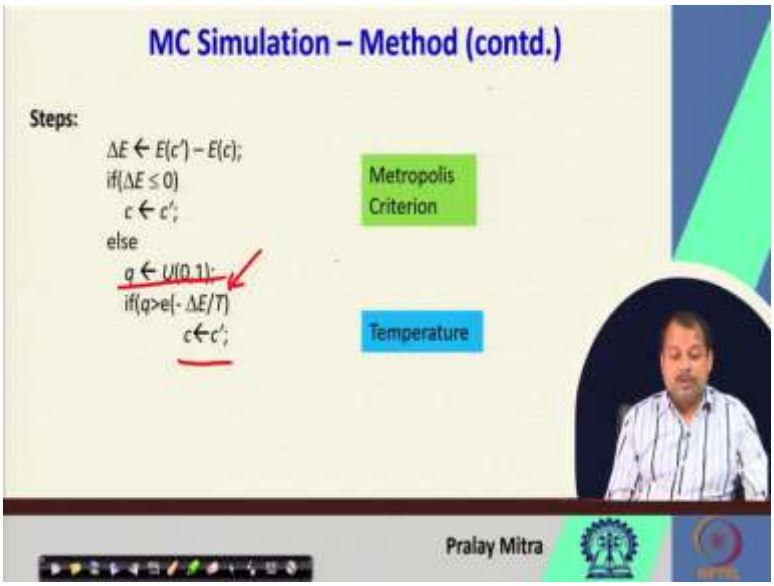
So, I am incorporating phi as the number of search steps. So, that way so, if I call that my Monte Carlo simulation is a box blue box here you can consider the input is phi current state and new and output is C prime a modified conformation. Now, the algorithm still same for each step do k first I am generating one random state and that random state this neighborhood information.

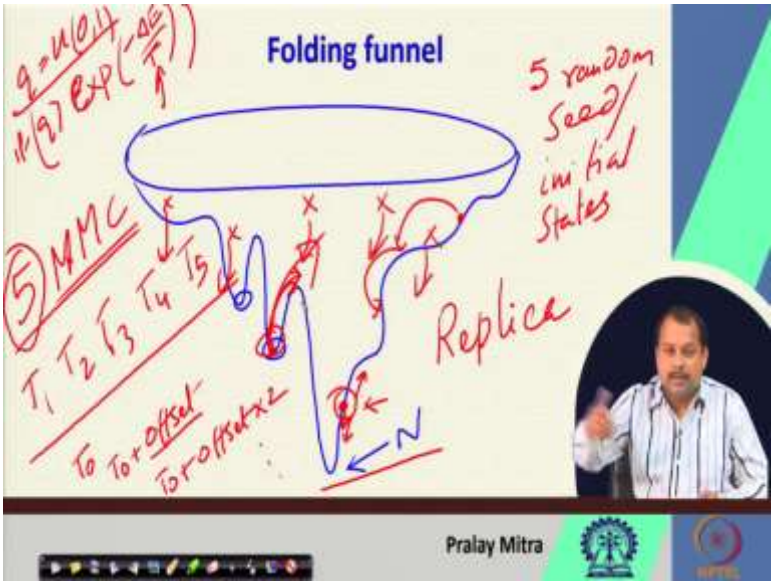
And the current state is giving me a new conformation then I am taking the difference of the energy for the new conformation and the previous conformation if my energy delta E is less than the is less than 0 which means the new conformation assumes lower energy state compared to the current conformation then I am accepting that one without any hesitation that is my Monte Carlo.

But we mentioned because of the local trapping, we wish to avoid that. So, we are also with some probability assuming some of the higher energy state and that probability is determined by this. So, q I am computing as one probability which varies from 0 to 1 and then if q greater than e to the power minus delta E by T this T is again Boltzmann temperature delta is the change of the energy if it is greater than then I am assuming that state as the current state.

So, this is my Monte Carlo simulation. Now, this is not enough in the context of the protein energy function that we discussed.

(Refer Slide Time: 05:05)





So, we are moving to replica exchange Monte Carlo the same folding funnel, I did not have to recap I believe you are aware about that one. The problem is started because of the nature of the folding upon a like this again purposefully, I am not considering the kinetically control that is amyloid fibrils situation this is my N or native state and these are my locally trapped regions.

I wish to avoid and I mentioned in order to avoid that one what I am going to do is I am going to generate say 1 2 3 5 random seed values or initial states. Now, when I generated 5 random seed

or initial state. So, 5 in 5 different locations they will be selected it can be here it can be here. So, I do not know it is random.

Now, what I am proposing as if I am running 5 metropolis say metropolis Monte Carlo simulations. So, this is running one this is running one this is running one this is running one this is running one if I do and this 5 I can control. So, currently I am generating 5 random seed but I can increase that one say 10 20 30 40 like that I can change.

So, if I go for that one then it is very unlikely that all my metropolis Monte Carlo run will be locally trapped. So, I am increasing the probability that my simulation all my simulation will not be locally trapped. So, that way I am running say 5 different metropolis Monte Carlo simulations. Now, that cannot be enough, along with that one what I am introducing is that you remember that my criteria was what if the energy of the new conformation is less than accept otherwise accept with some probability.

So, what is my probability q I generated U 0 1 then if q greater than exponential to the power minus delta E by T then I am accepting that one. Now, you look at this temperature T if the temperature varies, then the acceptance rate will vary assuming this q is a random number. So, it is generating some random number if I assume that it is generating the random number then delta E I cannot change but keeping q and delta E as cost and if I change T then the acceptance rate will change.

If I increase temperature, then the acceptance rate will change if I decrease the temperature then also acceptance rate will change. So, how many say higher energy state will be accepted is determined by this temperature factor. If that is the case, then I can consider as if or you can assume that as if.

So, this initial state from there I am making some random movement in order to find this valley N this is a hilly region I am on top say from helicopter I landed on top of some hill then I am searching for the valley. The green valley and I know there exists only one green valley that is this instead. Now, the movement I will make on small movement or large movement.

So, that movement will be controlled by this temperature T. That will be my acceptance rate whether I will go down or go up and down and how many times I will go up those things. Now,

based upon that one say 5 different metropolis Monte Carlo searches are running in parallel. So, if all of them with the same temperature value then all of them are making same progress.

But you can see that when I will reach to this region, then I have to be very much careful and I have to go downwards most backward less whereas, if I am here, then I have to make large steps. So, from here to here then here to here. So, that I can reach here quickly or when I will be trapped here, then I have to make a large movement.

So, that I can come out of the trap region. So, some region requires that the step will be high which means that here from here to here, I have to jump out. So, the acceptance rate indicates that you accept more and more situations where the energy is not minimum in this case, you accept in this case you accept less states whose energy is minimum.

So, this is one situation which demands that we are almost there. So, walk slowly and do not walk back. Back means that you do not accept the conformation which is with the larger energy state. In this case, it will be just opposite. So, that will be controlled by the temperature. So, what we will do that, we will instead of one temperature we will have see five different temperature values corresponding to each metropolis Monte Carlo simulation.

Now, these five temperatures what we can think of is with some offset and there is a minimum and maximum fine no problem. So, what we can think that. So, it will be T 0, T 0 plus offset, T 0 plus offset multiplied 2 that way it will go. So, you have to decide what will be the minimum temperature what will be the maximum temperature and how many replicas then you will get the offset and corresponding to each say metropolis Monte Carlo you can have one temperature since 5 Metropolis Monte Carlo runs are there which are running.

So, each one will be called as one replica or metropolis Monte Carlo. So, to summarize what we discussed as of now, looking at the folding funnel of the energy landscape or protein energy landscape or protein folding funnel, we noted that there can be a number of local minima along with the global minima in order to reach to the global minima, it may not be sufficient to start with one seed value and move randomly using Monte Carlo simulation techniques to reach to the global minima we may be locally trapped.

So, instead of one we are going for multiple that multiple how many that will also decide multiple random seed we will generate and corresponding to each seed we will assume that as if one metropolis Monte Carlo runs are going. Now when say in such metropolis Monte Carlo runs are going then similar thing are going on so we are calling that as if one replica.

So, N number of such replicas are going on each replica indicates one metropolis Monte Carlo search, metropolis criteria is there. So, you need to tell what will be the temperature because we are not going to accept only the minimum energy conformation but the conformation which will have more energy function, sometimes we have to accept that one using some probability.

If that is the case, then for N number of replicas, which are running we need to decide what is the minimum temperature what is the maximum temperature and if N number of replicas they are, we can distribute them evenly or we can decide that way that it will be not even so based upon that one we can decide. So, different number of temperatures will be there.

Each temperature indicates the acceptance right If I change that temperature acceptance rate will change. And by looking at the protein energy landscape or the folding funnel, we noted that in some region where we are almost there to the native state. So, our acceptance rate will be only to one energy state whereas, when say we are far away and say making local moving towards a local minima, then our temperature needs to be higher.

So, that we can come out of the local minima say from here we can come out. So, sometimes temperature high is required some time temperature low is required N number of replicas are running then the biggest question is how do I understand that where am I because you have one energy function that we did not discuss.

We will discuss that we have one energy function and based upon that and we are doing the simulations. Now, how do I understand that this particular replica is going towards locally trap region or not? How do I understand? So, there is no way to understand during the simulation. So, what we will do. We will not go for full execution or say N number of iterations for each replica and they will be mutually independent with each other and finally.

We will combine the result we are not going to do that one, what we are going to do that one, we are defining some local number of steps. And we are exchanging the information as if you can

assume the same example that I mentioned that from helicopter one person was dropped on top of some hilly region and he is asked to find out one valley that is N native state and that particular native state exists only once.

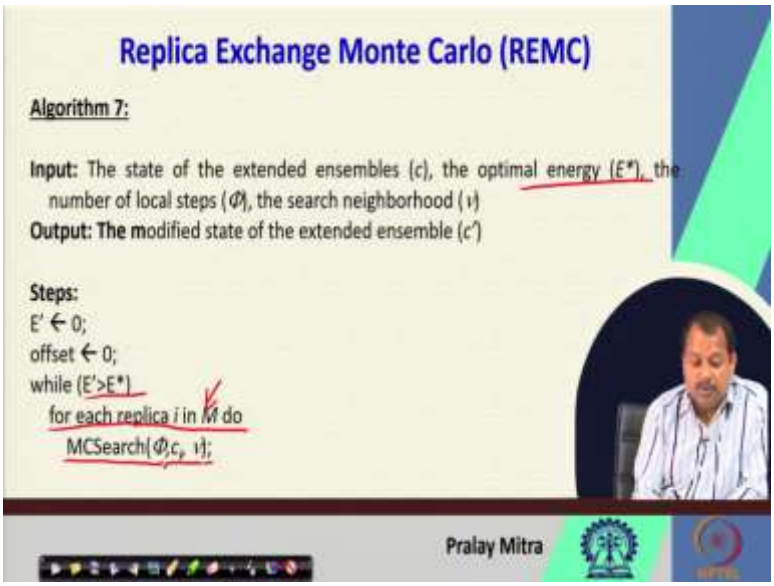
Now, that was the Monte Carlo simulation assuming that metropolis means sometimes when he will be in some say a local region which appears that there is no more no way to go down then he will climb little up to search for other area or other neighborhood and if finds that probably there can be another valley. So, he will jump, jump to that location or he will go to that location and from there he will search that was the metropolis Monte Carlo.

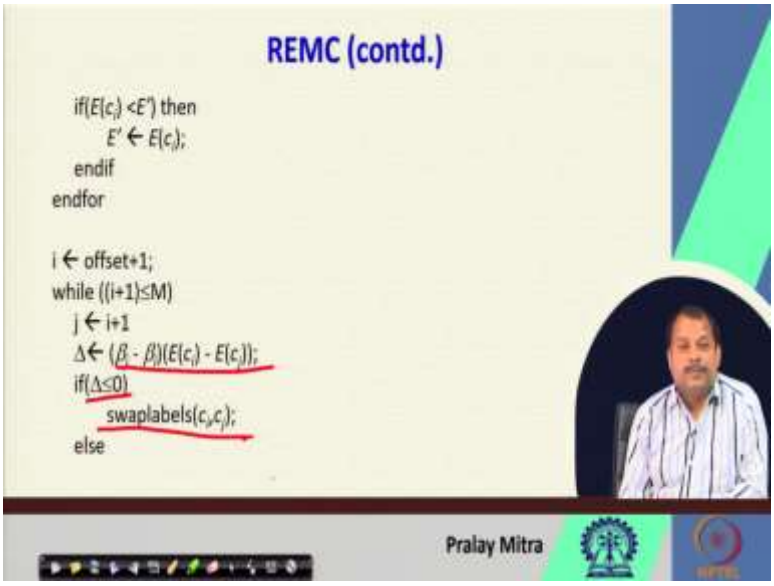
Now, in case of replica exchange, instead of one person say N number of persons was dropped randomly on some different locations. So, once situation, each of them will search by their own and one searching process is over then they will combine the result. But, I am telling no. So, first thing is that N number of persons are dropped on some different locations, they mentioned that you make this much movement, so, small movement or large movement.

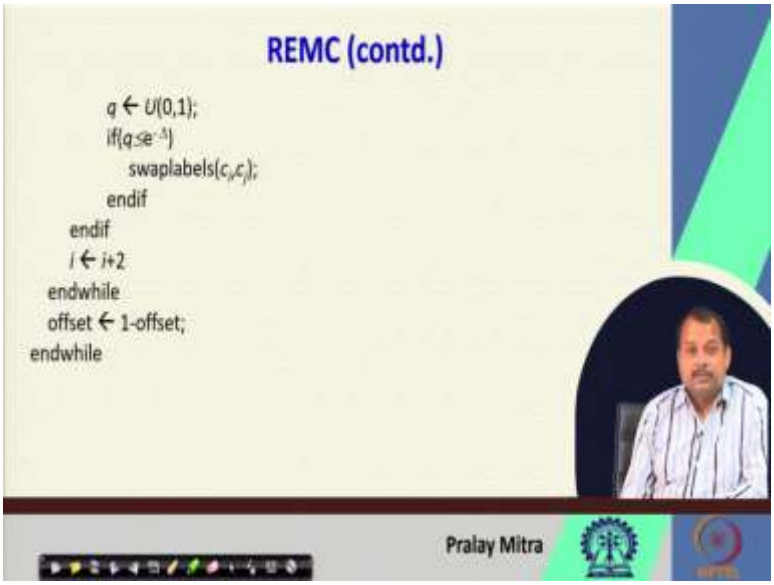
So, I am giving some temperature to them and I am also mentioning after say one hour two hour you exchange your information. So, that person who is making a big step. He will find that by big steps are more acceptance rate whether he is able to get lower energy compared to the person who is making a small step If yes, then exchange the information if no then it in that one.

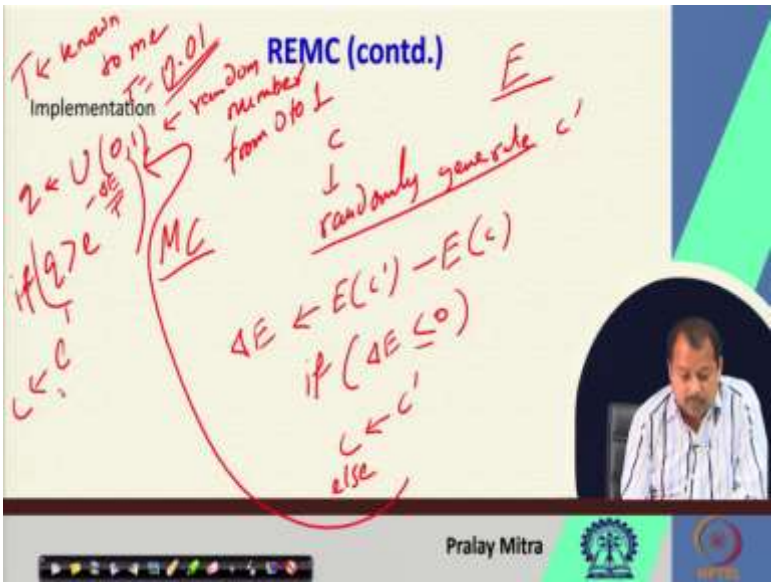
So, if somebody after one hour interval, if somebody one hour interval means after making a sufficient number of iteration or step in your Monte Carlo. So, local step if somebody is able to make a good progress then return that one if not, then you please exchange with your neighbor. Neighbor means another person who is in also doing the Monte Carlo search you do that one.

(Refer Slide Time: 19:23)









So, that thing combines in replica exchange Monte Carlo insort, we will call that as REMC. So, input the state of extended and symbol c extended means. So, different persons are dropped on the hilltop. The optimum energy E star the number of local steps phi the search neighborhoods new.

So, number of local step is mentioned which means that after say one hour or say after making 100 steps you please exchange your information with your friend or with another person and when we are making exchanging that information then we are supposed to exchange the information with my neighbor.

The neighbor is also defined which means that I am making say my step sizes say 2 fit and another step person step sizes say 1.5. So, I will communicate with him only not with the person who steps sizes is 5 fit. And the output is the modified state of the extended M symbol C prime. So, input output is more or less same with the Monte Carlo technique in the steps what we are doing first we are initializing E prime cost to say 0.

So, that way you understand that we are going to negative. So, offset is 0 also while E prime greater than E star my optimal energy E prime greater than for each replica i in M do MC search phi ci Nu this MC search you understand that we already discussed. So, Monte Carlo search. So, local step is phi say 100 number of local step ci is my current confirmation and Nu is the neighborhood with that one. So, you search for 100 steps.

So, this M is my model which includes the temperature information also. So, this temperature information I am supplying temperature, I am supplying you the current state, I am mentioning you that 100 steps, this Nu I, am this pie this, 100 step you have to search and the Nu which is a neighborhood that I am mentioning and in return what you will give me after 100 step what will be my current conformation and what will be the energy of that conformation you will give me that information.

If E ci less than E prime initially it was E prime. Now after Monte Carlo search MC search. So, I am getting a new conformation that is ci if that is less then I will accept that one like the previous one. If not, then I am changing my offset and in that offset this beta j and beta i that temperature I am multiplying with E ci minus E cj and if that energy is less than then swaplabels. So, I am changing their temperature.

So, I am changing their temperatures and when I am changing their temperatures then they are movement steps are also changing. So, that is it about the replica exchange Monte Carlo. So, this is the basic algorithm and this Metropolis criteria also I introduced here q equals to U 01 with some probability whether the levels will be swap or not that I am deciding and then I am changing the offset and I am coming out of that loop.

Now, let us discuss about how you can implement this replica exchange Monte Carlo because it will be useful for protein folding as well as for the protein design problem. So, first of all I believe that implementation of the Monte Carlo simulation is clear to you. So, you will assume

the existence of one energy function. That energy function we will discuss later. So, if such an energy function exists.

So, let us assume E then what you have to do, I am first talking about Monte Carlo search. So, first you have C one conformation. So, from C randomly generate C prime which means you have to implement one random function which will take C as an input and will give you C prime as an output after generating the C prime.

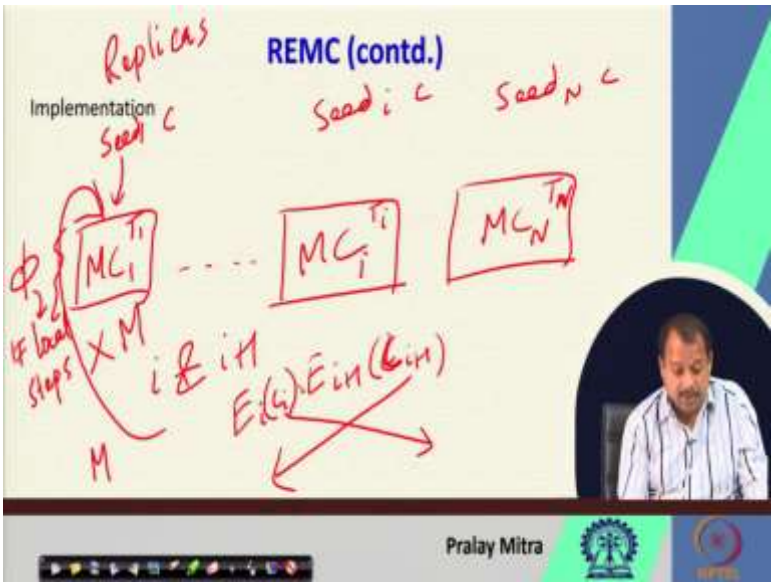
So, you can compute EC prime because I am assume the existence of one energy function it will vary from one implementation to another. So, when we will discuss the competition framework, I will give you that energy function but for the time being, you assume that one such function also exists in your implementation, which you will take C prime as an input will give you the energy it will take E as c also as an input and will give the energy of C then that will give you that delta E.

Now you know that if delta E less than equals to say or say less than that is based upon your choice then you are assuming C prime as your current state. If not then what you are doing else part. So, I am taking from here to here else I am generating q U 0 1 so, random number I am generating one random number from 0 1 the q then temperature T known. Definitely the temperature you need to decide.

So, I have seen that temperature value off say 0.01 is good enough but you can change this one again this temperature is not nothing to do with the environmental temperature of a protein it is regarding the distribution of your Monte Carlo simulations space. So, in Monte Carlo simulation you are generating the space assuming the Boltzmann distribution one temperature will fix so, that is your this temperature.

So, 0.01 may be good for your purpose then you are assuming if q greater than here exponential greater than e to the minus delta E by T simply you can calculate the, you accept C as the C prime otherwise you reject. So, that way you are accumulating N number of cases. Now, this is your Monte Carlo simulation with the metropolis criteria.

(Refer Slide Time: 27:00)



Now, you assume that this is now your box. So, what do you have to do that there are a number of such MC box say N i MC1. Corresponding to each box there is one temperature also attached T1 Ti Tn temperature is also attached. Now, what you have to do so, you have to start with some random position see seed1 seed i seed N and with that one this seed is nothing but my conformation C you execute this one.

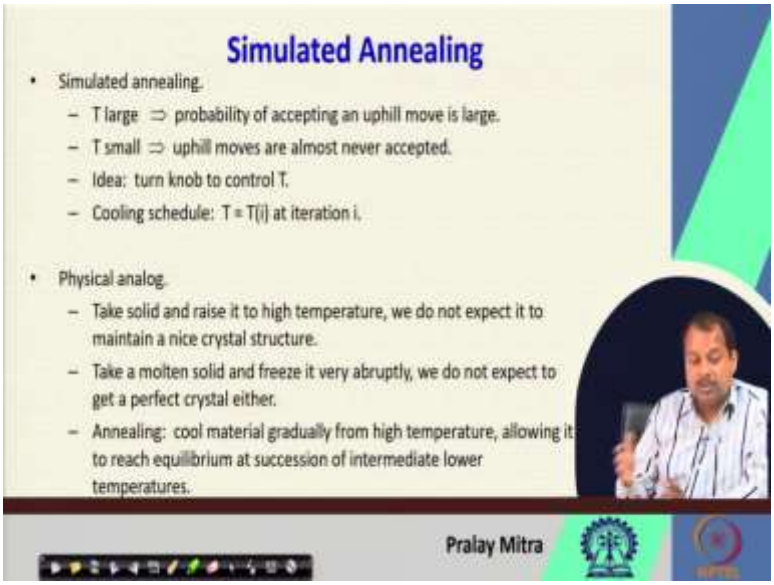
How many times it will run phi that is my number of local steps that you can assume say 100, 200 based upon your interest, after that what you will do that you exchange when you exchange you exchange with i and you compare first i and i plus 1, two consecutive these are my replicas. So, to consecutive replica, you compare their energy both on the current conformation.

Now, based upon that one, if you find that energy is reducing, it is fine. If you find energy is not reducing, then what you need to do that two things you can do either you can swap their temperature, but I will prefer swap their levels which means this conformation if I said this is my ci plus 1. This conformation will go here this conformation will go here swap levels are swap temperatures.

After swapping you go back and do the same thing means local steps and this process will keep on going until so this loop if say REMC for M the local number of steps multiplied with M will

be total number of steps are simulation in your REMC. So, after this number of steps, you come out, you accumulate or you take the union of all the conformations that you accepted during each replicas or during each Monte Carlo simulation technique and that complex. This very simple and easy to implement. You can try this at your home.

(Refer Slide Time: 30:02)



So, one thing I wish to touch on here since we detail, we did detailed discussion on replica exchange Monte Carlo another technique which is called as a simulated annealing. So, today just the cursorily I will mention, but later I will give a detailed discussion on this also. So, simulated annealing is also using the same concept of the replica exchange Monte Carlo.

Where T large is a probability of accepting an uphill move is large T small uphill moves are almost never accepted. Idea turn knob to control T. So, you control the temperature similar to replica exchange Monte Carlo. Cooling schedule T equals T i iteration i and physical analogue of similulating annealing is that take solid and raise it to high temperature keep on rising as if we do not expect it to maintain a nice crystal structure it will melt allow it take a molten solid and freeze it very abruptly, we do not expect to get a perfect crystal either.

So, because of this say increase in temperature and quickly down quickly because of that one it will take some particular say process and that annealing cool material gradually formed high temperature allowing it to reach equilibrium at succession of intermediate lower temperatures.

So, we will discuss in detail later. So, that is it for today. So, what we discussed in summary is we started with a protein energy landscape.

And looking at the protein energy landscape we understand that Monte Carlo method even the metropolis Monte Carlo method is not going to be sufficient for protein folding or protein design where we have to deal with the protein energy landscape, where they are can be a number of locally trapped situations or local minima. Although biologically we accept and understand the situation likes the formation of the amyloid fibrils which are going to be the global minima state.

But if for the time being we allowed that situation and then even the native state may not be reached if we start with one seed and that particular seed or the random initial point leads to some local state. So, solution we are trying to give that instead of one local seed, we will start with a number of local seeds.

Where each seeds will have a different temperature values and they will learn individually for a number of certain local steps after that one they will exchange the information with their neighborhood the neighbor we are describing or defining based upon their temperature. And if such seed values which we are calling as replica, also, when they will exchange their information after the local step then there is a possibility that locally trapped region we can avoid and we can go to some globally native state. So that is it. Thank you very much.