**Algorithms for Protein Modelling and Engineering**
**Professor Pralay Mitra**
**Department of Computer Science and Engineering**
**Indian Institute of Technology, Kharagpur**
**Lecture: 21**
**Protein Energy Landscape**

Welcome back, today we will start the topic on protein energy landscape.

(Refer Slide Time: 0:27)



So, I am planning to discuss the concept on protein folding, because the energy landscape is very much related to protein folding. And you remember that we started to discuss about this protein

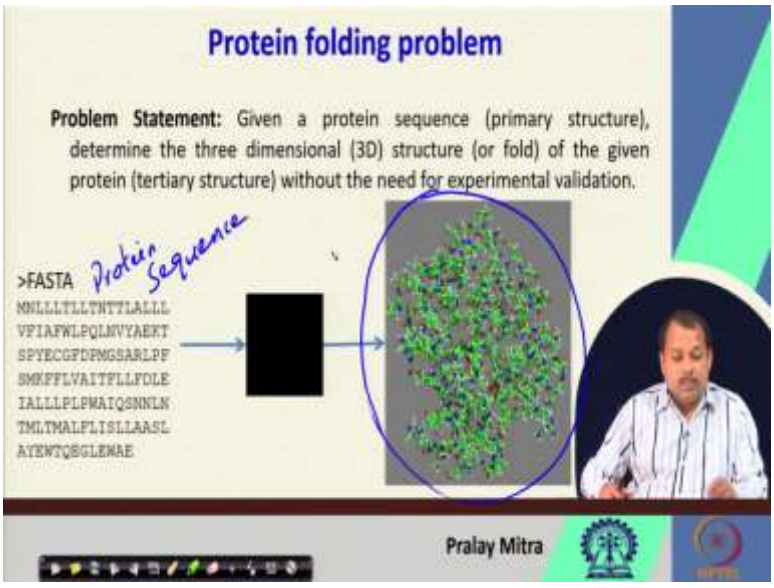
folding the definition I gave on the last week. Now, we also discuss the Monte Carlo similar simulation technique, but there are some problem with the Monte Carlo simulation technique.

And although we demonstrated that for the integration of an equation or say, for computing the value of the pie, the Monte Carlo simulation technique is perfect. But for our purpose, I mean for protein folding problem, this Monte Carlo simulation technique, even along with the metropolis criteria may create some problem. So, the problem is triggered by the nature of the protein energy function, although we designed one energy function.

But you remember that I mentioned when I say doing the integration of one function say integration of fx dx, sometimes it is difficult to have that effect. So, we approximate to gx and even if we approximate to gx, then we will find that there is some nature because of that nature actually it might be a problem to run only the Monte Carlo simulation technique wven if there is metropolis criteria.

So, those things we will discuss in today's lecture. So, that is why the protein folding protein energy and the protein energy landscape, what it, so is those things we will discuss today. And the keywords that is why I selected protein folding and energy landscape.

(Refer Slide Time: 02:12)




So, let us recap with the definition of the protein folding problem. It says that given one protein sequence as an input, which we can also call as protein primary structure, you need to determine

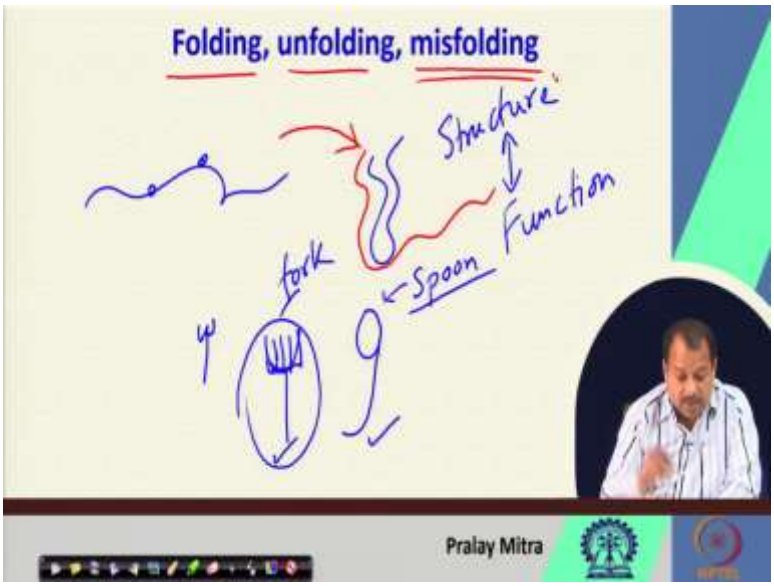
the three dimensional structure of it or you need to determine the atomic level resolution or the atomic model corresponding to that input sequence without the need for experimental verification or validation or testing.

So, that is about the protein folding problem. And, here the diagram indicates that the same diagram that we discussed on the last week. So, FASTA is my input, since it is a sequence. So, I am mentioning it as a FASTA, but you can consider that this FASTA is nothing but the protein sequence that also that you can also see, it is a single letter quote.

So, M, the first one indicates methionine N the second one indicates asparagine T L. The third, fourth, fifth one is indicating the leucine T indicates the threonine like that where it is. Now, if there exists one black box, that is the computational technique or a software tool, which will take this FASTA as an input and will give me this three dimensional model as an output.

So, that is my requirement that is the protein folding problem, however, when we are designing this black box or the software tool, which will take one sequence as an input, and will give me one structure as an output. So, there may be some possibilities, it may not be the correct one. And if I go back to the biology and or say chemical biology and try to see that, what is the behavior of a protein when say in vivo or in cell it is behaving then we will see some interesting features of that particular protein. So, that features we will discuss first.

(Refer Slide Time: 04:42)




So, first thing is that given one protein sequence, it is supposed to fold into a particular structure and that particular structure will have one function. So, for example, you can consider that once a sequence is given and it will fold to this structure and it will have one function, but not necessarily it will always fold to that one there may be several reasons say environmental presser or say mutations.

Mutation means that some point has changed because of that one, it may not take this particular shape or structure. Now, you know that if one particular shape or structure is there, then based upon that one. One function will be determined. So, for example, if I assume that so, the basic material is always is still now out of that still I can say make one shape like this, I can have something like this I can have something like this.

So, if you assume that say this is my fork, this is my spoon then this has one function this has one function and there is kind of a correspondence between the structure and functions. Say instead of say forming the spoon if I form a fork, so, that is also a structure that you see that say I am taking some melted steel and instead of making a spoon, I am make one fork then all the functions of a spoon.

So, for example, say drinking soup. So, that property will not be satisfied by this fork. So, if that is the situation then during this sequence to structure folding process, if something goes bad or goes wrong. Then what will happen that because of this mutation environment pressure or say change of PAH, etcetera etcetra, that we will discuss then if this particular shape is not assumed then one particular function will be absent.

I believe you all know that inside the ribosome from DNA to MIRNA and then to protein the transcription translation process will take place and when say MIRNA is translating to protein then it will take some shape that we will call as a fold and as if that this will move from this sequence to this structure.

Now, if because of anything this particular structure is not assume and say it is assuming one structure like this. Then clearly you can see that the blue structure and the red structure is different also their functionality will be different. That is why you say if I assume that this particular protein sequence is a protein sequence of a human body, then if instead of blue structure, which is native structure.

Native means that it is the functional form and it has one particular function instead of that one if it assumes rate structure, then according to that rate structure, it will have some function, but that function is not same as the blue function whereas, your body is waiting for the blue structure and corresponding to that blue structure there is a function and that function will be absent in your body.

So, you may lead to some situation which can be a disease condition. So, that may happen if a protein misses two fold or misses to assume it is correct folding structure. Related to that, there are three concepts folding, unfolding, misfolding. So, when a particular sequence assumes a correct structure. When I say correct, which means that it is known that this sequence is supposed to take this particular structure and this process.

I am talking in the context of the in vivo which means inside the cell, I am not talking about developing one algorithm who is folding. So, in biology or in chemical biology, when it assumes one particular structure corresponding to one sequence, when it assumes that particular structure then it is called as a folding.

However, in the process if it happens that instead of blue structure it is taking the red structure, because of that one it folds, but that is not a correct fold. So, I will call that as a misfolding. Now, you will be surprised to know that in our body, there is a mechanism through which it can senses that whether a protein is assuming a correct structure or not.

Most of the time and the moment it will recognize that probably it is not assuming the correct structure, then it will allow to misfold, I mean open that folding, which was leading to some incorrect folding and then allow it to fold it again. Now, if for the second chance also it will not able to fold correctly then definitely it is going to misfolding state. And unfolding indicates that when a sequence is not in the fold it forms.

(Refer Slide Time: 11:26)


So, regarding this few features, we are going to discuss the correct three dimensional structure is essential to function although some parts of functional proteins may remain unfolded. So, this is one unique situation which can be found in protein say it is assuming one structure like this. Now, you see that this region so, this region is folded but this part is not folded that may possible. But the function will be then corresponding to this one.

Next failure to fold into native structure generally produces inactive proteins, but in some instances misfolded proteins have modified or toxic functionality. So, when it will fail to fold to native structure again native means that it has a specific function to do and for that specific

function it should assume one particular structure and if that particular structure it cannot able to assume or it cannot fold to that particular structure.

Then what will happen that it will assume another structure. Now, that particular structure may be completely inactive do not have any function at all or it may have some function. Now, that function mostly be not good forward body that is why it will be with that toxic functionality. Which is unwanted not required forward body.

Next, several neurodegenerative and other diseases are believed to result from the accumulation of amyloid fibrils fromed by misfolded proteins. So, when there is the misfolding, because of that, one that amyloid fibrils are formed, this particular amyloid fibrils is more stable compared to the native state from stability point of view.

But, that, although we are talking about stability of one protein molecule, but of course, the functionality is also related to that stability, if we reach to some situation where it is of utmost stable, but there is no function then it is useless. So, something like amyloid fibrils it will reach to some region where the energy is even less compared to the native state.

But it do not have any function. That is why make a note of this that is why computionaly when we are designing some software tool, then definitely we will try to optimize our energy function or we will try to reach to the minimum energy state or stable energy state corresponding to one protein structure that we are modeling, but that must have one function, if we reach to some situation. So, that you do not have any function then that is useless.

Then, the question may come in your mind that how do I discriminate that whether it will have a function or not. So, that thing you can fix after your modeling you can do some corrections or some cross checking. Cross checking means whether that is assuming a particular structure following that ramachandran plot or say whether it is having some fold or not. So, those things you can check and then perhaps you can say whether it is a good modeling or not.

Many allergies are caused by incorrect folding of some proteins, which means, incorrect folding means that. That protein is supposed to take some structure and have some function but it folded incorrectly and because of that incorrect folding two situation occurs one is that you are starving for the correctly folded protein or you are starving for some known or essential protein because,

that particular fold it is not assuming and you are generating some incorrectly for led protein whose function is not required for your body.

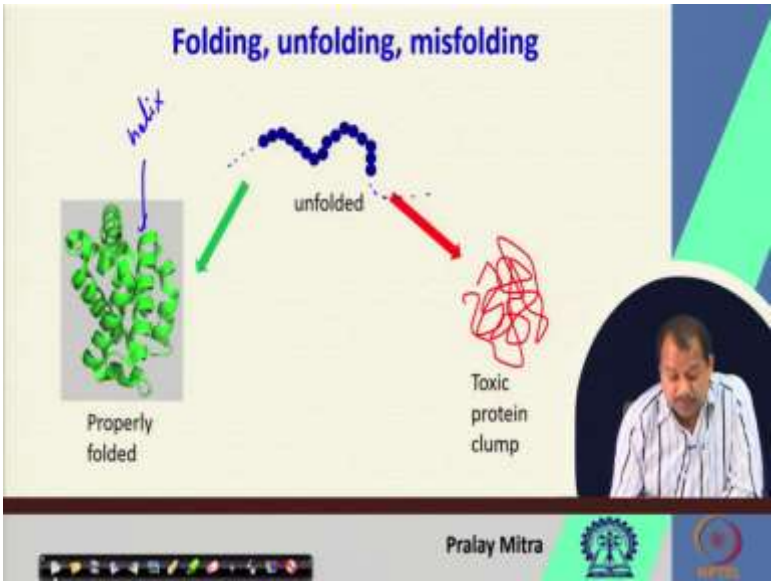
So, unwanted thing mostly leads to some allergy for the immune system does not produce antibodies for certain protein structures. So, if it is a misfolded or are taking some wrong safe, then it is not possible for say human body to identify that it is misfolded. So, throw it out of the body like it used to do for say foreign bodies.

So, when foreign body enters to our body then our immune system identifies that it is a foreign body. So, throw it out of the body it may create some problem in our body again, some disease, viral disease or other disease when they said seen then they cripples the immune system fast and the immune system will not able to track or detect that particular foreign body and that way they progresses and stays in our body.

So, that is one issue different one, but in this context what I am trying to say inside the body, no foreign body is coming inside the body, the protein instead of taking its native state, it is taking some incorrect state and because of that incorrect state, it is unwanted in our body, sometimes our immune system may identify this is not required from where it is coming, I do not care from where it is coming say it is not required to throw it out of our body.

Sometimes it is possible but not always. So, when it is not possible, then it leads to some disease or say some allergy condition.

(Refer Slide Time: 17:59)




So, here is the schematic the diagram for folding, unfolding and misfolding you see this is one sequence. So, each circle indicates a one amino acid, but you should remember that it is not complete it is extending in both ways, so N-terminus and C-terminal, so only a small part I have drawn but that is a long portion of it. So, one situation that this is unfolded and it will take a properly folded step that you can see here, this spiral indicates the helix that is one regular secondary structure.

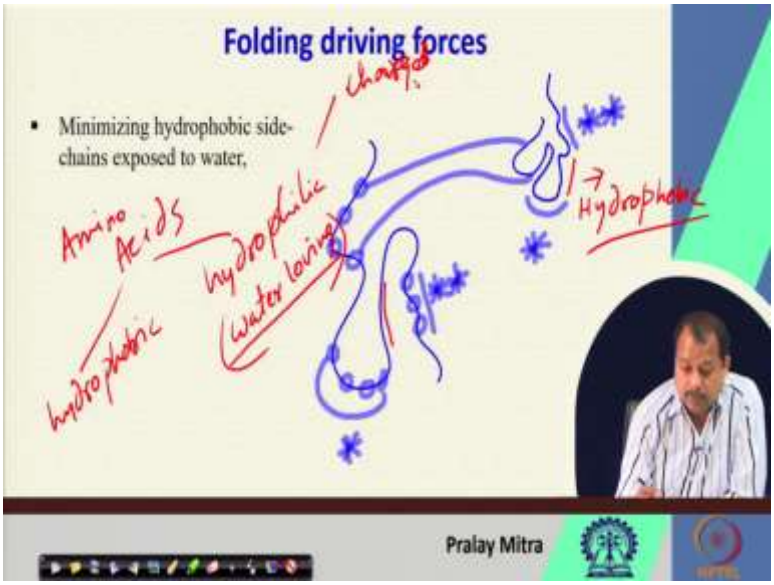
Now, it has some shape or because of some factor it may take some toxic protein clump. Now you see that this shape is kind of irregular in nature, so that spirality, etcetera. I cannot able to see here. Now this toxic protein clump definitely is unwanted in our body and may lead to some disease condition. May lead to sometimes our immune system may identify that this is unwanted and so throw it out of our body. If it is then it is fine if it is not then it may create some problem.

Now this protein toxic protein clump may occur because of several reasons. One situation may be that some amino acid during the evolution process or translation process inside our body is changed because of that change. It may happen during the folding process, the others or the environment which is supposed to allow it to assume it is correct folded structure like this green one is not able to support it properly and because of that one it is going to that red state here it is toxic protein clump.

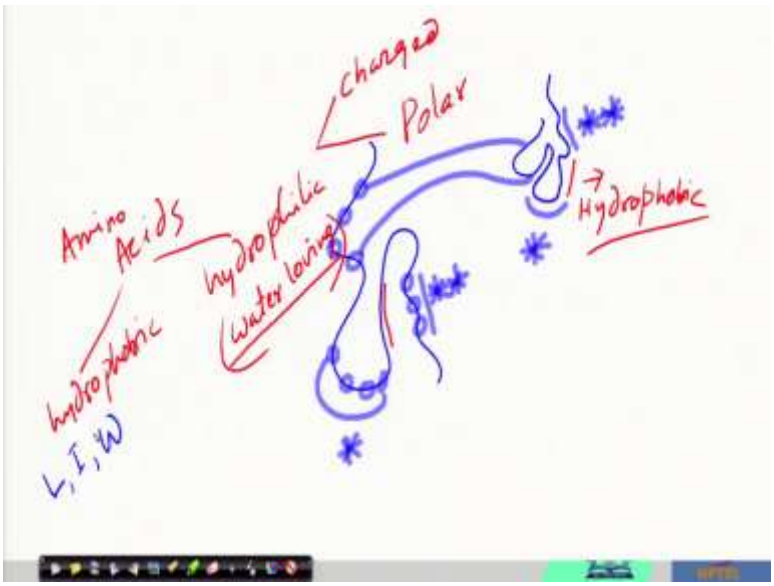
(Refer Slide Time: 20:12)

## Folding driving forces

- Minimizing hydrophobic side-chains exposed to water,




Pralay Mitra



## Folding driving forces

- Minimizing hydrophobic side-chains exposed to water,
- solvent (water or lipid bilayer),
- concentration of salts,

Charge AA
R, K
Basic
D, E
Acidic
Na+
H

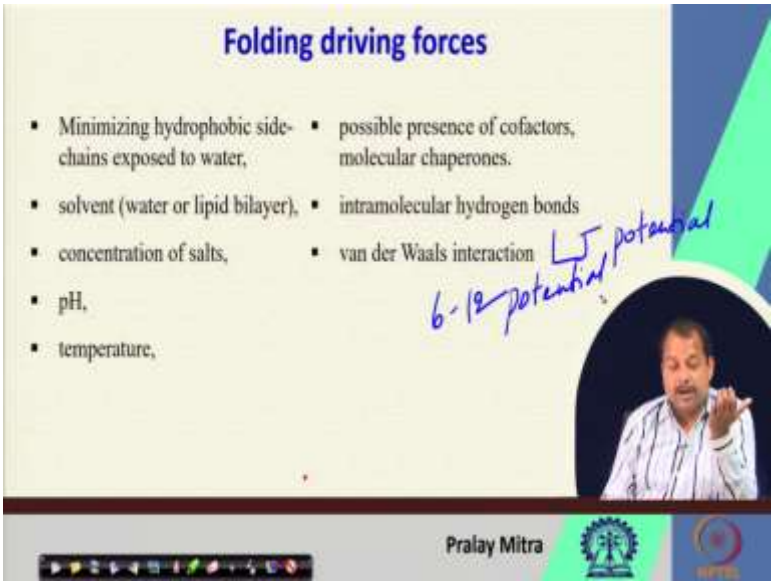
Pralay Mitra

## Folding driving forces

- Minimizing hydrophobic side-chains exposed to water,
- solvent (water or lipid bilayer),
- concentration of salts,
- pH,
- temperature,
- possible presence of cofactors, molecular chaperones.
- intramolecular hydrogen bonds

Salt bridge
disulfide bond
ionic bond
S-S

Pralay Mitra





Now, who are the protein driving forces? So, first one although its individual contribution is very small that is the hydrophobic side chain. So, hydrophobicity: So, in a protein say if I assume this is my unfolded state. So, let us as you that there are some these are highlighted parts are say hydrophilic the rest of the parts are hydrophobic.

Then if you assume that this particular unfolded protein or partially folded protein not completed and is not in a straight string form if I put it into a water that is my solvent then it is going to take some say like. So, here this corresponds to say this region this region corresponds to this region this region corresponds to this region and apart from that you see that there are some regions say

like this region which is the sum of this part and this red region is say hydrophobic they are also exposed to water.

But you have to understand that it might be but the aim is that optimizing the fold in such a way that contribution of the hydrophobic being inside the code or away from the water is more compared to that hydrophobic will be on the surface, so hydrophobic. So, this amino acids and classified into two groups as of now, hydrophobic and hydrophilic, so this is water loving and this is just opposite this hydrophobic.

Now, inside this hydrophilic there are variations like charged, polar. Now, there are some hydrophobic residues and hydrophilic residues. So, hydrophilic loves water molecule. So, it will try to be on the surface but hydrophobic do not like water molecules. So, examples here I can give as a leucine, lysine, isoleucine. Then say what else to some extent you can consider.

So, TRP. TRP means W. So, these are some of the hydrophobic residues. So, there are some hydrophobic scale through which you can measure and you can identify, but the point is that. That is although the contribution of that hydrophobic residue individual contribution is less compared to say charged charges interaction or say disulfide bond interaction but the collective or the if I take the summation over the individual contributions.

Then if I take the individual contributions then it will be very high. So, going back to that slide, so minimizing hydrophobic sidechain exposed to water is one. Next, solvent water or lipid bilayer – Now, the point is if it is water then what will happen hydrophobic will be at the core and hydrophobic or only on the surface if it is lipid bilayer the nature will be opposite.

So, you see that based upon what is the solvent the nature of the folded state will change and that way you may think and you can think correctly that one particular amino acid sequence based upon what is the solvent may assume different folded structure and accordingly they have multiple functions. Concentration of salts: So, salt will give you more ions in the solution because salt the simple salt you can assume is Nacl.

So, it will contribute Na and cl. Now, all the charged molecule who are the charge amino acid, so one is aspartic acid glutamic acid, arginine and then K. So, these are basic and these are acidic sometimes H or histidine is also considered as basic. Now, if the concentration of the salt in the

environment increases then whether percentage of plus increases or basic part ion increases or the acidic ion increases based upon that one who will be on the surface will be determined.

So, the concentration of the salt will also matter then pH. pH and concentration of salt is to some extent related with each other. So, water has a neutral pH on if you go in one side you will get base and other side you will get acid. Then temperature, so each atom has some sub defactor and based upon that one it vibrates.

Now, if temperature increases then the fold state will also change possible presence of cofactor or molecular chaperones which helps to take a particular fold say in presence of their particular cofactor or molecular chaperones the fold may be different compared to in absence of that molecular chaperones or cofactor.

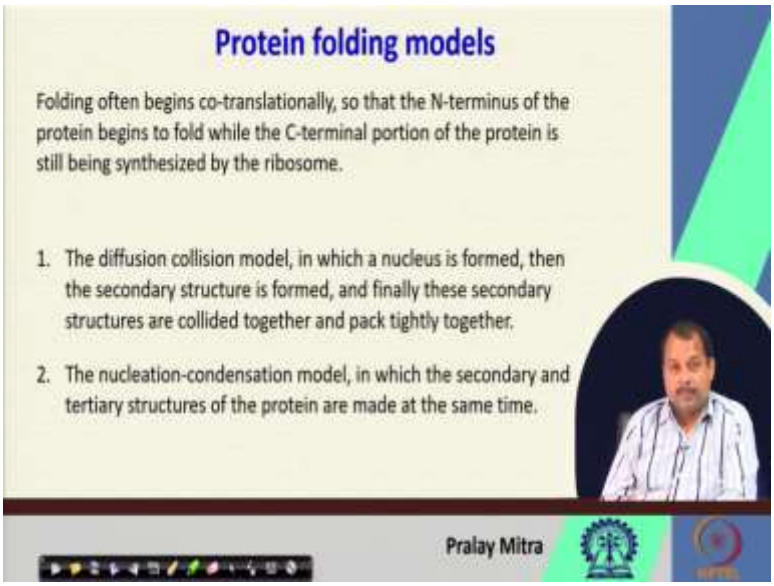
Intramolecular hydrogen bonds that will make an a make a difference again this particular hydrogen bond or their salt bridge. How do you bond hydrogen bond salt bridge disulfide bond. So, these are called as the ionic bond. So, this kind, so disulfide will always occur between to cysteiene or where there is a sulfur, but salt bridge will formed between say one acid and other basics. So, acid I may mention that aspartic acid glutamic acid base indicates that lies arginine or lysine.

So, between this there will be salt bridge and hydrogen bond can be with the main chain with the side chain whenever there is a hydrogen donor and then hydrogen acceptor then there can be a hydrogen bond. So, the complete some of this effect will also be an driving force for protein folding process van der Waals interaction.

So, they are this is also called as the LJ potential or 6 12 potential because there is a 6 and 12 part. So, the equations says that A divided by (())(29:49) 12 minus b divided by (())(29:51) 6. So, that part is also there. So, those are the driving force and many other these electrostatic interactions that I mentioned as the ionic one and many more out there, but these are more or less enough for us when we will formulate it to some energy function.

So, these are in a nutshell the driving forces behind the protein folding. So, when we will say design one protein folding software or tool then in order to design the energy function, then we have to keep all those things in our mind

(Refer Slide Time: 30:41)




Now, I will conclude by mentioning two protein folding models, one is the diffusion coallision model another is the nucleation condensation model. So, folding often begins co translationally. So, that the N-terminus of the protein begins to fold while the C-terminal portion of the protein is still being synthesized by the ribosome.

The diffusion coallision model in which a nucleus is formed, then the secondary structure is formed and finally, the secondary structures are collided together and packed tightly together. Whereas for nucleation-condensation model in which the secondary and tertiary structure of the protein are made at the same time. So, that is it. Now, we will continue to the next lecture the rest of the part. Thank you.