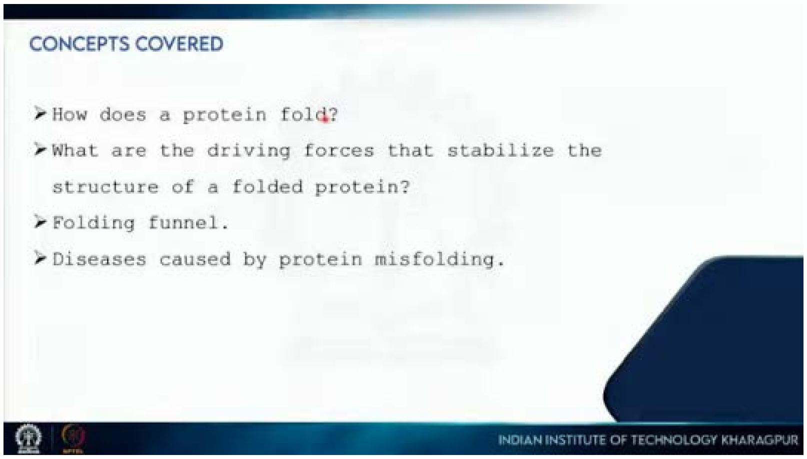
Introduction to Complex Biological Systems
Professor Dibyendu Samanta and Professor Soumya De
Department of Bioscience and Biotechnology
Indian Institute of Technology, Kharagpur
Lecture 12
Protein folding, Folding funnel, Anfinsen’s experiment

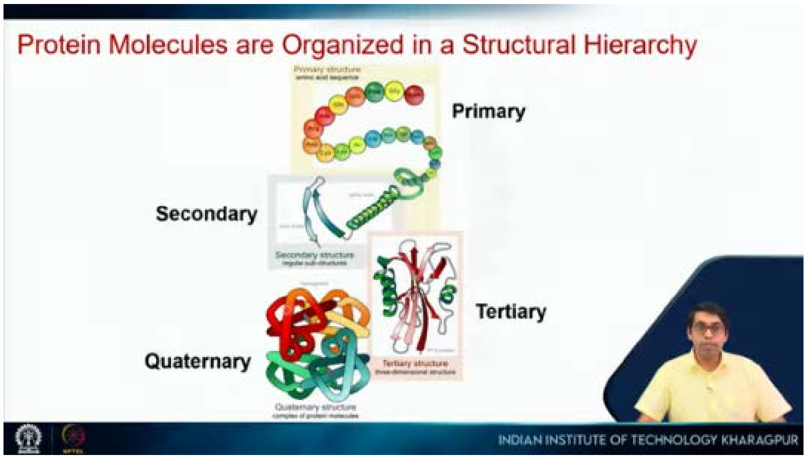
Welcome back. So, today I am going to continue with the second lecture on proteins, and today I am going to discuss protein folding in more detail. So, we will see how proteins fold, and we will see what the driving forces are that stabilize the structure of a protein. I will introduce the concept of the folding funnel, and then I will briefly talk about diseases that are caused by protein misfolding, but this will be the major topic of the next lecture.


CONCEPTS COVERED

- How does a protein fold?
- What are the driving forces that stabilize the structure of a folded protein?
- Folding funnel.
- Diseases caused by protein misfolding.

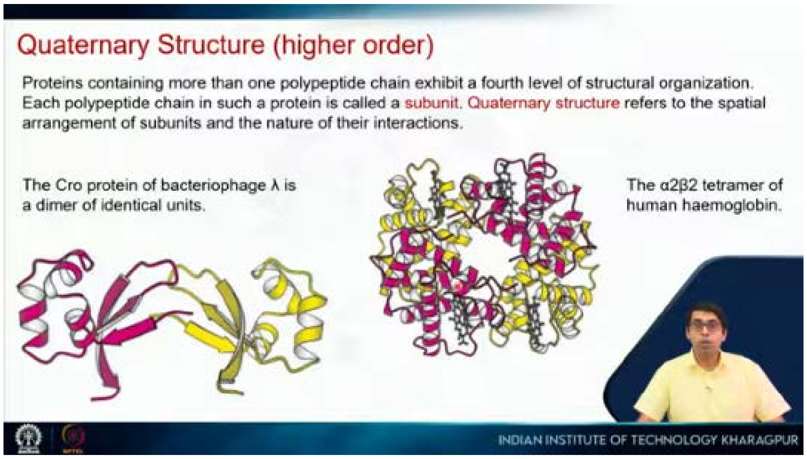
INDIAN INSTITUTE OF TECHNOLOGY KHARAGPUR

So, in the previous lecture, we have already seen that proteins have this particular hierarchy of structure. The sequence of amino acids in which they are connected is called the primary structure of the protein. Then, consecutive amino acids can fold and form either a helical form or this extended form, which are called alpha helices or beta strands, and these are the two secondary structures of proteins. Now, in the three-dimensional space, these secondary structures can come together and form the tertiary structure of a protein. For most proteins, it stops here, but some proteins can also have this higher-order organization, which is called the quaternary structure of a protein. The example that is shown here is myoglobin, where four polypeptide chains fold and come together to form this quaternary structure. Sorry, this is hemoglobin.



Protein Molecules are Organized in a Structural Hierarchy

So, what is a quaternary structure? Proteins containing more than one polypeptide chain exhibit a fourth level of structural organization, where each of these polypeptide chains is called a subunit. So, quaternary structure refers to the spatial arrangement of subunits and the nature of their interactions. Another example of a quaternary structure will be this. So, this is the cro protein from the bacteriophage lambda. So, it's a virus, and it forms a dimer. The same peptide chain interacts with each other to form a dimer, so it's again a quaternary structure. We have already seen this. So, in the case of hemoglobin, there are two polypeptide chains, the alpha and the beta.



Quaternary Structure (higher order)

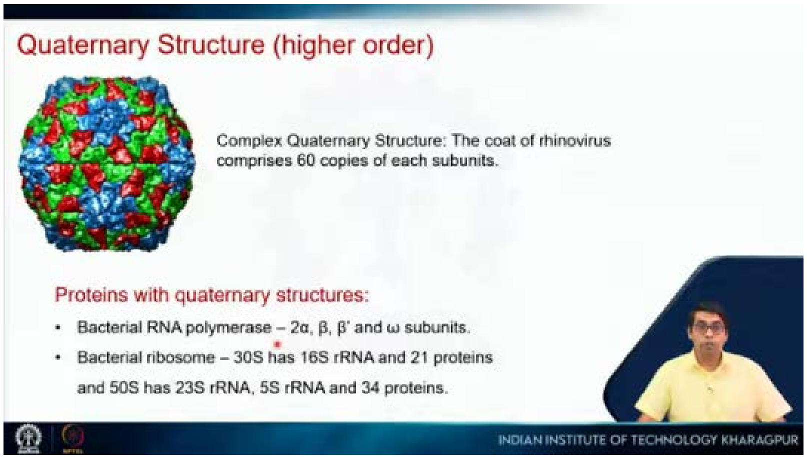
Proteins containing more than one polypeptide chain exhibit a fourth level of structural organization. Each polypeptide chain in such a protein is called a subunit. Quaternary structure refers to the spatial arrangement of subunits and the nature of their interactions.

The Cro protein of bacteriophage λ is a dimer of identical units.

The α2β2 tetramer of human haemoglobin.

So, two alpha subunits and two beta subunits come together to form the tetramer called hemoglobin. This is a more complex structure. So, this is the coat of the rhinovirus, which comprises 60 copies of each subunit. You can see that there are different subunits color-coded differently, and there are 60 copies of each of these subunits, which form this coat of the virus. In this case, it is a rhinovirus, which causes the common cough and cold.

We have already seen examples of proteins or large machineries with quaternary structures. For example, we have seen bacterial RNA polymerase, which has two alpha subunits, a beta and beta dash, and an omega subunit. So, these five subunits come together to form the bacterial RNA polymerase. We have also seen the bacterial ribosome, which has two large subunits, the 30S and the 50S, and individually, these 30S and 50S consist of ribosomal RNA and 21 proteins. In this case, two ribosomal RNAs and 34 proteins.



Quaternary Structure (higher order)

Complex Quaternary Structure: The coat of rhinovirus comprises 60 copies of each subunits.

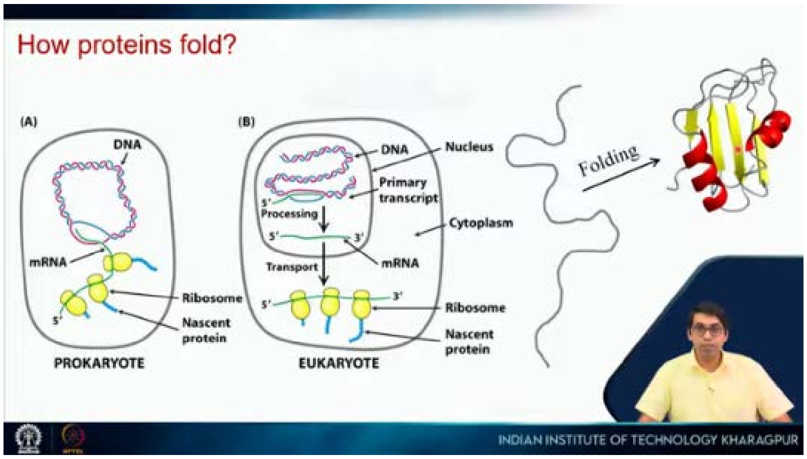
Proteins with quaternary structures:

- Bacterial RNA polymerase – 2α, β, β' and ω subunits.
- Bacterial ribosome – 30S has 16S rRNA and 21 proteins and 50S has 23S rRNA, 5S rRNA and 34 proteins.

INDIAN INSTITUTE OF TECHNOLOGY KHARAGPUR

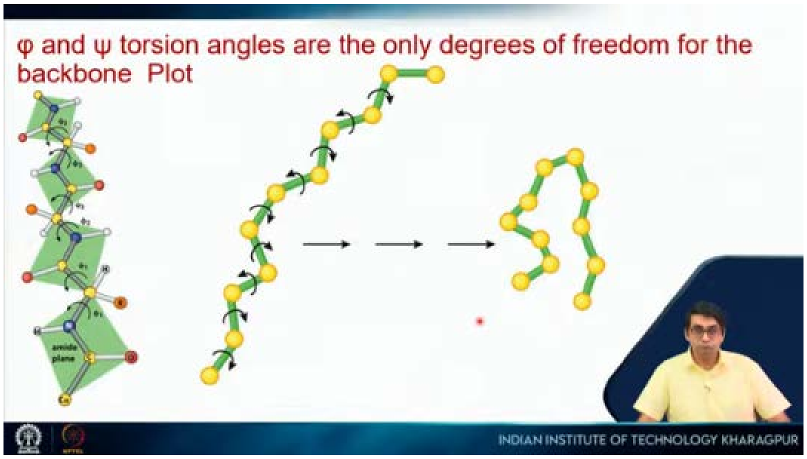
So, in this case, the ribosome not only has proteins but also has ribosomal RNAs as its subunits. So, proteins are, as we have already seen, synthesized by ribosomes like this. In the case of bacteria, transcription and translation occur simultaneously, whereas they happen in two different compartments in the case of eukaryotes. However, what the ribosome does is connect the consecutive amino acids depending on the mRNA sequence. So, when the protein is synthesized, it most probably looks something like this.

But then, how do we go from this to something like this? This process is called folding. So, something like this, which is an unstructured polypeptide chain, how does it fold into a precise structure like this? So, this is the protein folding problem. So in other words, we can think of it like this: this is the extended polypeptide chain, and we have already seen this torsion angle. There is no rotation about the amide bond, but we can have rotations about the phi and psi torsion angles. So, if all these bonds rotate, then this extended polypeptide chain can collapse into a compact structure like this, a very precise structure like this.



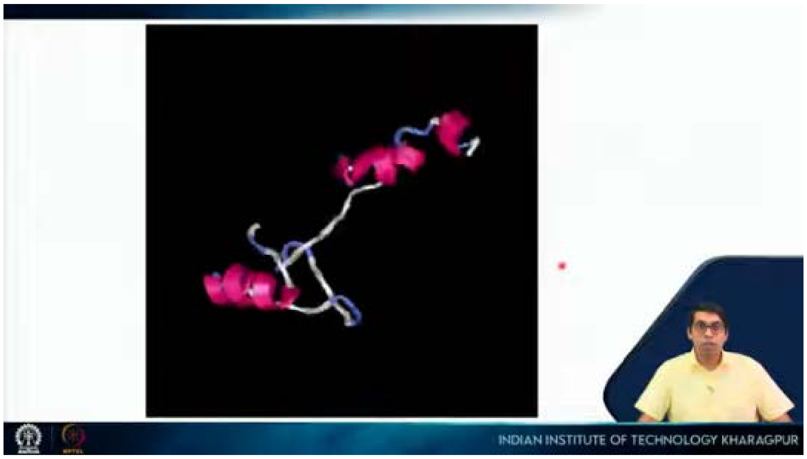
How proteins fold?

Now, one of the important questions is that is this a spontaneous process, meaning that the protein can do it by itself, or does it require help from other entities? We will see by the end of this lecture that the answer to this question is both are correct. So, here is a simulation of the folding of a small protein. So, you can see that it started with an extended structure. And here, what we have done is we can see that the secondary structures are shown in magenta and yellow, the alpha helix and the beta strand.



φ and ψ torsion angles are the only degrees of freedom for the backbone Plot

So, you can see that the extended structure has somewhat collapsed, and it is trying to form all these hydrogen bonds and tertiary contacts where these secondary structures are formed. So, you can see that the three beta strands are forming here, and now it is more collapsed, and finally, this is the final folded form. So, if you run this simulation again, you will see that alpha helices are the ones which form first, and the beta takes more time to form the beta strand and even more time to form the tertiary contacts to get the final folded form. The reason for that is, in the case of the alpha helix, the interactions are very localized because the hydrogen bonds are formed between amino acids which are separated by only 4. So, i+4.



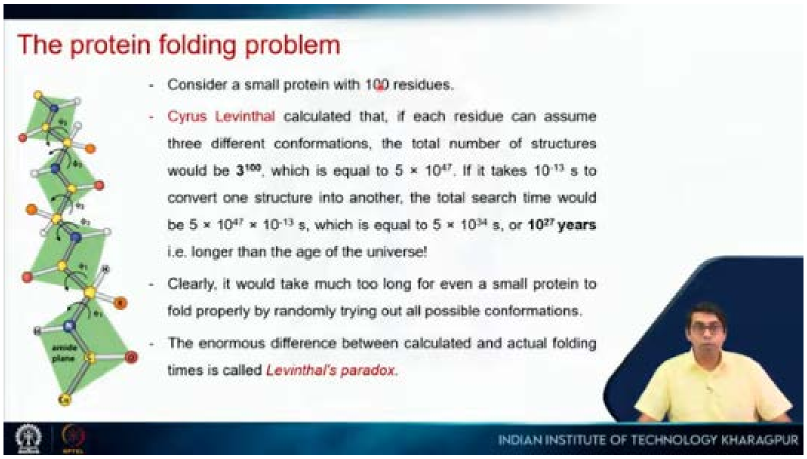
In beta strands, it can be a bit longer, and the tertiary contacts can be even longer. So, the probability of these interactions becomes lower. So, alpha helices form first, then the beta strand, and then finally all the tertiary contacts. So, now let us consider a very small problem; it is a thought problem where, let us say, we are thinking about a protein which is 100 residues long, and we know that for each amino acid, it can have various phi(φ)-psi(ψ) torsion angles. So, let us say in the Ramachandran plot, this phi(φ)-psi(ψ) torsion angle can occur either in the alpha-helical region or in the beta-stranded region or in the

left-handed helical region. So, for each amino acid, there can be three different phi-psi torsion angles or three conformations. So, if we consider a 100-residue protein, each amino acid will have 3 conformations, so for the first amino acid with 3 conformations, the second amino acid can also have 3 conformations, so together they can have 3 times 3, totaling 9 conformations. So, if we keep doing that for a 100-residue protein, we can have 3 to the power So, 3 times 3 times 3 goes on 100 times.

So, 3^{100} different conformations, which equals 5×10^{47}. It is a very big number. Now, from experiments, we know that these torsion angles, from going from one angle to another angle, take roughly 10^{-13} seconds. This is very, very fast. So, let us say the protein, if it samples all these conformations, and it takes this time to go from one conformation to another conformation, the total time taken will be this multiplied by this, which equals 5 × 10^{34} seconds or 10^{27} years, which is actually longer than the age of the universe.

Which means that if a protein which is only 100 residues long samples all possible conformations, it will never fold. And this is what is referred to as Levinthal's paradox, because this was first introduced by Cyrus Levinthal. So it means that, of course, proteins do not sample all possible conformations, like not even a small fraction of this. They sample very few conformations and go to the correctly folded form. Now, you might think that maybe I have taken up 100 residues as a big number, but that is not true.

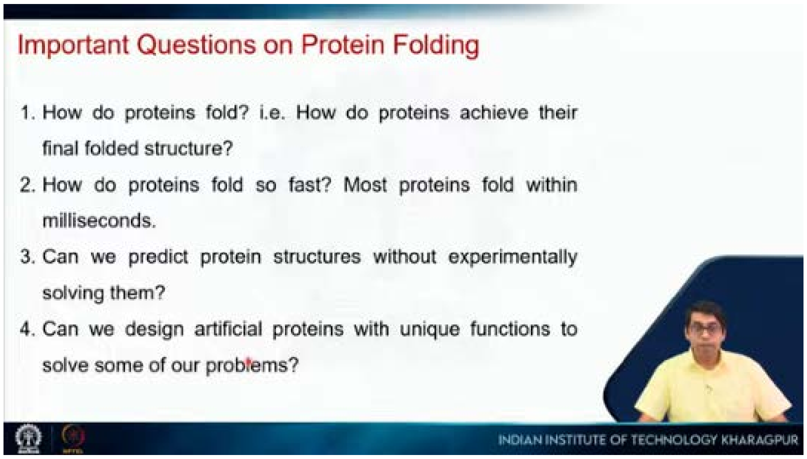
100 residues is a very small number. There are proteins which have 1000 residues. So, if we think about protein folding, we can divide all the issues and all the questions that are there into four major questions. The first question will be how do proteins fold? That is, how do proteins achieve their final folded form? How do proteins know that this is the final structure? And how do proteins fold so fast? How do they reach that final folded form so fast? Most proteins fold within milliseconds. We have already seen the simulation; it happens within microseconds or milliseconds. So it means that proteins are not sampling all the possible conformations. Then there is the third question, which is: can we predict protein structures without experimentally solving them?



The protein folding problem

- Consider a small protein with 100 residues.
- Cyrus Levinthal calculated that, if each residue can assume three different conformations, the total number of structures would be $\mathbf{3^{100}}$, which is equal to 5 × 10^{47}. If it takes 10^{-13} s to convert one structure into another, the total search time would be 5 × 10^{47} × 10^{-13} s, which is equal to 5 × 10^{34} s, or **10^{27} years** i.e. longer than the age of the universe!
- Clearly, it would take much too long for even a small protein to fold properly by randomly trying out all possible conformations.
- The enormous difference between calculated and actual folding times is called *Levinthal's paradox*.

This is a very important question because, as I mentioned earlier, protein structures can be solved by three different experimental techniques: X-ray crystallography, NMR spectroscopy, and cryo-electron microscopy. However, all these techniques take a lot of time, require a lot of expertise, and they are also very expensive. So, if we can theoretically model protein structures, then that will be very fast and also less expensive. And finally, can we design artificial proteins to solve some of our problems? So, there are so many issues that can be solved by designing artificial proteins.

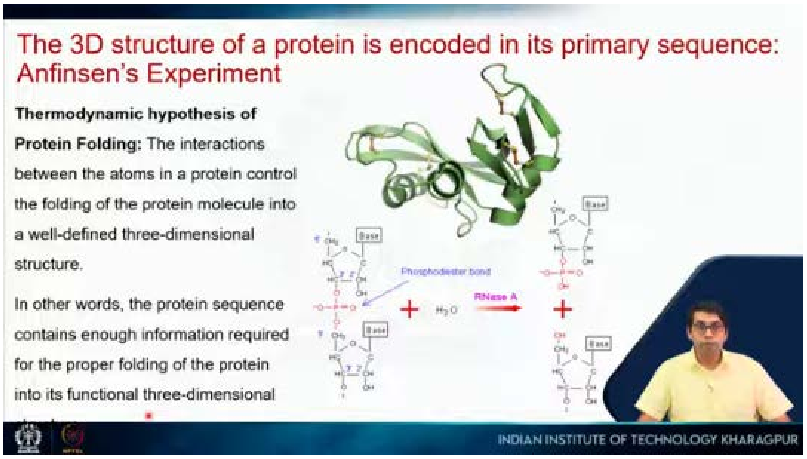


Important Questions on Protein Folding

1. How do proteins fold? i.e. How do proteins achieve their final folded structure?
2. How do proteins fold so fast? Most proteins fold within milliseconds.
3. Can we predict protein structures without experimentally solving them?
4. Can we design artificial proteins with unique functions to solve some of our problems?

So I will tell you about the current status of where we stand in regards to these four problems. So let us take the first problem. To answer that, a very elegant experiment was done by Christian Anfinsen in the 1960s and 70s, and that proved the thermodynamic hypothesis of protein folding. So what is the thermodynamic hypothesis of protein folding? It states that interactions between atoms in a protein control the folding of the protein molecule into a well-defined three-dimensional structure, or in other words, the primary sequence of the protein contains enough information that allows the protein to fold into its

functional three-dimensional form. So, to show what Anfinsen did, he and his colleagues used this particular enzyme called ribonuclease A. I will talk about enzymes in next week's lecture, and this enzyme catalyzes this particular reaction. So, this is an RNA; it catalyzes the hydrolysis of this RNA into its subunits. A particular amino acid will become very important in this, which is called cysteine. So, cysteine and cysteine.

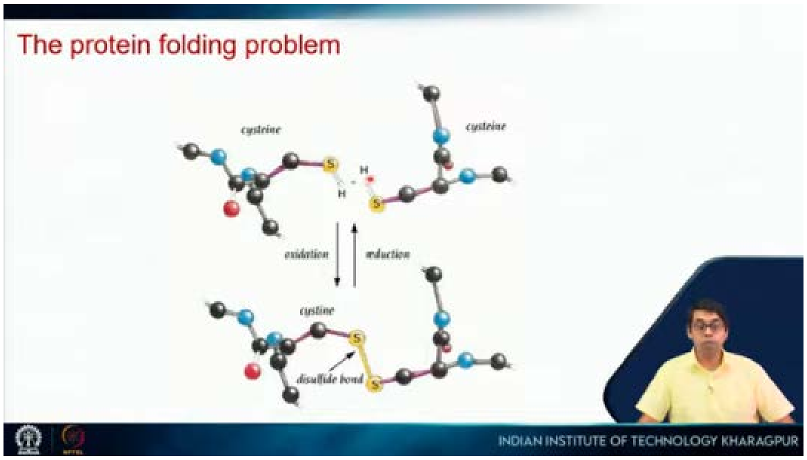


The 3D structure of a protein is encoded in its primary sequence: Anfinsen's Experiment

Thermodynamic hypothesis of Protein Folding: The interactions between the atoms in a protein control the folding of the protein molecule into a well-defined three-dimensional structure.

In other words, the protein sequence contains enough information required for the proper folding of the protein into its functional three-dimensional

So, it can occur in these two forms. So, cysteine's side chain has a thiol group or SH group. When this gets oxidized, it forms this disulfide bond, and when it is reduced, it goes back into this reduced form, which is the thiol form. So, ribonuclease A has 8 cysteines, which form 10 such disulfide bonds. So, what Anfinsen did was he took this ribonuclease A and checked its activity, and he found that the protein has a certain activity.

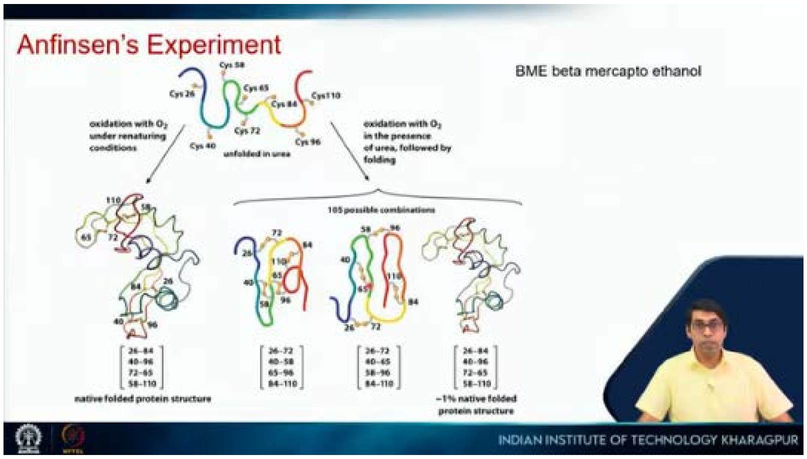


The protein folding problem

Then he used urea, which breaks all the hydrogen bonds and hydrophobic interactions. We will talk about this in a few minutes. So, it denatures the protein; it completely destroys the structure. And he also used a reducing agent like beta-mercaptoethanol to break all the disulfide bonds. So, the protein becomes this. So, it is not nicely folded like this; it becomes something like this.

And then he divided it into two parts. In the first part, he removed urea first and then bubbled oxygen, and then he checked the activity; he found 100 percent activity. In the second part, he removed urea first so that the protein would collapse, and then he bubbled oxygen first so that all the disulfide bonds were formed, and then he removed urea. So, in this case, he found only 1% activity. So, he is doing the exact same thing. He is removing urea and passing oxygen.

All he did was change the order. So here, he removed urea first and then oxidized. In this case, he oxidized first and then removed urea. Here, he got 100% activity. Here, he got 1% activity, which was an amazing result.



Anfinsen's Experiment

BME beta mercapto ethanol

Cys 58
Cys 26
Cys 65
Cys 84
Cys110
Cys 72
Cys 40
Cys 96
unfolded in urea
oxidation with O_2 under renaturing conditions
oxidation with O_2 in the presence of urea, followed by folding
105 possible combinations
110
58
65
72
84
26
40
96
26-84
40-96
72-65
58-110
native folded protein structure
72
84
26
110
65
40
58
96
26-72
40-58
65-96
84-110
58
96
40
110
65
84
26
72
26-72
40-65
58-96
84-110
26-84
40-96
72-65
58-110
~1% native folded protein structure
INDIAN INSTITUTE OF TECHNOLOGY KHARAGPUR

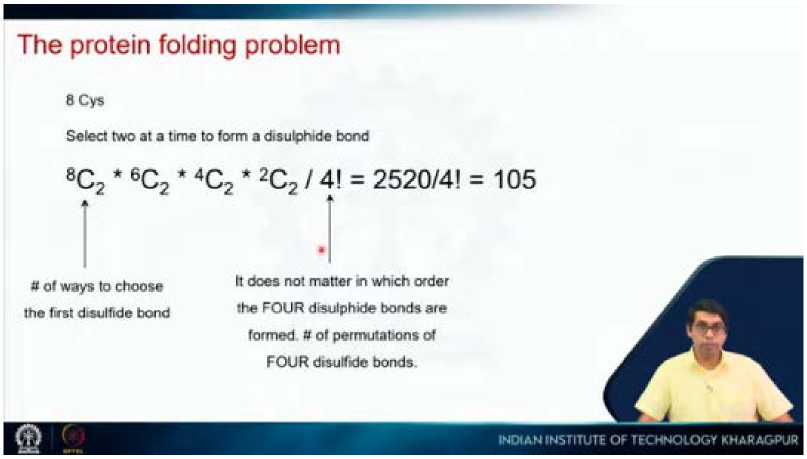
So why did this happen? So, let us go back to this. So, when he removed urea first, the protein collapsed into this structure, which is the final folded form. Then he passed oxygen, which means he oxidized, so all the cysteines which are close to each other formed a disulfide bond, and it locked the structure into this active form. So, he got 100% activity. But when he didn't remove urea first but oxidized the protein first, the protein was in this denatured form. So, cysteines can randomly form disulfide bonds. So, this cysteine 26 can form a disulfide bond with 58, 40, 72, or any of these, right?

So, all possible combinations of these four disulfide bonds were formed, and only one out of all these possible combinations was correct. Only that one was able to fold properly when urea was removed; the remaining could not. It turns out that if we do our calculations correctly, there are 105 different forms in which these four disulfide bonds can be formed. And only one of those is correct. So, one out of 105 gives you 1% activity. So, how do we get this 105?

$$^{8}C_{2} * {}^{6}C_{2} * {}^{4}C_{2} * {}^{2}C_{2} / 4! = 2520/4! = 105$$

So, this is the calculation. We can form the first disulfide bond. So, we are choosing two cysteines to form the first disulfide bond. So, ${}^{8}C_2$ will be the first. ${}^{6}C_2$, now we are left with 6 cysteines. So, we are choosing 2 out of those 6, ${}^{6}C_2$, ${}^{4}C_2$, and ${}^{2}C_2$.

So, that is how we form the 4 disulfide bonds. Now, it does not matter in which manner we form these 4 disulfide bonds. So, which disulfide bond is formed first? So, we have to divide it by factorial 4, and if you do your calculation, you will get 105. So, 105 is the number of different ways in which 4 disulfide bonds are formed in the unstructured protein, and only one of those will be correct.



The protein folding problem

8 Cys

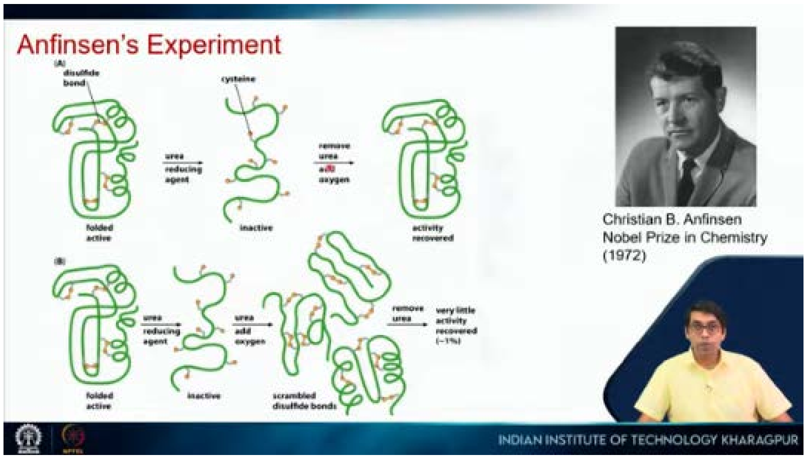
Select two at a time to form a disulphide bond

$${}^{8}C_2 * {}^{6}C_2 * {}^{4}C_2 * {}^{2}C_2 / 4! = 2520/4! = 105$$

# of ways to choose the first disulfide bond

It does not matter in which order the FOUR disulphide bonds are formed. # of permutations of FOUR disulfide bonds.

Which is again shown here. So, only one of those will be correct; all these different disulfide bonds are formed, but only one is the correct form. So, if I add oxygen first and then remove urea, I will get 1% activity but if I remove urea first and then add oxygen, I will get 100% activity. So, what this experiment showed is that this protein can fold on its own; it does not need any help. So, all the necessary information is present here, and this was actually shown by Christian Anfinsen, who got the Nobel Prize for this work in 1972. Since proteins can fold and since all the information is present in the primary sequence of the protein, can we predict protein structure if we know the sequence of a protein?

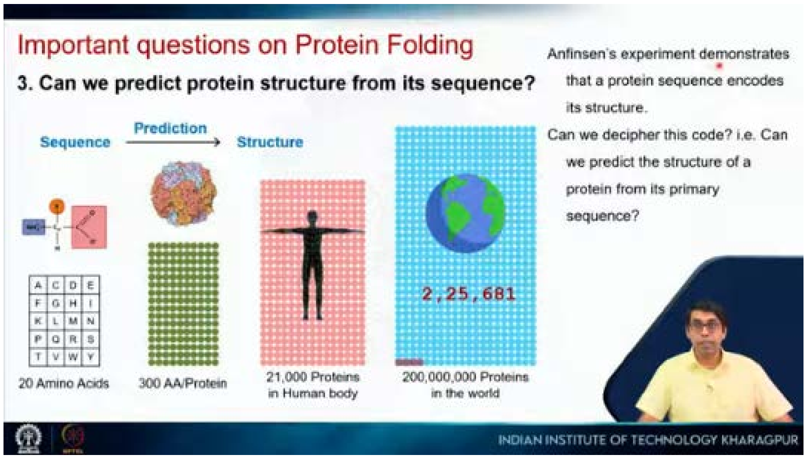


Anfinsen's Experiment

Christian B. Anfinsen
Nobel Prize in Chemistry
(1972)

INDIAN INSTITUTE OF TECHNOLOGY KHARAGPUR

And that is very important because it turns out that these are just some numbers. So, proteins are made up of 20 amino acids; in bacteria, the average size of a protein is 300, so there are 300 amino acids per protein. In humans, we have 21,000 proteins. And if we consider all proteins on this planet, there are 200 million proteins. Right now, when we solve protein structures by experiments, it takes time, and in the last 70 years, we have solved only 225,000 protein structures, which is a very small amount out of all these possible protein structures, so it will be very helpful if we can predict structure from the primary sequence of a protein. So, Anfinsen's experiment has already shown that the information is there; can we use this information? And this is exactly what many scientists have been doing for the last 20-30 years. So, there is this particular worldwide competition that happens every two years since 1994; it is called the Critical Assessment of Structure Prediction.



Important questions on Protein Folding

3. Can we predict protein structure from its sequence?

20 Amino Acids

300 AA/Protein

21,000 Proteins in Human body

200,000,000 Proteins in the world

Anfinsen's experiment demonstrates that a protein sequence encodes its structure.

Can we decipher this code? i.e. Can we predict the structure of a protein from its primary sequence?

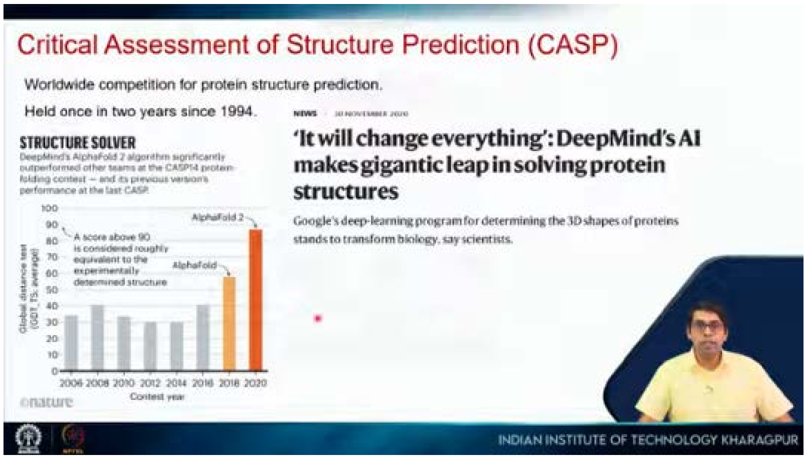
INDIAN INSTITUTE OF TECHNOLOGY KHARAGPUR

So, here, all these groups worldwide who are working on this problem have developed their algorithms. So, they are given sequences whose structures have been solved by different groups, but they have not been released. So, they are given the sequences and asked to predict the structure, and then the predicted structures are compared with the experimental structure, and then some score is given. A 90% score is the goal of this competition. If you can get something that is close to 90%, it means that your predicted structure is as good as the experimental structure. So, over the years, you can see that these are the best groups, so these are the best performances by any group, and it never crossed 40%.

So that was the best that was achieved. In 2018, DeepMind entered this competition and introduced something called AlphaFold. It uses artificial intelligence, and you can see that it reached almost 60%. Then, in 2020, the second version of AlphaFold came out, and they achieved almost 90%. Accuracy in prediction.

So this is as good as an experimentally determined structure. This created a lot of interest. It was shown, it was predicted that this would change everything. This AI AlphaFold tool makes a gigantic leap in solving protein structures. In the last four years, people have analyzed this.

Many more algorithms have come up similar to this. It turns out that AI can actually solve this problem very efficiently. So we can say that out of these four major questions, this third one is now solved. But then it does not answer the question of how proteins fold because what AlphaFold does is it takes the sequence and gives you the final folded structure. It does not tell you how proteins fold, how it reaches that structure so fast, or what the interactions are by which it reaches there.



Critical Assessment of Structure Prediction (CASP)

Worldwide competition for protein structure prediction.

Held once in two years since 1994.

STRUCTURE SOLVER

DeepMind's AlphaFold 2 algorithm significantly outperformed other teams at the CASP14 protein-folding contest — and its previous version's performance at the last CASP.

100
90
80
70
60
50
40
30
20
10
0
Global distance test (GDT_TS, average)
A score above 90 is considered roughly equivalent to the experimentally determined structure
AlphaFold
AlphaFold 2
2006
2008
2010
2012
2014
2016
2018
2020
Contest year
©nature

NEWS · 30 NOVEMBER 2020

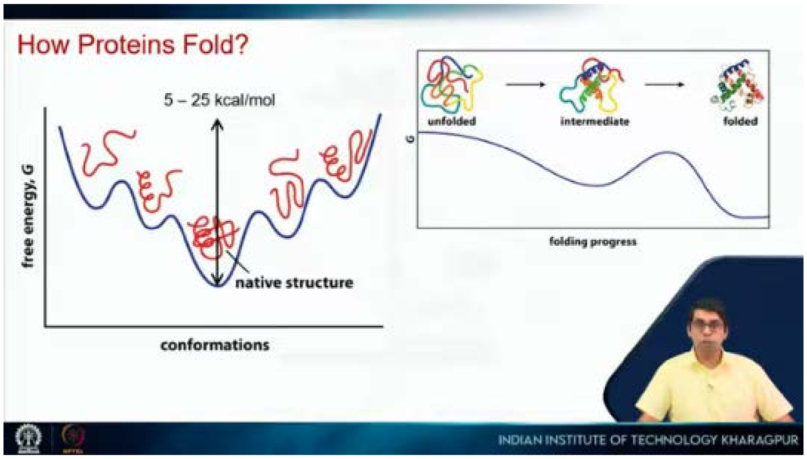
'It will change everything': DeepMind's AI makes gigantic leap in solving protein structures

Google's deep-learning program for determining the 3D shapes of proteins stands to transform biology, say scientists.

INDIAN INSTITUTE OF TECHNOLOGY KHARAGPUR

So if we think about protein folding, we can think in terms of energetics. These are the different conformations, and this is the free energy. Unfolded structures will have higher energy, and the folded structure will have the lowest energy. We can have all these different structures, and this is the final folded form. It turns out that the difference between unfolded and folded is not much; it is somewhere between 5 to 25 kilocalories per mole.

In terms of time, we can think of it like this: the unfolded structure looks like this, and over time, it goes through all these different intermediate steps and then achieves this final folded form. So, we have seen something similar in the simulation that we saw a few slides back. So, what stabilizes this final folded form? These are the few important interactions that stabilize the folded form of a protein. There are hydrophobic interactions.

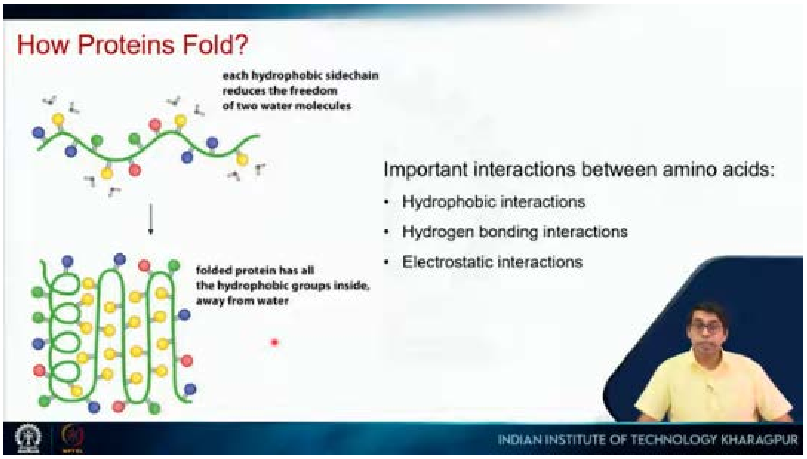


How Proteins Fold?

INDIAN INSTITUTE OF TECHNOLOGY KHARAGPUR

We have already seen that amino acids, including hydrophobic amino acids like those with aliphatic side chains or aromatic side chains, exist. They form hydrophobic interactions, and it turns out that these hydrophobic side chains tend to go inside the core of the protein because they do not like to interact with water. Then there are hydrogen bonds, which we have already seen in the case of secondary structures. Alpha helices have a pattern of hydrogen bonds, and beta strands have another pattern of hydrogen bonding interactions. Additionally, there are electrostatic interactions between charged groups. If this is a negatively charged side chain and this is a positively charged side chain, they can interact with each other and stabilize this final folded form. They can also interact with water and with each other. So, hydrophobic interactions, hydrogen bonding patterns, and electrostatic interactions together stabilize the final folded form of the protein. What drives this final folded form?

These interactions and also something called hydrophobic collapse. So, it turns out that when in this unstructured form, water molecules close to this protein have restricted motions, so they have restricted degrees of freedom. Additionally, these hydrophobic side chains do not like to interact with water molecules; they prefer to interact with each other, which is also energetically favorable. So, together, the driving out of water molecules from here and the interaction of these hydrophobic residues is something called hydrophobic collapse, and that drives protein folding. So, the protein will not sample all possible conformations.



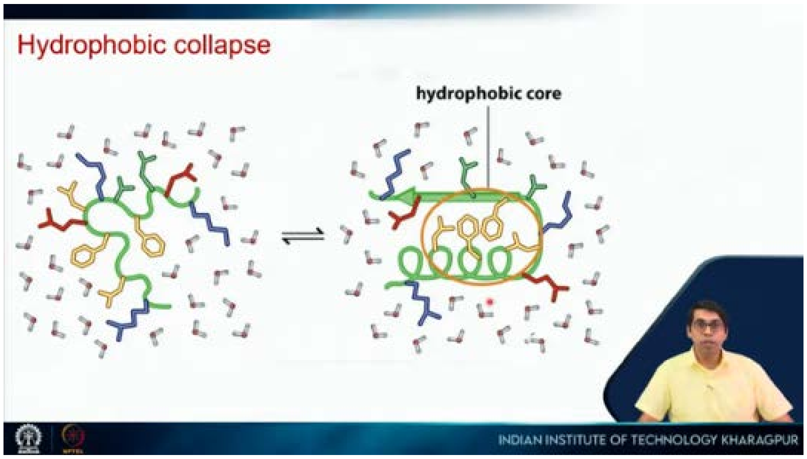
How Proteins Fold?

Important interactions between amino acids:

- Hydrophobic interactions
- Hydrogen bonding interactions
- Electrostatic interactions

each hydrophobic sidechain reduces the freedom of two water molecules
folded protein has all the hydrophobic groups inside, away from water
INDIAN INSTITUTE OF TECHNOLOGY KHARAGPUR

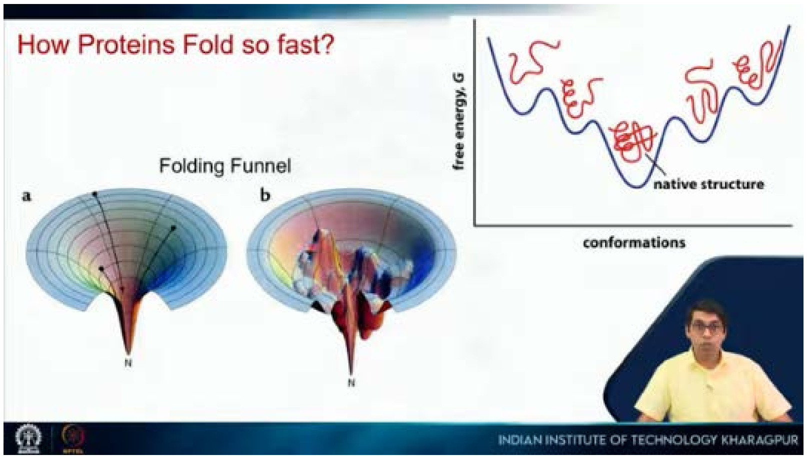
It will quickly go from this to this because of hydrophobic collapse, and we will have this final folded form. So, this sampling of only a few conformations is very elegantly shown by something called the folding funnel. So, we have seen this energy diagram. So, free energy is plotted on the y-axis and conformations on the x-axis. So, instead of just one axis, we can have two axes, x and y, for conformations, and the z-axis can be for the energy, and that will give us a diagram like this, which is called a folding funnel.



Hydrophobic collapse

INDIAN INSTITUTE OF TECHNOLOGY KHARAGPUR

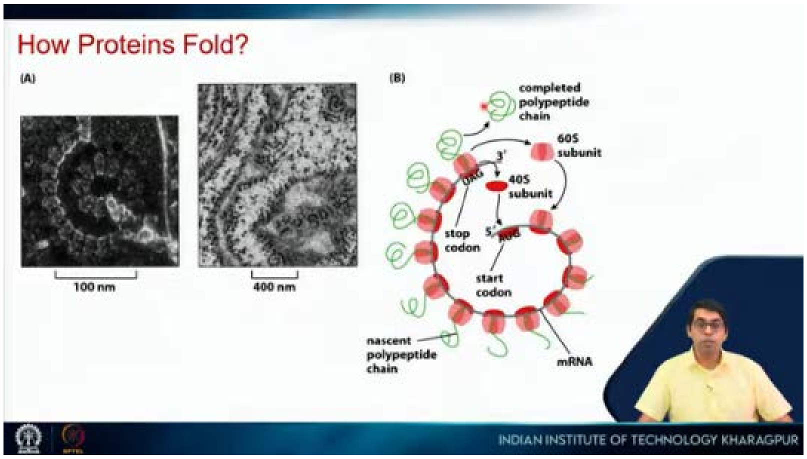
So, let us say the unfolded structure is here, which is energetically very high. Instead of sampling all these possible conformations, the protein will quickly go through only these few conformations and reach the native folded form, which is the lowest energy state. So it can start with all these different conformations and quickly go to this native folded form. So, some proteins have a very smooth funnel like this, where the folding is very fast, and some funnels have this rugged funnel structure like this. So, the protein will go through all sorts of intermediate states and finally reach the final folded form.



How Proteins Fold so fast?

INDIAN INSTITUTE OF TECHNOLOGY KHARAGPUR

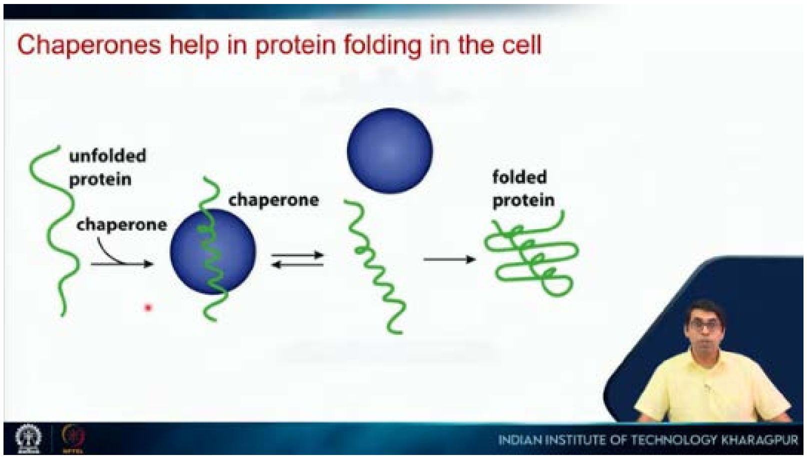
So, it turns out that if we think of bacteria, then this folding, this synthesis, and folding can happen on the ribosome itself. Because now we know that proteins can fold by themselves, so while the protein is getting synthesized, it will start folding, and then the complete folding happens here. So this is the final folded form of the protein, and this is true for the most small proteins. However, in some cases, if there are slightly bigger proteins or proteins with a lot of hydrophobic residues, these proteins might not fold properly. So, or it can happen that most of the proteins, most molecules fold properly, but there are some molecules which do not fold. So in that case, extra help is needed.



How Proteins Fold?

(A)
100 nm
400 nm
(B)
completed polypeptide chain
60S subunit
3'
UAG
40S subunit
stop codon
5'
AUG
start codon
nascent polypeptide chain
mRNA
INDIAN INSTITUTE OF TECHNOLOGY KHARAGPUR

And that help is provided by other types of molecules. These are also proteins called chaperones. So chaperones bind to this unfolded protein, and they prevent their aggregation. Because when these proteins are unfolded, hydrophobic residues are exposed, so hydrophobic residues from one chain can stick to the hydrophobic residues of another chain, and they can form aggregation and precipitate. So they will not get a chance to fold. So that kind of aggregation is prevented by the chaperones. They keep them separate and give them enough time to fold properly and get into the folded state. Apart from that, there is another type of protein called GroEL and GroES. So this is mRNA. This is a ribosome.



Chaperones help in protein folding in the cell

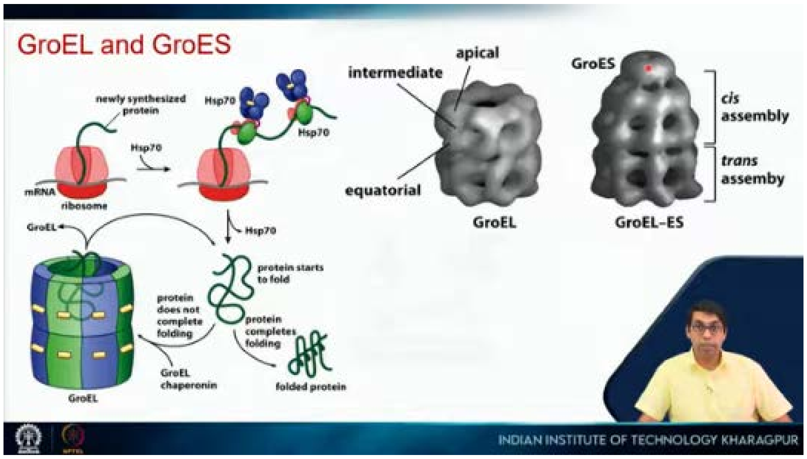
Protein is getting synthesized. Now the chaperones, in this case, we are seeing one particular type of chaperone called HSP70. So HSP stands for heat shock protein, and 70 is its molecular weight. So it is 70 kilodalton in size. So this chaperone binds it and prevents it from aggregating.

Now the chaperone will go out, and it will allow this protein to fold, and it will complete its folding. Some protein molecules might not fold properly, and they will need additional help, which is provided by this large protein complex called GroEL. So GroEL looks like this huge barrel. It has 14 subunits.

So you can see there are two rings, up and down. So the upper ring has 7 subunits, and the lower ring has 7 subunits. So in total, there are 14 subunits. And each of these subunits binds to this yellow thing, which is an ATP. So ATP is hydrolyzed to ADP, which provides energy, and that energy is used to change the shape of this big barrel.

So what does it do? I'll come to that. So this is a cryo-EM structure of GroEL. And you can see that it looks very similar to this. And then on top of that, there is this small structure which looks like a lid.

This is GroES. So GroEL and GroES help in protein folding like this. So this is the upper barrel, and this is the lower barrel. So in the upper barrel, this is GroEL. It takes in the unstructured protein.



GroEL and GroES

INDIAN INSTITUTE OF TECHNOLOGY KHARAGPUR

This is a misfolded or unfolded protein. It goes completely inside the barrel. And then it will bind seven molecules of ATP. Each subunit will bind one ATP, and the GroES forms the lid. So the lid comes here.

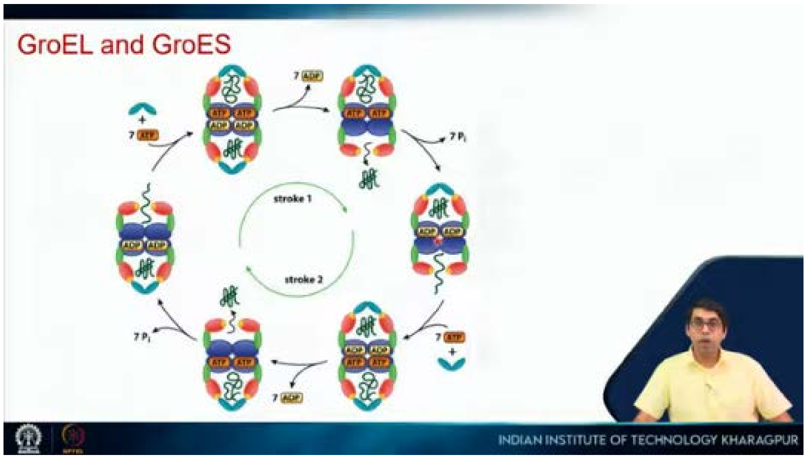
This provides a completely separate environment for this protein molecule. And this is also referred to as an Anfinsen cage. This prevents the interaction of this protein with other protein molecules. So hydrophobic-hydrophobic interactions are prevented here. Now, inside of this

chamber, it is hydrophobic to begin with. Once this lid is closed, the seven ATP molecules are hydrolyzed to ADP, and that energy is used to change the conformation of the seven subunits, where initially the inside is hydrophobic and now it becomes hydrophilic. So this mimics the aqueous environment. Now since this is hydrophilic, it will induce this hydrophobic collapse because the hydrophobic residues would want to go out of this environment, and they will stick to each other. So, hydrophobic collapse happens, and the protein will fold. Now we get this properly folded protein. And finally, this folded protein is released from this chamber. So this is what happens in the top chamber, right?

And you can see the same thing happens in the bottom chamber also. So GroEL and GroES act like a double-stroke engine. So, in one stroke, the top chamber is folding one molecule. In the second stroke, the bottom chamber is folding another molecule.

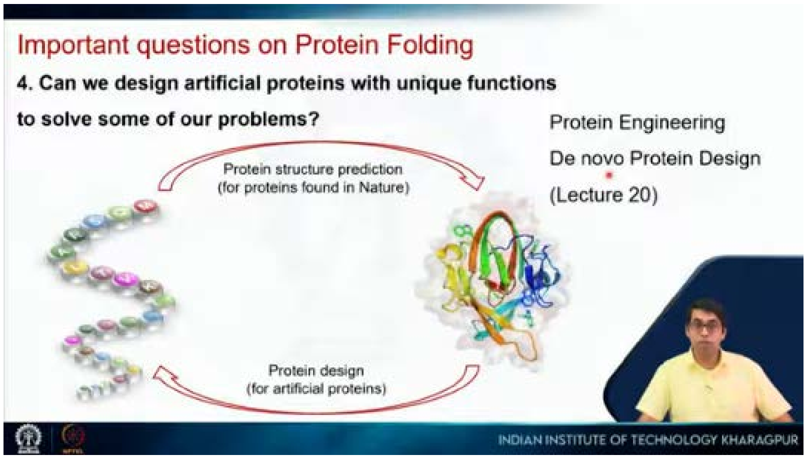
So what GroEL and GroES do is provide an isolated environment for the protein to fold properly. The basic principle is exactly the same. The sequence of the protein determines its function and its final folded form. GroEL and GroES are not going to change the final folded form. They are not going to change the thermodynamics of folding.

All they are doing is preventing aggregation and providing an isolated chamber where the protein can fold. So, what we have seen so far is we know how proteins fold, we know how they fold fast, we can explain it by a folding funnel, but again, I will not say these are complete, a lot of research is still going on regarding these two questions. The third one is more or less complete, and then comes the fourth important question. Since we know all of this, can we use this knowledge to design artificial proteins with unique functions to solve some of our problems? So, let's think of it like this. The third problem was, if we know a protein sequence, can we predict its structure? This is for the natural proteins which are found in nature. The fourth problem is sort of the reverse of this. We have some function in mind, and based on that, let's say we have designed some three-dimensional structure of a protein. That's the machine we think will perform this function. Now, can we design a sequence? Which will fold into this three-dimensional structure.



GroEL and GroES

So, we have some knowledge of this structure, or we have designed this structure. Now, can we predict the primary sequence which will fold into this three-dimensional structure, right? So, this protein design or artificial protein design, this problem can be divided into two parts. The first one is called protein engineering, where we take naturally occurring proteins and make changes to them so that instead of performing their natural function, they will do something else, something that we want them to do. So, that is protein engineering. We are taking an existing protein and engineering it to perform a new function. And the second one is called de novo protein design. So, de novo protein design means that this sequence does not exist.



Important questions on Protein Folding

4. Can we design artificial proteins with unique functions to solve some of our problems?

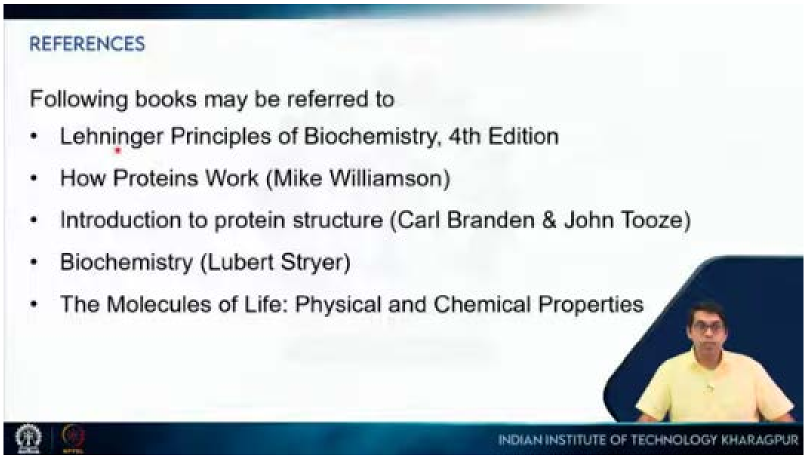
Protein Engineering

De novo Protein Design

(Lecture 20)

Protein structure prediction
(for proteins found in Nature)
Protein design
(for artificial proteins)
INDIAN INSTITUTE OF TECHNOLOGY KHARAGPUR

We are completely designing a new structure with a new function. So, I will talk about these in more detail in Lecture 20. So, that will be the last lecture next week. So again, these are the books that I have followed. You can follow any one or more of these books.



REFERENCES

Following books may be referred to

- Lehninger Principles of Biochemistry, 4th Edition
- How Proteins Work (Mike Williamson)
- Introduction to protein structure (Carl Branden & John Tooze)
- Biochemistry (Lubert Stryer)
- The Molecules of Life: Physical and Chemical Properties

So, that is all for now. Thank you.