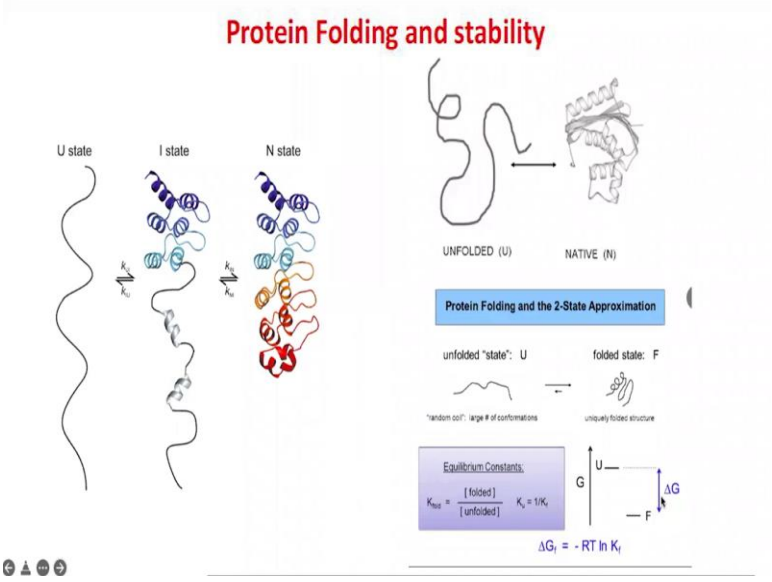
Introduction to Cell Biology
Professor Girish Ratnaparkhi
Professor Nagaraj Balasubramanian
Department of Biology
Indian Institute of Science Education and Research, Pune
Protein Structure, Folding and Function - Part 4

(Refer Slide Time: 00:15)



So, at this point, you have a fair idea of the main chain structure of proteins, you have an idea of conformational space, you have an idea of what folded proteins look like, even though the representation shown to you are cartoon representations of real atomic structures, which have been solved, maybe ball and stick arrangements or ribbon arrangements, but they are all representative what a protein looks like.

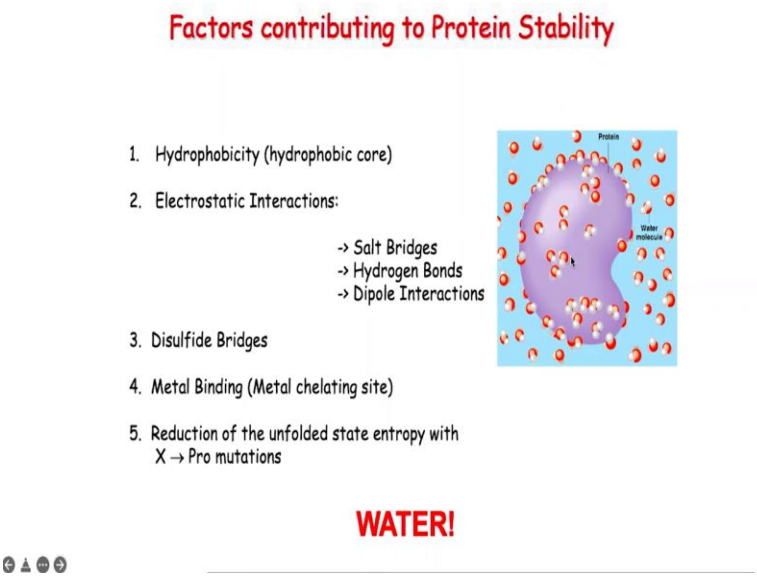
So, let us now transit over to the idea of protein folding and the idea of protein stability. So, now, this is a representation of a unfolded state, a linear sequence of amino acids. And on the right side is a so called folded state and we get these folded states from crystallographic experiments. And we know that during the process of unfolded state going to a folded state, there are these what we call as intermediate or partially folded states.

And as the protein goes from its unfolded state to a precise folded state, a unique folded state, it samples a lot of conformational space. And part of the sampling involves it making secondary structures and tertiary structures, which are not the correct structure, and finally finding its way to the correct state.

Now, I am not going to go great detail, because that is part of the biochemistry course you will be going through in your third year. But basically we say there is an unfolded state and there is a folded state. And we can talk about a lot of protein stability, measuring how stable a protein is, by making this assumption that there are only two states. This is not completely true. But this is an assumption we make that the intermediate states are transient and they really are not stable for long periods of time.

Now, with this two state approximation, we basically can have an equilibrium constant K, which relates to the concentration of the folded state over the unfolded state. And if we use this equilibrium constant K, we can get a Gibbs free energy value of using this formula, which anybody who is done basic chemistry is familiar with, which talks about the difference in the Gibbs free energy, which is called delta G, between the unfolded and the folded states. And this difference, which we can measure by experimentation, tells us how stable a folded state is over the unfolded state. So, it is a very focused thing I am going to talk about.

(Refer Slide Time: 02:52)

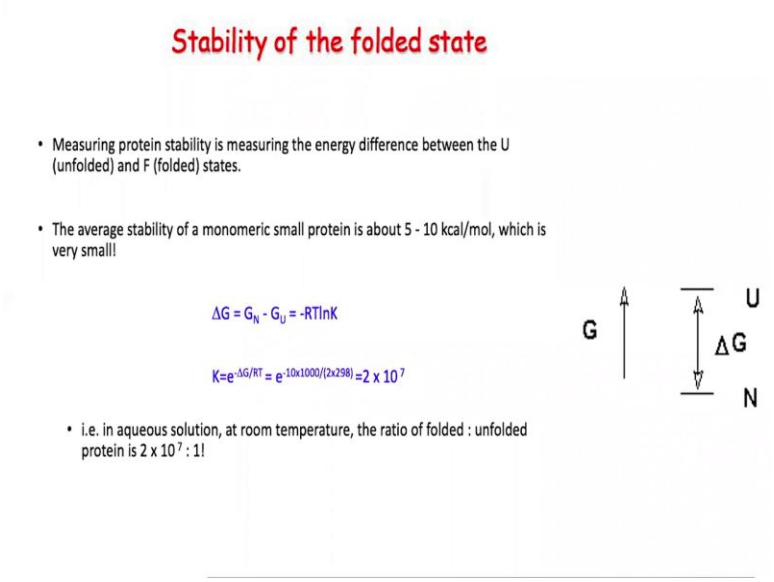


Now, this delta G defines the ratio of folded states over unfolded states, and this delta G usually is negative, because folding is a spontaneous reaction. You put an unfolded polypeptide in water and it will go and fold in a very fast timescale. This is in the microsecond time scale. And there has been a lot of research over the very fuzzy nature of why a protein is stable, what makes it stable, what are the stabilizing forces for the folded state.

And these are some of the actors which have been talked about hydrophobicity, electrostatic interactions, including hydrogen bonds, salt bridges, disulfide bridges, which you would have heard about, metal binding, very few proteins have that. And many others where there are at least half a dozen to a dozen things listed in the different forces which stabilize proteins.

And I will emphasize here and it is very important that, whenever we talk about protein folding, we are always talking about protein folding in water. Water is always around and water has a major role to play in protein folding.

(Refer Slide Time: 04:07)



So, as I have told you that you can talk about a folded state by looking at the energy differences between the unfolded and the folded or the native states. And what is done over here is, if I just plug in the average numbers for temperature, for the gas constant, for a protein, which is about let us say 10 kilo cals per mole, because that is what is the, is what the experimental measurements are telling us that the folded state is stable over the unfolded state by about 10 kilo cals per mole.

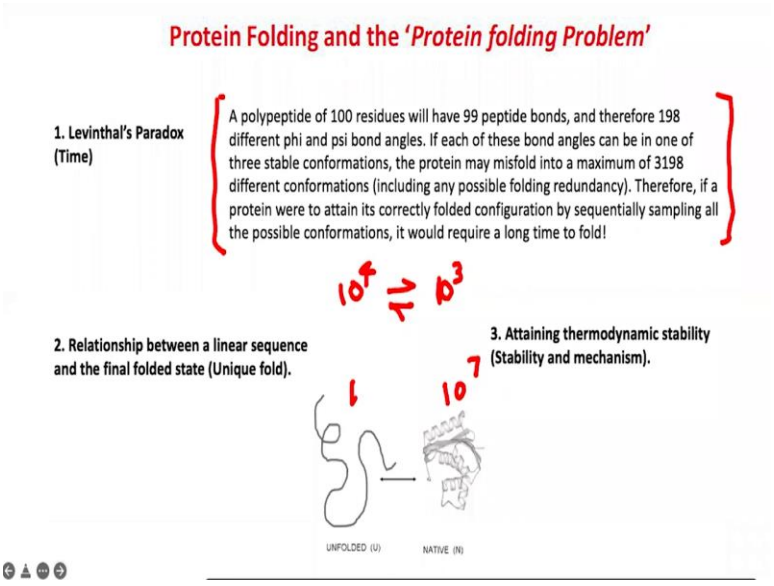
And the number we get is about 2 raise to 10 raise to 7 which tells you that at any point in solution for protein which we have experimentally verified as 10 kilo cal per mole, they are 2 into 10 raise to 7 or 10 raise to 7 molecules which are folded and one molecule which is unfolded, and the folded and unfolded states are pretty much always in equilibrium with each other, which means there is a dynamic equilibrium. The unfolded state goes to the folded state,

and the folded state goes back to the unfolded state. But there is a heavy bias towards the folded state as compared to the unfolded state for average protein, which is about 10 kilo cal per mole.

Now, 10 kilo cal per mole it seems very nice these numbers that a large number of molecules are always folded as compared to the unfolded state. But the strange thing is when you start thinking about what these energy values mean. Now, hydrogen bond is about 1 to 3 kilo cals per mole. Basically, if you break a hydrogen bond, that is pretty much the amount of energy you release.

Now, a covalent bond is to the scale of about 80 to 100 kilo cals per mole. So, it is very strange that a protein which is stable to the level of about 10 kilo cal per mole, the value is so low that compared to a covalent bond, it appears to be a very weak stability over the unfolded state. In this weak or easily disruptable stable nature of the folded state has been a matter of conversation for the past 40 to 50 years. And I will talk about it a little bit more in the next few slides.

(Refer Slide Time: 06:15)



Now, for many years, people have talked about what is called as the protein folding problem. And the protein folding problem has many aspects. And I will just list these. So you have an unfolded state. The unfolded state goes to the folded state. It goes through certain intermediates, and this is a kinetic process. And it finally ends up with a folded state. And at equilibrium, you have let us say and this is an approximation, you have one molecule of this and 10 raise to 7 molecules of this.

Now, when the protein is made of the ribosome, it is an unfolded state, and it has go to these folded states. So, part of the Levinthal's paradox, and I will tell you why it is a paradox is what are the pathways, which are, which the protein goes through to make the folded state and over here is just a small calculation. I have just cut and pasted it from the net, which talks about the fact that since for a protein of 100 amino acids, there are so many different phi-psi angles.

And if a protein has to come to the right combination of phi-psi angles, which define the folded state, and it explores each one of these phi-psi angles even in, let us say, three confirmations, forget 360 confirmations, the amount of time, even if you give nanoseconds or pico seconds to each determination is greater than the time of, since the start of the universe, since the start of the Big Bang.

So, the numbers are huge. And somehow magically a protein can very quickly fold into its native state and define a set of phi-psi values which define its folded state. So, the protein folding problem incorporates the uncertainty about what pathways is a protein using. It is about the time required for a protein to fold. Protein can fold in microseconds. How does a protein find its native conformation so quickly, knowing that the unfolded state is itself not a single state, it has multiple unfolded stage. It is like a snake which is vibrating all the time. A snake can be in many, many different positions, but all positions lead to this final state.

We know from work by Anfinsen and I will talk about Anfinsen in a minute, that the linear sequence of amino acids is very important to determine this particular unique state. If you take random amino acids, you are not going to get it to fold into a unique state. It will sometimes remain in unfolded or partially folded state.

And the last aspect of the protein folding problem is the fact that average protein is stable to around 10 kilo cals per mole, even though the number of molecules compared to the unfolded state are very, very high. But if you break one covalent bond, and remember, there are many, many covalent bonds over here, you effectively are destabilizing the protein back to the unfolded state. If you break two hydrogen bonds, you lose about 8 kilo cals per mole, 6 kilo cals per mole, and immediately, instead of having 10 raise to 7 molecules in the folded state, we will have 10 raise to 3 and 10 raise to 4. So, you are basically changing the equilibrium again.

And the moment you have very few molecules in the folded state, the activity of the protein is affected because the activity depends on the folded state of the protein. So these are all these aspects, which for the past 5-10 years we have been trying to understand understand. There is a question.

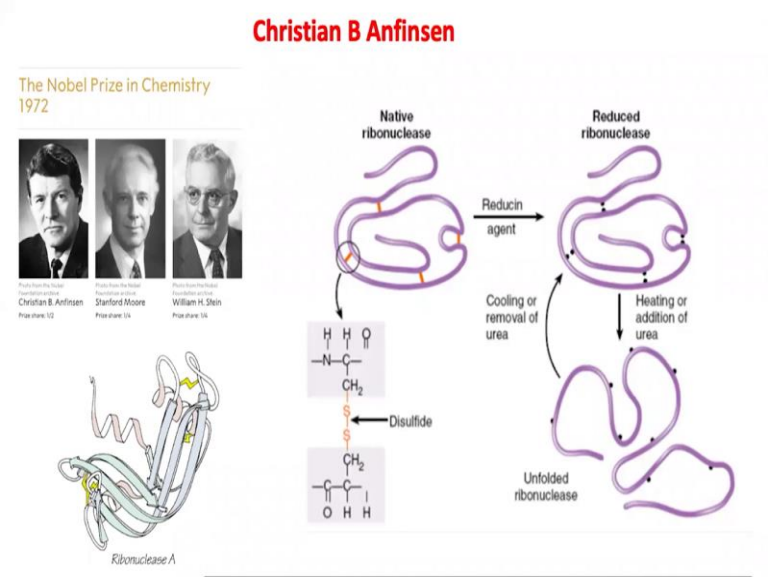
Student: Please repeat this part, this breakage of bond from which…

Professor: So, the protein folding problem has multiple aspects. One is the linear sequence of amino acids defines the folded state. You cannot have a random sequence of amino acids. Hemoglobin has a certain defined sequence. If you change the sequence, it will not fold into hemoglobin and it will not do its function.

Second part is time, there are so many combinations of protein has to explore around the phi-psi angles as an example as it folds. How does a protein fold so quickly, that is a sort of a mystery. The third part is that if you measure the stability of a protein, it is only about 5 kilo cals per mole. But breaking two hydrogen bonds will remove, will basically make it 0 kilo cals per mole. And you need this free energy of stability to have a decent number of folded molecules in solution.

So, it is weakly stable. It can be, it seems to be very finely tuned and balanced with even a breakage of a few bonds, converting it back to the unfolded state. The unfolded state is unfortunately not a functional state. So, it is not the state you want the protein in your body to be in. And a single mutation can destabilize the protein enough to take it back to the unfolded state. And if most of the molecules of hemoglobin in your body are in the unfolded state, they would not be able to bind iron and if they do not bind iron they would not be able to carry oxygen to your blood and you are going to die and this is true for every functional protein in your body. Is that clear?

(Refer Slide Time: 11:22)



So, now, there was a Nobel Prize in, for discovering the relationships between the amino acid sequence given to Christian Anfinsen in 1972 for work he did in the 1950s and 60s. And he took a protein called ribonuclease. And he basically broke the disulfide bonds. Then he either put the protein in a denaturant like urea or he heated it to 80, 90 degrees, then he cooled it and he found that after a certain period of time, the molecule could refold back to its native state, form its disulfide bonds, and do its activity. And the activity of ribonuclease is it cuts, chops RNA into small pieces. So, he basically found a pure solution of protein. Ribonuclease is his example.

When denatured and renatured could retain go back to its native state, which meant that the information for the folded state of a protein was contained within the sequence of amino acids. And he got a Nobel Prize for this discovery. This is still true, but many, many proteins in the cell cannot fold by themselves. They need help. And the help which is given to them is by a class of proteins called as chaperones, which I am not going to really talk about at this point.

Student: Sir, can you repeat that last slide again? I did not catch it.

Professor: So, just work from the 1960s showing that if you take ribonuclease, you denature it by temperature and then you reduce the temperature back to room temperature, the ribonuclease will fold back to its native state. And the native state is the state which has the ability to cut RNA. The unfolded state cannot cut RNA. So, denaturation and renaturation of proteins was possible

in-vitro without the addition of anything else indicating that the sequence of amino acids had the information necessary to help refold a protein.

Student: So, what is this addition of urea?

Professor: It is just a denaturant. So, you need to unfold the protein in some way right. The protein will remain folded in solution till you force it to unfold. Under normal conditions a protein will not unfold. So, if you, to force it to unfold, you can take the temperature to 80 degrees. At 80 degrees, it will unfold. Or you add denaturant. Denaturant will cause it to unfold. But when you remove denaturant or bring the temperature back to room temperature either cases, the protein will fold back to its final state and start becoming active again, without having any other information apart from the linear sequence of its amino acids.

Student: Disulfide bond formation takes place again.

Professor: Yes, so for that you have to add a tiny bit of an agent. This is not something I am emphasizing on. But, yes, you do need to give a little bit of assistance. So, in Anfinsen's experiment, the refolded protein was not, refolded protein, when you added reducing agent, the refolded protein was not efficiently got back into the folded state. However, if you do not add a reducing agent, there is no problem at all. The protein would not unfold completely. But when you denature it by heat, it will lose its native state, but it will quickly regain it because disulfide bonds are not broken just by increasing temperature.

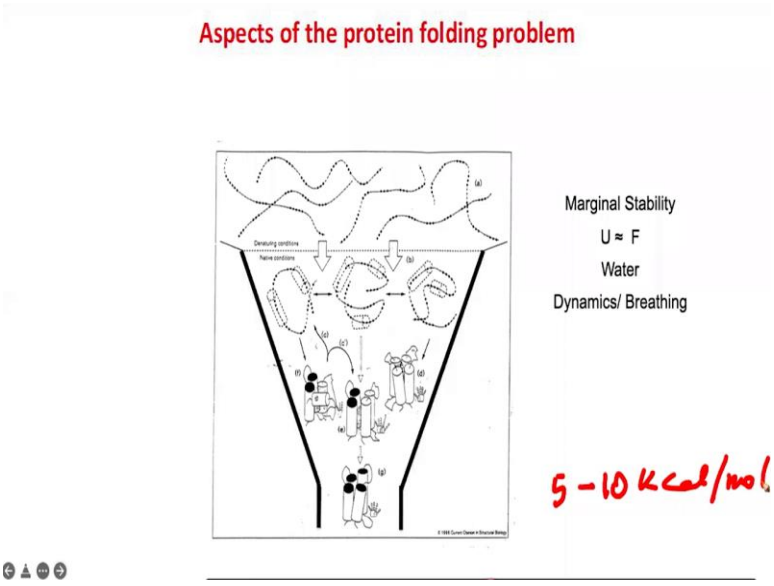
Student: Excuse me, sir. Sir, I do not know if you cleared this confusion or not because there was some connection error on my side, but I will ask this question again. Sir, we learned that, if the temperature is increased and the protein is denatured once it cannot regain its activity again, like it is destroyed then. So, how come the protein will again become active as you said?

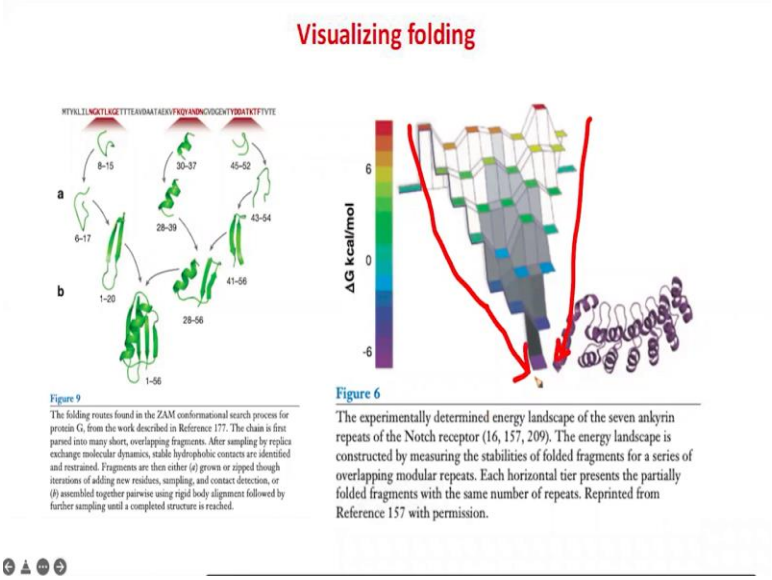
Professor: No. So, you see, if you take any small globular protein, let us take 100 amino acids, ribonuclease is about 123 amino acids, there are proteins which are 50, 60, 70, 80 amino acid range, many of them, most of them, not all of them, if you denature, you can renature. Indicating that, as a general thumb rule, this is what Anfinsen said, proteins can refold back their native states. But this is not true for very large proteins and many other proteins which are multi-

domain proteins, because these proteins, it turns out, cannot find their way back again, unless they are helped.

But if you provide them with the help, and the help is another protein called as a chaperone, which is sort of a, which gives moral support, which tells these proteins go on I know you can do it, I know you can fold, the proteins can still fold. So, we still agree even after 70 years after Anfinsen discovered the refolding abilities of ribonuclease, we still agree that the information for the three dimensional state of a protein is within the amino acid sequence of that protein and amino acid sequences of proteins are fairly unique, and they relate to the folded state.

(Refer Slide Time: 16:40)





So, now, very quickly, how do you define part of the protein folding problem, which is now the fast folding of a protein? And the current view, I am telling you the current view, is the following. I am showing you a picture on the left hand side.

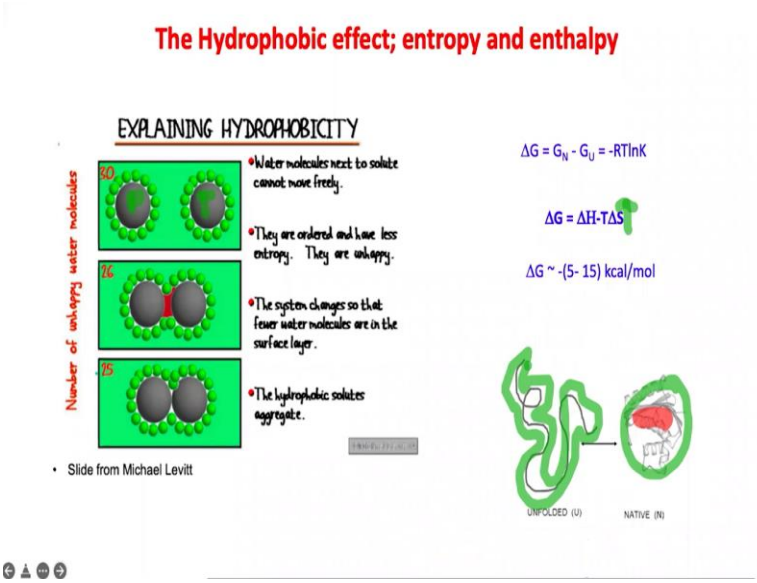
The current idea is that as an unfolded state goes towards the folded state, the secondary structures form very, very quickly. And because they form very, very quickly, the slow step is usually the secondary structures coming together to form tertiary structures. And since there is a lot of space to explore, because there are many, many parts of the protein which may become helices, many different beta sheets can come together in different combinations.

I know this is difficult for you to visualize, especially since you are having problems with dealing with the concept of folding unfolding. The generic idea is that the unfolded state which is shown as a positive delta G, which is unfavorable, negative is sort of spontaneous, it forms certain platforms or certain restricted secondary structure elements, which form a sort of a funnel, which is why a funnel is the picture usually shown, and this funnel drives the protein to its, to a single unique confirmation, which is why when you start reading literature, you will see the representation of fast folding as a linear funnel as shown over here.

And I think I will come back to this in the next class, because this is not going to be easy for you to decipher. But remember the process we are talking about unfolded state, partially folded states and finally, the unique folded state, which usually is as a, which has a stability of around 5 to 10

kilo cals per mole. So, this is what, which is marginally stable and can be disrupted by even breaking a single disulfide bond.

(Refer Slide Time: 18:41)



Another concept which is important is this whole idea of water. Now, I told you this equation which relates the ratio of folded to unfolded states to the free energy. The free energy for a spontaneous folding is negative obviously, 5 to 15 kilo cals per mole. We also have this equation which relates the free energy to two parameters, enthalpy and entropy.

Now, usually, when you imagine hydrophobic entity and remember many of the amino acids are hydrophobic, tryptophan, tyrosine, phenylalanine, even alanine is not very hydrophilic, all of them tend to fold in such a way that the hydrophobic amino acids are found in the core of the protein, in the center of the protein rather than outside the protein. So, you will have the core of the protein being hydrophobic and you will have the outside of the protein being mostly hydrophilic.

Now, think about what happens when you have individual amino acids. Imagine this, imagine that this is a phenylalanine and imagine that this is the tryptophan. Now, individual amino acids will have a large amount of water molecules which are following, forming shells around them, because these water molecules cannot really, they are sort of isolating the hydrophobic entity from the hydrophilic solution which is around them. And imagine that there are about 30 water molecules for each amino acid.

When these two amino acids come together in the folded state of a protein, the shells merge and you have fewer water molecules which are now surrounding the protein. And this is pretty much what is happening. Usually you have, let us say around an average protein, you have a large number of water molecules, which are forming shells around the entire structure of the protein. And when you fold the protein, the shell of water molecules is actually in terms of numbers much, much little.

So, in some ways, in terms of the disorder in the water molecules, which is pretty much I am focusing on right now, the water molecules less water molecules are ordered, more water molecules are free to go into solution. So, overall, entropy is increasing, which becomes makes it a positive term. And I will come back to this. And because this is a positive term, it adds to the negative term, and which improves the stability of the protein. I realize these are difficult concepts, but we will go over them again tomorrow.

(Refer Slide Time: 21:15)

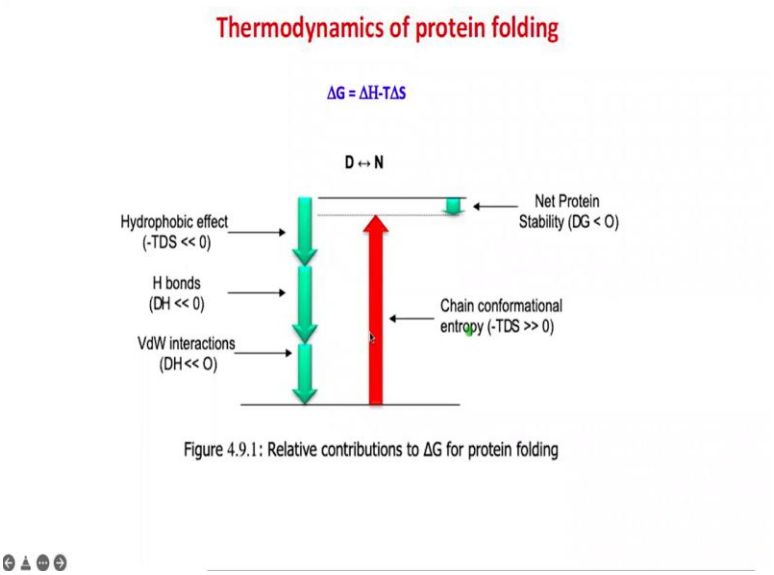


Figure 4.9.1: Relative contributions to ΔG for protein folding

So, to end this lecture, the net protein stability is actually very, very weak. It is about 5 to 10 kilo cals per mole. It is a negative value for a spontaneous folding of a protein. The forces which are involved in this net negative value include the hydrophobic effect, hydrogen bonds, and Van der Waal interactions, which are all favorable. And there is the chain conformational entropy, which is unfavorable, because the disordered protein chain in multiple confirmations has to fold so there is a decrease in entropy into a single folded state. And this opposes this entire process.

However, between these two opposing forces, you end up with a delta G which is negative, which causes the spontaneous folding of a protein into its folded state. So, I am going to stop over here.