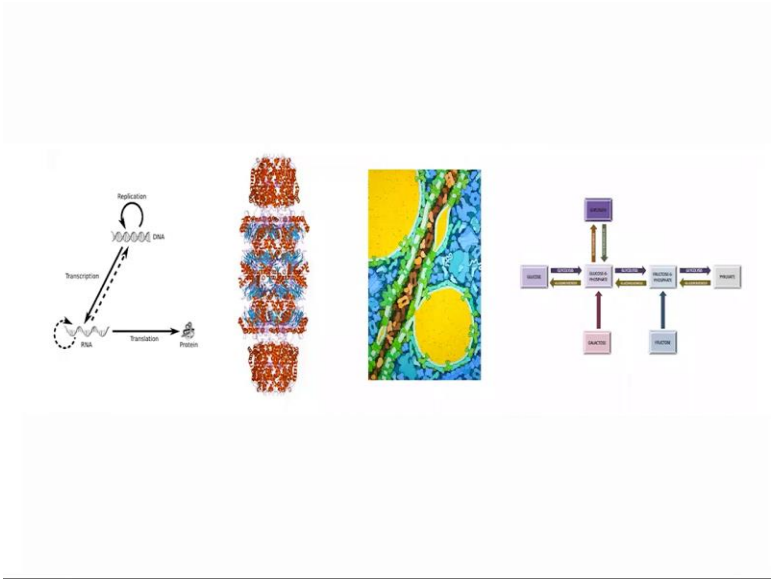
Introduction To Cell Biology

Professor Girish Ratnaparkhi and Nagaraj Balasubramanian
Department of Biology
Indian Institute of Science Education and Research Pune

Protein Structure, Folding And Function: Part 6

(Refer Slide Time: 00:15)



So, what we have been doing for some time, in a very broad overview with a lot of emphasis on concepts is to look at the central dogma, which is what we started with. And while doing the central dogma, I have introduced you to proteins, and DNA and RNA as macromolecules as major Biomolecules.

And you have also seen some of the machinery in the sell and how information is transferred from the storage molecule to the, two execute molecules. And that has been a big chunk of what I have been discussing. In the last few lectures I have concentrated a little bit more on proteins, though I want give it too much time will also touch upon lipids and also touch upon carbohydrates.

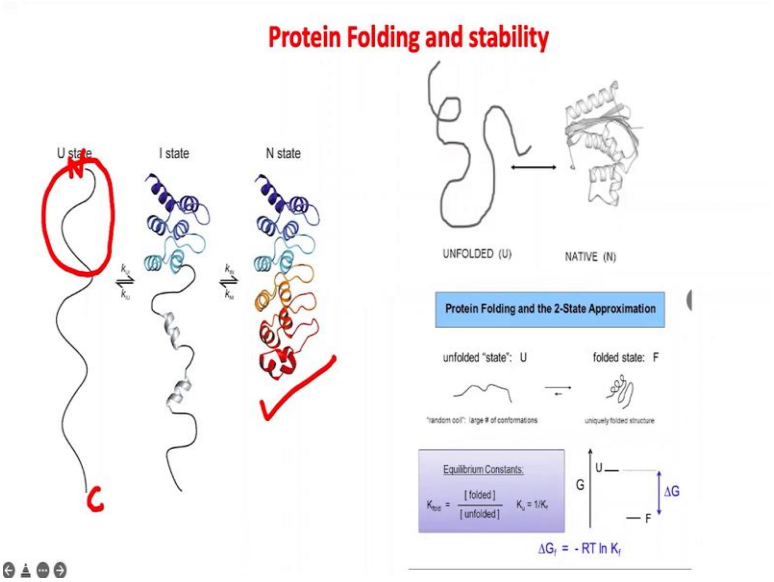
So, the picture over here is the broad area, which I will be covering central dogma, this is a, again, a ribbon representation of a protein, very large complex inside the cell called the proteasome and there are 100 and 1000s of interesting protein, proteins. This is interesting, because it is a recycling system.

It basically takes up proteins which are old and which need to be recycled, takes it inside, there is a pore inside, which is not obvious. Think of it as a thermos flask with a hole at both ends, the protein goes in through the hole in the top gets chopped up into little pieces, and gets chucked out from the, from the other side the pore on the other side.

Here are lipid fat droplets inside a cell. And one of the things which I hope I have conveyed to you through the various animations is that the environment inside the cell is very crowded with these macro and micro molecules. It is not like a beaker of water with a little bit of sugar in it.

It is extremely crowded environment. And what you see here in yellow is fat. Basically lipids, which we are going to talk about today and in blue are mostly proteins. And of course, there are water molecules pretty much everywhere. And on the right is the idea that carbohydrates are energy source. And this talks about the basic conversions between the major carbohydrates like glucose, fructose, galactose, inside the cell, which again is part of metabolism.

(Refer Slide Time: 02:27)



So, the linear sequence of amino acids, when it comes out of the ribosome is not in a folded state, it is in what we call as a unfolded state. It has a N-terminus and it has a C-terminus, just like nucleic acids have a 5 prime 3 prime, there is polarity in the polypeptide chain, with N and C terminus. So, let us assume synthesis starts from the N terminus side.

And let us say this much part of the protein has been synthesized while the rest is being synthesized this part as it comes out starts to fold. And this is happening inside the cell in vitro, if you take an unfolded state and put it in solution, it will start to fold, but it will start to fold as a single large entity.

Now, each region of this polypeptide chain will try and form secondary structure and the two major secondary structures as you are not aware of beta sheets, and alpha helices. And this particular example is alpha protein, which is just alpha helices and nothing else there is no beta sheets.

So, this protein starts to show local folding. And they will be different steps in this process of this protein folding locally, then the local folds that is the alpha helices coming together to form the folded state. And this is the so called folded state, which is the functional state of this particular protein.

Now, as the molecule transits from the unfolded state to the folded state, and this is a very fast process happens in microseconds, let us say there is a huge rate, it is from nanoseconds to milliseconds, depending on the protein, the larger the protein, the slower it is to fold. But broadly, it is a very fast process.

And it is exploring a lot of possibilities as it keeps on folding. And these possibilities include maybe transiently during the intermediate states, some of the chains trying to form a transient, let us say beta sheet, but finally, the protein will not show a beta sheet. Also in the path to forming the native or the unique native state, there will be many, many different ways by which it is like a Rubik's Cube. There will be many permutations and combinations by which you will attain the final state and you have to explore all of these states as you go from the U state to the N state.

Now, when biophysicist study protein folding, they simplify things and say much of the time there are only two states present, because once a protein folds, it is mostly in the folded state. And spontaneity of folding is driven by a negative delta G, which is basically in the range of 5 to 15 Kilocals per mole. And when we did a calculation.

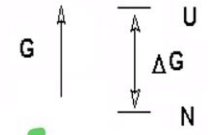
(Refer Slide Time: 05:18)

Stability of the folded state

- Measuring protein stability is measuring the energy difference between the U (unfolded) and F (folded) states.
- The average stability of a monomeric small protein is about 5 - 10 kcal/mol, which is very small!

$$\Delta G = G_N - G_U = -RT\ln K$$

$$K = e^{-\Delta G/RT} = e^{-10 \times 1000/(2 \times 298)} = 2 \times 10^7$$



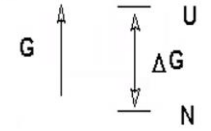
- i.e. in aqueous solution, at room temperature, the ratio of folded : unfolded protein is $2 \times 10^7 : 1$!

Stability of the folded state

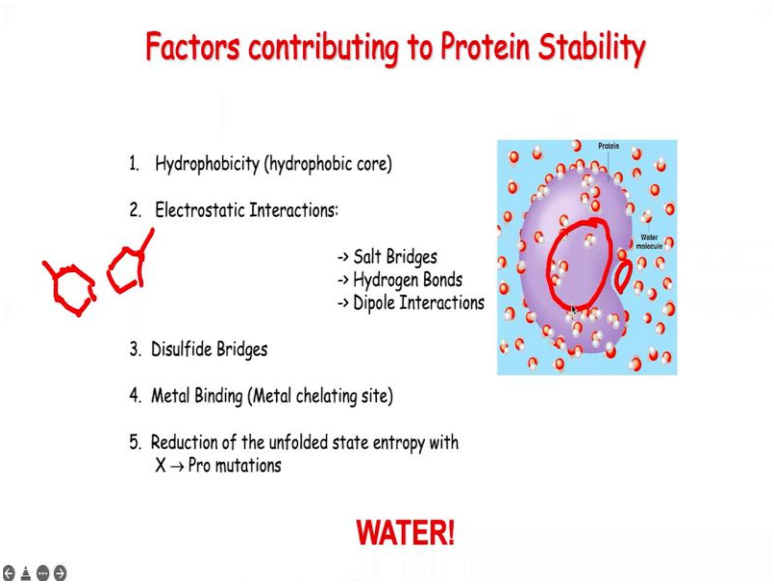
- Measuring protein stability is measuring the energy difference between the U (unfolded) and F (folded) states.
- The average stability of a monomeric small protein is about 5 - 10 kcal/mol, which is very small!

$$\Delta G = G_N - G_U = -RT\ln K$$

$$K = e^{-\Delta G/RT} = e^{-10 \times 1000/(2 \times 298)} = 2 \times 10^7$$



- i.e. in aqueous solution, at room temperature, the ratio of folded : unfolded protein is $2 \times 10^7 : 1$!



We basically found that if you for a protein, which has a delta G of minus 5 to 15 Kilocal per mole, the ratio when you calculate between the folded state and unfolded state is 1 into 10 raised to 7, basically indicating that at any point in solution with a delta G of minus 10, 10 raised to 7 molecules are folded and one molecule is unfolded.

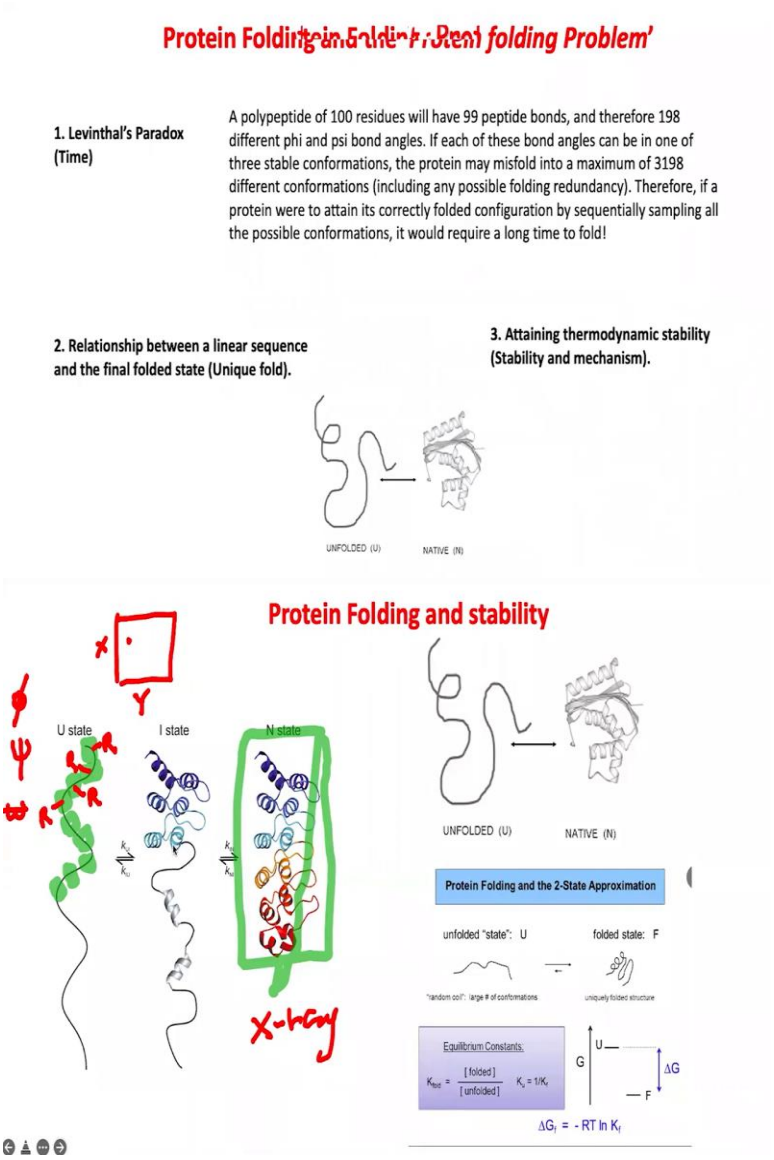
And there is a equilibrium between these molecules, so this was a very, fairly simple cause. Now, all this seems very well and good. But I also told you that the bonds which are stabilizing the protein, and remember proteins are stabilized in water. These bonds are of variety of types. These include salt bridges, these include hydrogen bonds within the protein, especially secondary structure, their wander walls interactions.

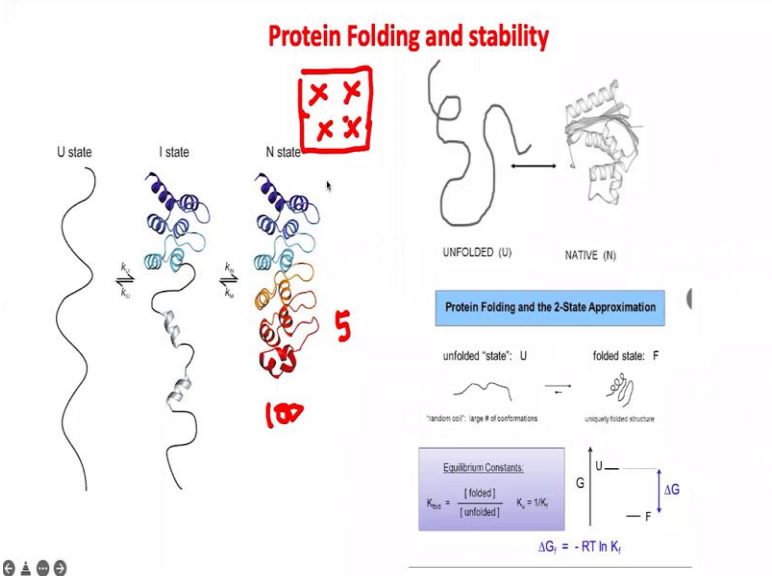
And there is also a packing in the core of the protein. And I will try and draw it for you. At the core of the protein, many of the hydrophobic groups packed together, away from water, and there is usually not much water inside the core of the protein, water is completely excluded. Most of the water is either on the surface or in crevices of the protein.

So, all we know, as of now, that experimentally, we know that the proteins are stable to the tune of 5 to 10 Kilocals per mole. We know that hydrogen bonds have a strength of around, let us say, 3 to 4 Kilocals per mole, we know breaking of a covalent bond is about 80 Kilocals per mole.

So, in theory, breaking of a single covalent bond can destabilize the folded state breaking of two hydrogen bonds can severely destabilize the folded state, in spite of the fact, any mutation can also destabilize the folded state. So, in spite of that, proteins are folded they are functional.

(Refer Slide Time: 07:13)





Now, I will wait for questions before I go ahead. Is there anything over here which you did not understand?

Student: Sir, could you move to the previous slides where the folded unfolded state of protein was there?

Professor: Yeah,

Student: this one?

Professor: Yeah

Student: I see that there is a equilibrium between U state and I state, and then the N state, what does the partially folded protein do in our body like? What is the purpose of it? Because, it would be unstable.

Professor: Right? So, partially folded states so, let me tell you the traditional viewpoint, the traditional viewpoint is the viewpoint from about 10-15 years ago, the traditional viewpoint is that proteins are made as unfolded states, and they need to function as folded states. And the only way to go to the native state is to go through partially folded states. So, partially folded states do not need to have any function, they are transiently present immediately after the protein comes off the ribosome. So, that is the classical view.

Student: And very short lived.

Professor: Very. So, the, remember the entire process of folding is in microseconds.

Student: Okay

Professor: So, intermediate state, since there will be maybe a 1000 intermediate states during the folding of a protein of different types, each intermediate state will probably be there for a pico second.

Student: And so one more thing from the previous lecture. Could I ask?

Professor: Yeah, please.

Student: Sir, while you were explaining about Ramachandran Map, you explained while drawing sequence of protein in which you took different amino acids, and then you told that I do understand that we plot the coordinates of the phi psi and omega and in the Ramachandran map, but how can we infer the configuration or where the density of that particular amino acid increases? If we are taking a protein which consists of different amino acid, I mean, it should be a same amino acid, and then we should try the permutations. And then we can plot.

Professor: So, let me try and give you a short explanation. I think there is a disconnect over here. So, this particular unfolded state over here, which you can see brown over here, you can see that. There is basically a sequence of amino acids fine. Now, all I am saying is that if each block over here, now I will just use some other color if each block over here represents one amino acid. You agree with that?

Student: One same amino acid?

Professor: No, it is not a same amino acid, different amino acids. Each block represents a amino acid and they are 100 of them in this unfolded state, all I am saying is, for each block, which contains a phi now that is better psi and omega, and we ignore the omega because it is always 180 degrees. For each block, we have a phi and we have a psi value. And in the Ramachandran map, we just plot for each block, x-axis is one y-axis is another, we just plot a dot so that is the Ramachandran map. And the next amino acid is a completely different amino acid. But the phi psi are main chain torsion angles, they are really nothing to do with the side chain.

Well, indirectly, they have something to do with the side chain, but the atom is used for plotting phi and psi all main chain atoms, what the R group is, is not used for the calculation at all.

Student: So, that gives the, tells about the protein structure, not in how many configuration that particular?

Professor: No, so the Ramachandran map is only a representation of the native state. Because that is the.

Student: Okay I got it.

Professor: That is the state, represented by x ray crystallography. During folding, the protein will follow basic rules of chemistry, but it may have a more, broader representation and Ramchandran space than the folded state.

Student: Yes sir, yes sir thank you.

Professor: Are there any questions?

Student: A question about Ramchandran map, should ask it now?

Professor: Yes

Student: The map for individual amino acids right?

Professor: Yes

Student: For glycine, and proline. So, over there, the glycine one, for example, represents all of the phi psi angles glycine can take with some being more favorable than others.

Professor: Right.

Student: But then and you also showed as a more generic map which had all of the amino acids of all proteins plotted. So, how is it that the glycine map covers more area than the generic map? Because, everything that is plotted on the glycine map, should also be on the map with all other amino acids.

Professor: So, let us see if I think here again, there is a different way of looking at things. This particular let me take this unfolded state again, or the folded state over here, let us talk about the folded state, the folded state over here, this molecule, let us assume that it is made up of 100 amino acids. There will be not more than let us say 5 glycines in this linear sequence.

So, if you draw a Ramchandra map for this particular protein, you will have 99 dots and only 5 of them will be glycine, and those 5 glycines can be here can be here can be here, or can be here in the Ramchandra map. So, if you look at dots one, only one will be here, the others will be in Ramachandran, normal Ramachandran space. Is that clear?

Student: Yes sir but the map you showed, I think all of the amino acids did not have any dots at all in the fourth quadrant.

Professor: So, that is see glycine is free to explore more areas of conformational space than any other amino acid. Just because it is free to explore does not mean it always explores, because it is also part of a main chain surrounded by tryptophans, tyrosines bulky amino acids.

Student: Yeah, understood

Professor: Can and will is are two different things over here.

Student: Okay, sir I understood.

Student: Sir?

Professor: Go ahead.

Student: Sir, I have a doubt, in what is this 5 to 10 kilo calorie energy, it is delta G of protein, or it is the energy of peptide bond.

Professor: So, the minus the 5 to 10 Kilocals per mole is a representation of the stability of the folded state over the unfolded state. And it is basically the overall stability of the folded state with respect to the unfolded state. And you can relate the delta G through two main equations in biochemistry, one is a relationship between delta G with the equilibrium between the two states and one is the relationship between delta G with enthalpy and entropy. Not clear enough even now.

Student: Ya sir, okay.

Student: Sir, whatever map we drew was for one particular protein or like when you talk about glycine, and we do the Ramachandran map, it was a poly glycine.

Professor: Yes so the poly glycine map will show you dots in all four quadrants.

Student: That is fine, but we never drew a generic graph for all the amino acid of all the proteins that exist.

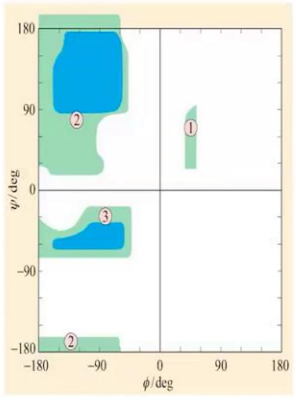
Professor: Yes, we did.

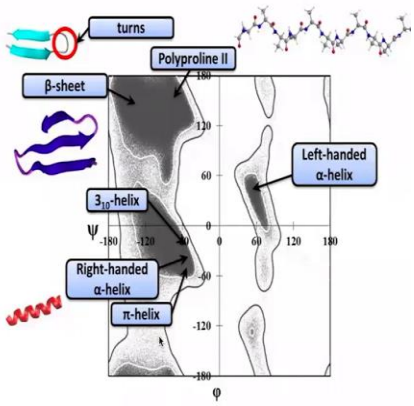
Student: For all the proteins.

Professor: Pretty much means all the proteins means whoever drew it at whatever point drew it for pretty much all the proteins.

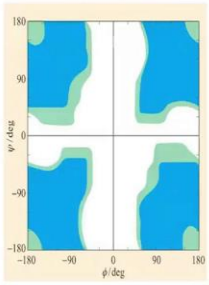
(Refer Slide Time: 15:13)

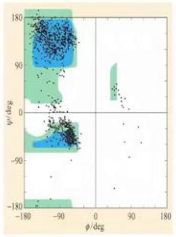
An appreciation of a 3D structure in 2D → The Ramachandran Map

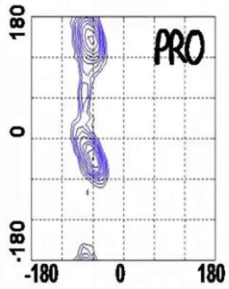


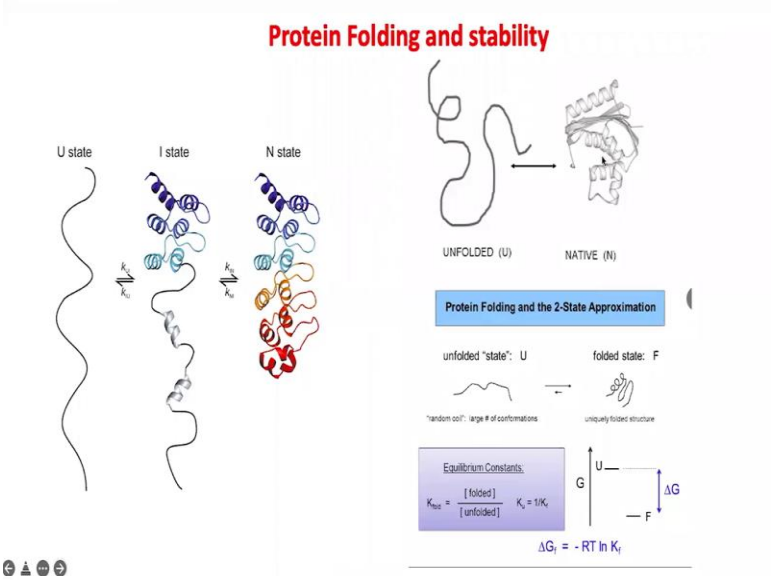


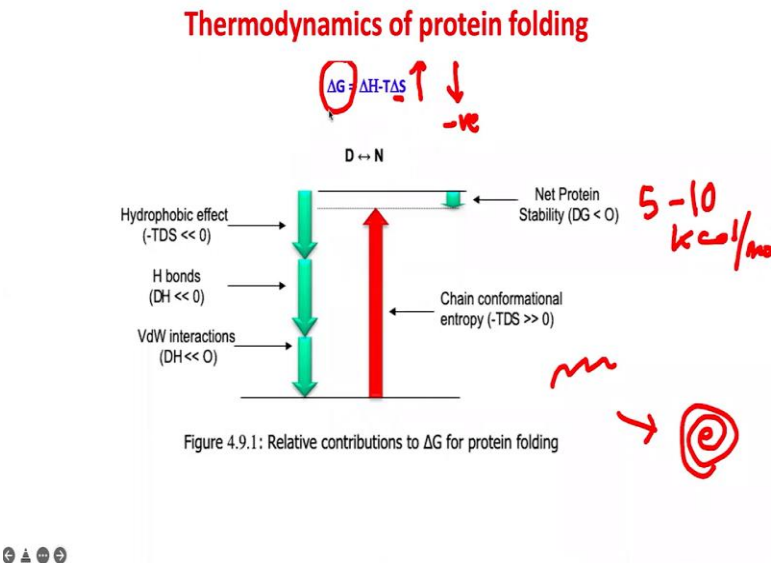
The main-chain conformation structure in 2D -> cylinder, and restricted Proline











Professor: For example, if you look at this one over here, if you look at the number of dots on the, picture, there are probably 100,000 dots over here many, many dots. Look at all the little little dots behind these, these are not shading these are dots. So, this is a representation of many, many, many proteins.

This is probably a representation of about 10 or 12 proteins, or a very one large single protein. Is that clear? I need you to conceptually understand that the ribosome makes a polypeptide, it goes through intermediate forms, it forms a native state the native state is the functional state. And we have ways to measure how stable a protein is. And surprisingly, the protein is stable, moderately, because 5 to 10 Kilocals per mole, which is the stability of this protein over here.

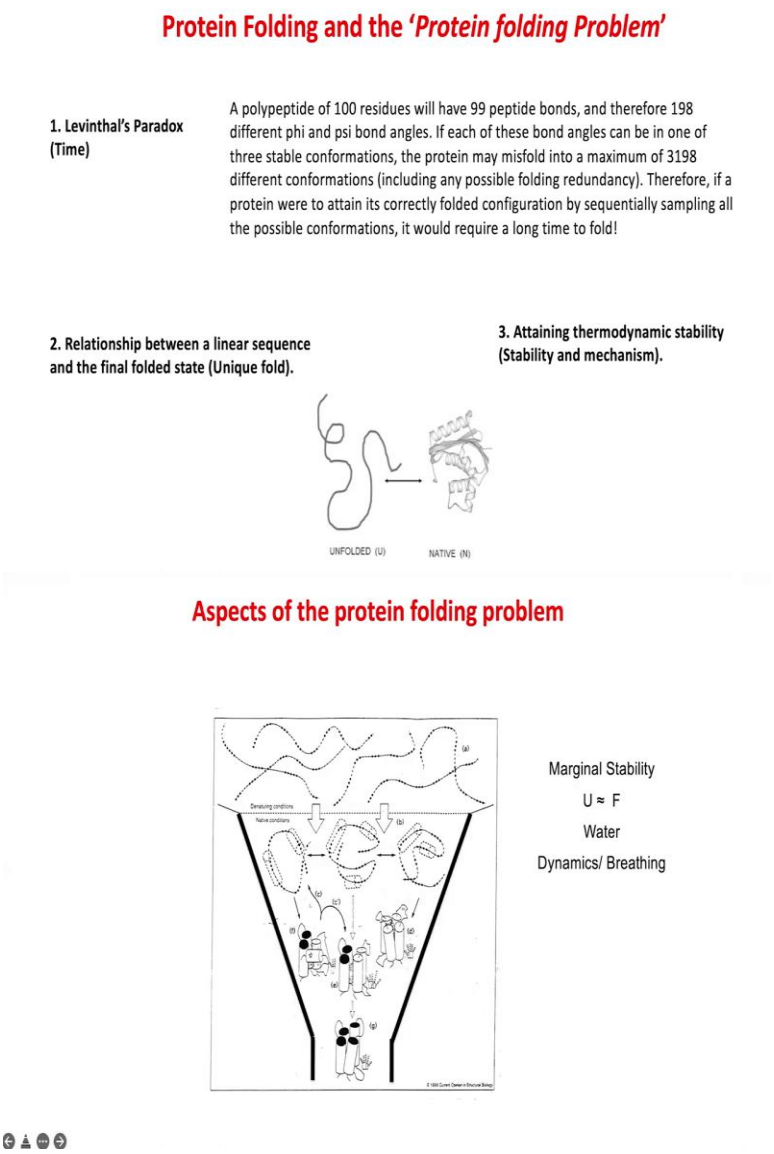
If I break one hydrogen bond, the stability goes from 10 Kilocals per mole to 8 kilocals per mole, I break two hydrogen bonds, it goes from 8 Kilocals per mole to 6 Kilocals per mole. So, I make one mutation, it goes from a stability of 10 Kilocals per mole to about let us say 3 Kilocals per mole.

So, this is stable and but weakly stable, and the point I was trying to make is that when you look at the different forces, so let us not go into the thermodynamics of it for this course. So, this slide is basically not part of the course, it is a balance of different forces, which are enthalpy and entropy, which we can discuss tomorrow morning.

And when you add up all the forces, which are involved in stabilizing and destabilizing a protein, the net stability is as is written over here is very small. So, this is the number which is used, whereas the forces which are against and for are very, very large forces, fine with, with the fact that an unfolded state with a large entropy folds into a folded state with a, very low entropy there is a increase in order over here opposes, the spontaneous free energy folding.

Simply, because the entropic term over here, it has to if it goes up, there is more disorder, if it goes down there is less disorder. And the folding of a chain basically lowers entropy, and if it lowers entropy, it is a negative term, it is a negative term, the negative or negative is positive. And because this is this term becomes positive, delta G becomes positive. And for a spontaneous folding, we want the delta G to be negative. So, all this is really not part of the course, but I am just trying to give you a quick explanation.

(Refer Slide Time: 18:02)



So, I just want to leave you with this idea of the variety of proteins, large proteins, small proteins, many of them doing functional roles, and the broad ideas about protein stability, the fact that there are multiple stabilizing forces involved, everything is in water. And there is something called as a protein folding problem, which has existed for over 60-80 years.

And there are different aspects of the protein folding problem. One is the time aspect. How is a protein folding so quickly, because one assumes it has to explore a lot of conformational space to find its final native state, it is like a Rubik's Cube, again, you start with a Rubik's cube which has been shuffled and if you are not, if you have not done Rubik's Cube before, you can spend hours

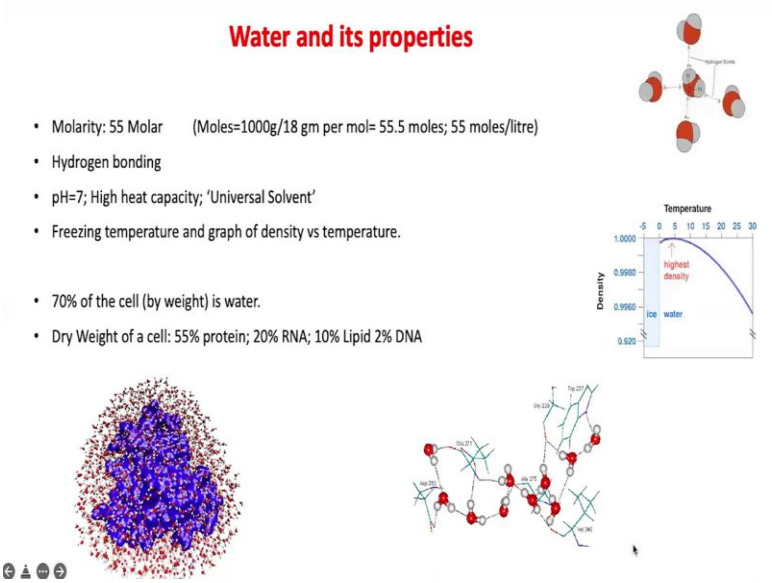
and hours trying different combinations trying to solve the Rubik's Cube and if each moment of your hands takes you let us say a 10th of a second and it takes you 10 seconds to make 10 moves and you keep on doing these moves and you never find, find the solution.

That is what a protein that is what can happen to a protein. Second thing I have been telling you is there is a relationship between the sequences of amino acids, these are not random. Each folded state is in its folded state because of a unique sequence which is there. So, hemoglobin in all species, starting from hemoglobin like entities in let us say, invertebrates and myoglobin.

They all have very conserved sequence of amino acids and this conservation is very important, because if you lose this conservation of the linear sequence of amino acids, you will not get this the folded state. And the third part of the protein folding problem is this whole idea that proteins are weekly, weekly stable, any small push on the wrong side can force them into going to up to the unfolded state, and going to the unfolded state is a disaster.

Because the, unfolded state is not a, functional state. For hemoglobin to work, it has to be in its folded state in the unfolded state, it will not be able to bind oxygen. So, folding is a very, very critical physical chemistry phenomena, biological phenomena, physics phenomena in going on inside the cell, and every second 100s and 1000s of proteins in your body are being made by the ribosome and are folding. And if they do not fold properly, basically, life does not exist.

(Refer Slide Time: 20:32)



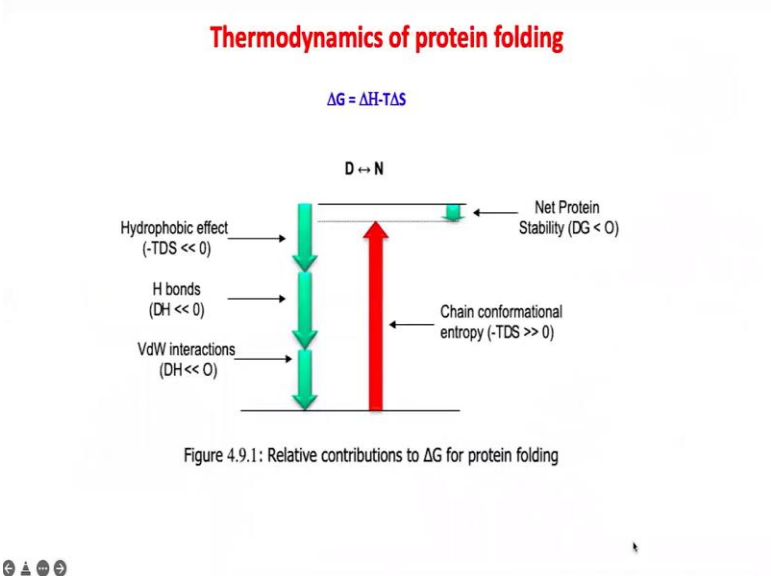


Figure 4.9.1: Relative contributions to ΔG for protein folding

Student: Sir, sorry to worry you again about the Ramachandran map, but one thing is not clear now also. Sir, for example, I am taking a lysine amino acid and I want to know in how many configurations that particular amino acid can exist in terms of how many angles of phi psi and omega is same 180 degree it can undergo naturally in our body for that, for plotting that particular thing I need to know one particular protein in which all that possible configuration should exist, am I right?

Professor: Not really, you see.

Student: Sir, then how did we got the idea about like, we maybe we have all the maps of amino acid, all the amino acid in which what all configuration they can exist. So, how can we get the idea if we are taking up protein in which there are different amino acids then how can we plot and be very sure about that, okay this particular amino acid cannot go into that phi psi angle and only this one like we plotted.

Professor: Nobody saying an amino acid, well people are saying an amino acid cannot go into a certain phi psi configuration, the generic statement is except glycine and proline, all other amino acids can only go into 30 percent of the conformational space available. And the reason they can only go into the 30 percent of the conformational space available is because they are in a linear sequence one after the other, if you rotate around the torsion angle, the R groups bang against each other.

And because they bang against each other some areas of torsional space which are allowed to glycine are not allowed to any amino acid alanine onwards, alanine to tryptophan, no amino acid can go into certain areas of conformation space.

Student: So, scientists have plotted all the proteins and then they have figured out now I got it thank you.

Professor: Ramachandran did this without having structures he used it he used models. So, he predicted this without even having the data we have today. But every year when another 1000 structures are solved, we know Ramachandran was perfectly correct. Every structure validates his initial modeling.

Student: So, here D means I mean that D to N part, D means destabilized?

Professor: Yes, I am sorry different nomenclature is used. D is the denatured state or the unfolded state.

Student: Denatured, okay sir.

Professor: N is the native state or the folded state.