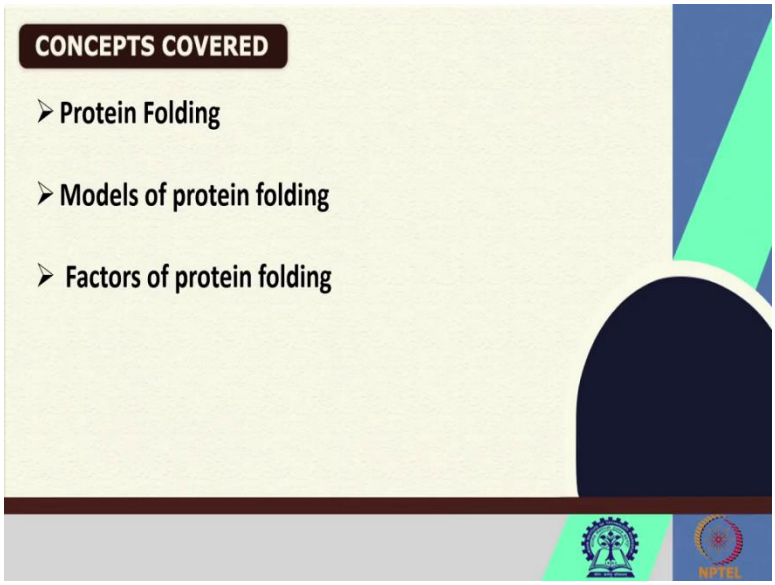
Fundamentals of Protein Chemistry
Prof. Swagata Dasgupta
Department of Chemistry
Indian Institute of Technology, Kharagpur

Module - 03
Protein Stability
Lecture - 11
Protein Folding and Structure

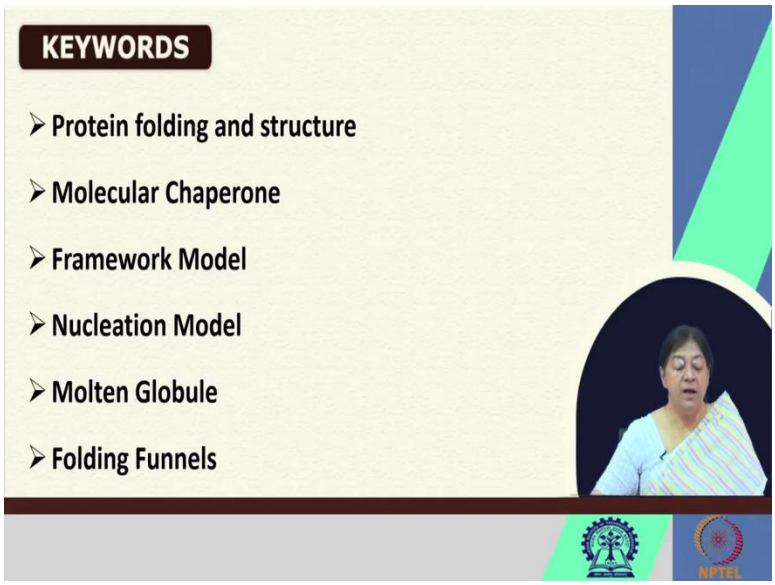
We begin module 3 of the course, fundamentals of protein chemistry. And in module 3 that is based on protein stability, we will be revisiting protein folding and structure. And then go on to understand the thermodynamics of protein folding and then see which particular actions could disrupt this particular folding or the interactions that are involved in the folding.

(Refer Slide Time: 00:45)



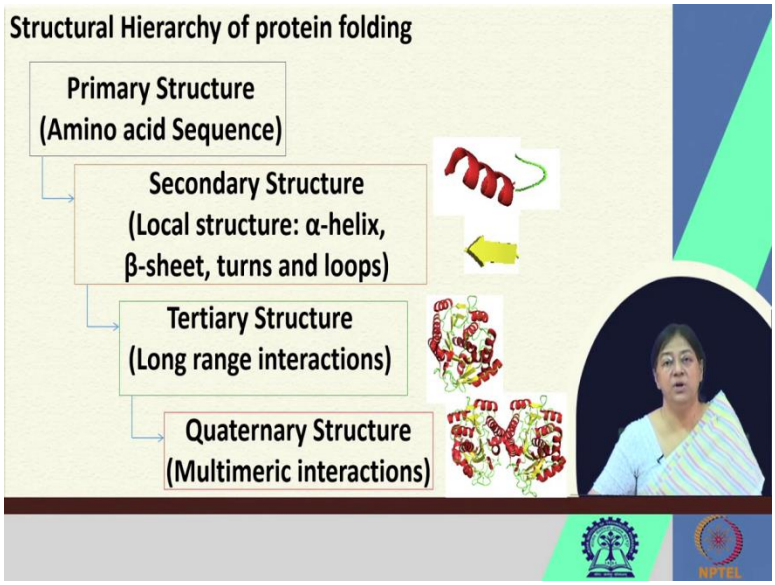
So the concepts that we are going to cover are going to be protein folding, the models of protein folding and the factors that affect protein folding.

(Refer Slide Time: 00:55)



There are specific associations in terms of understanding how the protein is going to fold.

(Refer Slide Time: 01:09)

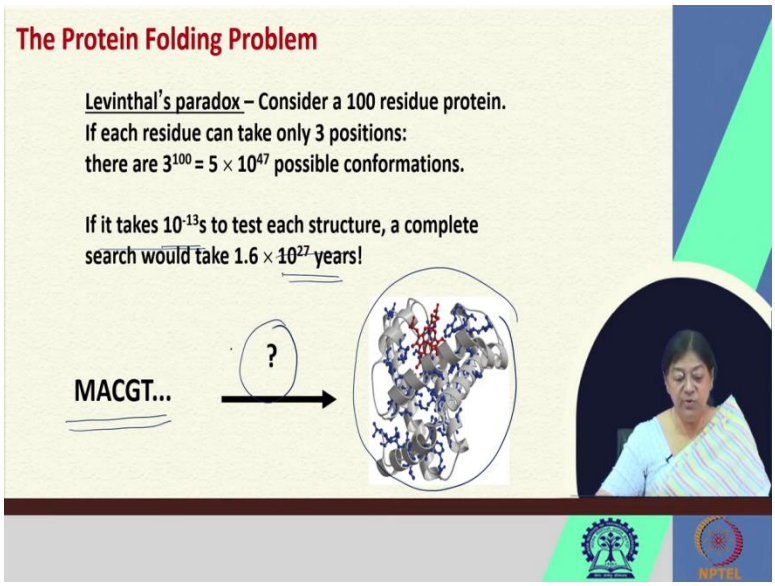


If we look at the structural hierarchy of protein folding, which we had discussed in the previous module, we see that there is the primary structure that comprises the amino acid sequence followed by the secondary structural elements, where we have the α-helix the β-sheet. And also the turns and the loops that link the secondary structural elements together. We will be looking at a bit more detail in terms of these structural units or what are called super secondary structures or motifs and domains.

In the tertiary structure, we have long range interactions which we looked at where we have a folded form, where different parts of the protein molecule come together in non-covalent interactions. Then we have the quaternary structure, giving us multimeric interactions between

several subunits that could range to a small even dimeric protein to very large 24 mer as was shown in one of the lectures.

(Refer Slide Time: 02:13)



There is a protein folding problem and it is called and Levinthal's paradox. In the sense, that if we consider a 100 residue protein, we already have a knowledge of the possible rotations about the single bonds that we have.

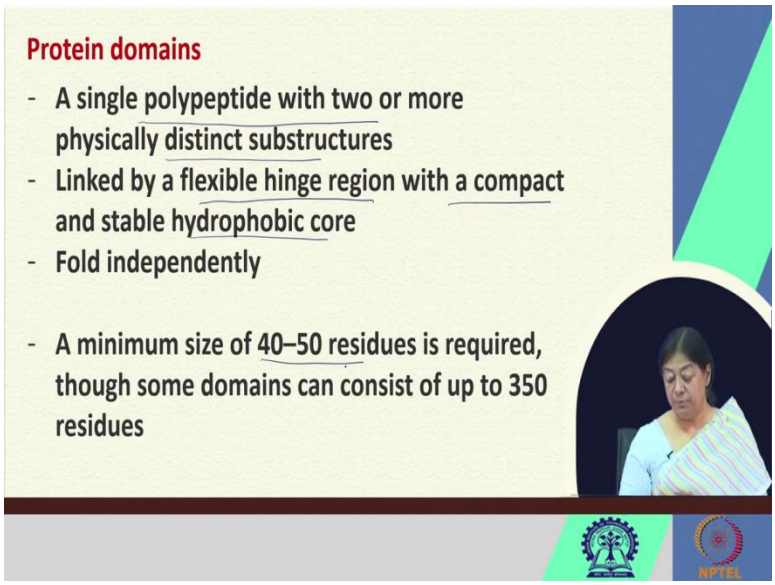
Given that we have only three conformations per residue, for this 100 residue protein in this protein folding problem. If we say that each residue can take only 3 positions, though we know it can take very many positions depending upon the rotations. But then again we have the geometric constraints on the rotations, but nevertheless we are considering only 3 specific positions.

This means there are 3 to the 100 possible conformations that this primary sequence of a 100 residues can have. Considering only 3 possible positions per residue, means that there are 10 to the power 47 possible conformations. Now when we look at our specific native folded structure, what we actually see is, we see that we have the primary sequence and we have the folded structure.

Now, the question is that if it takes 10 to the power minus 13 seconds to test each of these possible folded structures, that means even less than a picosecond to test each structure, then a complete search would take 10 to the power 27 years. And how long does it take for a protein to fold? Less than a second some milliseconds sometimes.

This is what the intriguing protein folding problem is.

(Refer Slide Time: 04:25)

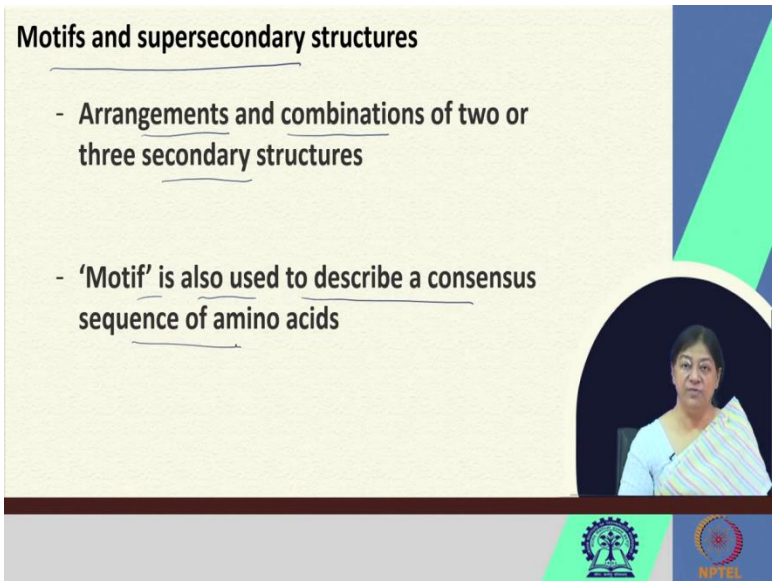


Now how does a protein fold? For this we need to understand the protein domains. We have to think of how the protein, the polypeptide chain approaches this problem, in a sense that, when we look at a single polypeptide chain with two or more physically distinct substructures.

We looked at the primary sequence which was the sequence of amino acids, we have the α-helix the β-sheet. Now we are trying to look at sub structures. Would substructures initially form before the native structure is formed. What happens in these cases? We have linkers as we saw and we have a compact and stable hydrophobic core for globular proteins, and they can fold independently.

These are the protein domains that we are talking about. And there are specific sizes of these domains that could range from a 40-50 residue sequence, sequence part of the polypeptide chain, or even up to 350 or more.

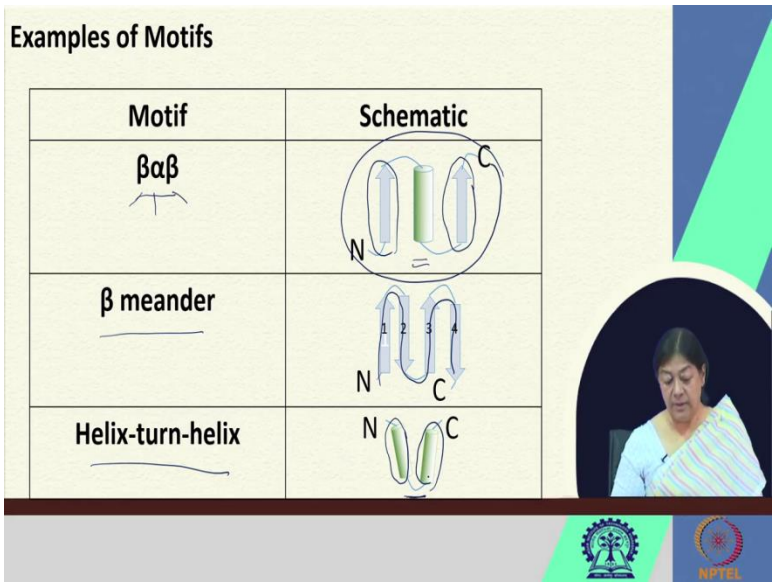
(Refer Slide Time: 05:29)



When we look at these motifs [refer to slide] and these super secondary structures, they are going to be arrangements and combinations of two or three secondary structures put together, where we would have a combination of the secondary structure elements in forming a unit by itself.

So this motif is used to describe a consensus sequence of amino acids.

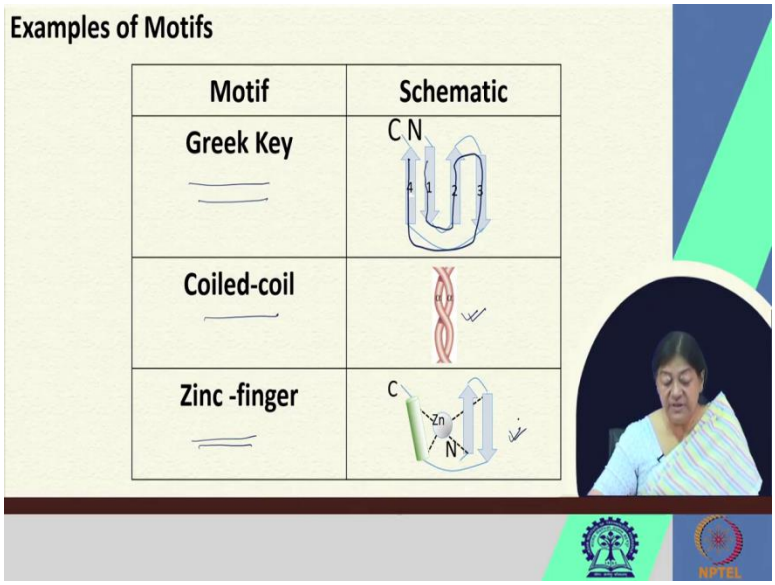
(Refer Slide Time: 05:57)



For example if we look at a motif called a βαβ, we have a beta strand here, we have a strand here and in between we have an α-helix. This is a motif. So, we see that together if we look at a schematic diagram of this what do we have? We have the β strands and we have an α unit sandwiched between the strands.

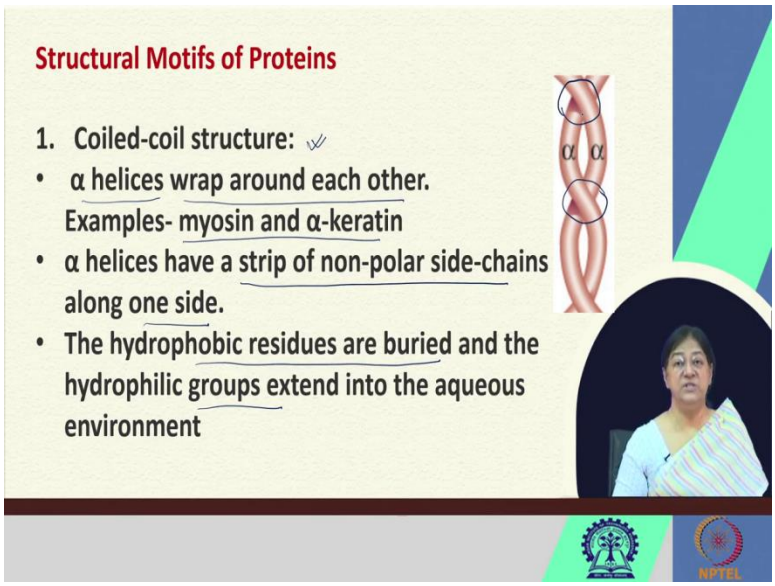
We can have what is called a β meander, where we have the β strands go in this direction. We can have a helix-turn-helix, where in this case we have a helix, we have a turn that is connecting the two helices together.

(Refer Slide Time: 06:51)



We can have what is called a Greek key motif. In this case, we have strands to a β-sheet which form in this fashion 1, 2, 3, 4. This is a Greek key motif. We have what is a coiled-coil. This is where we have α-helices that form this coiled-coil structure, another well-known motif is the zinc finger motif, which we will visit when we study metalloproteins.

(Refer Slide Time: 07:27)

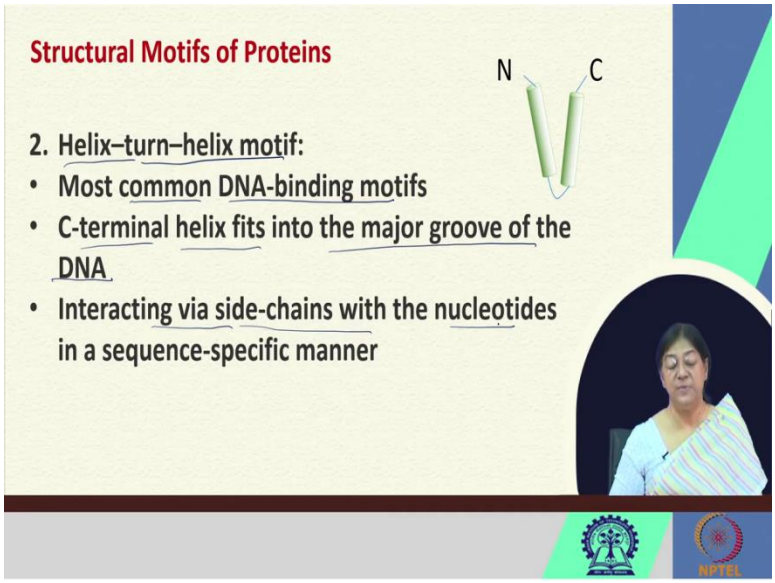


These are different examples of motifs, we will look at specific types now.

In the coiled-coil structure that we saw, the α-helices wrap around each other. The examples are myosin and α-keratin, and when we look at α-keratin and myosin, we remember they were present in the muscles and in hair. So these are the fibrous type proteins. That means, they have a strip of non-polar side chains along one chain and the hydrophobic residues are buried and the hydrophilic groups extend.

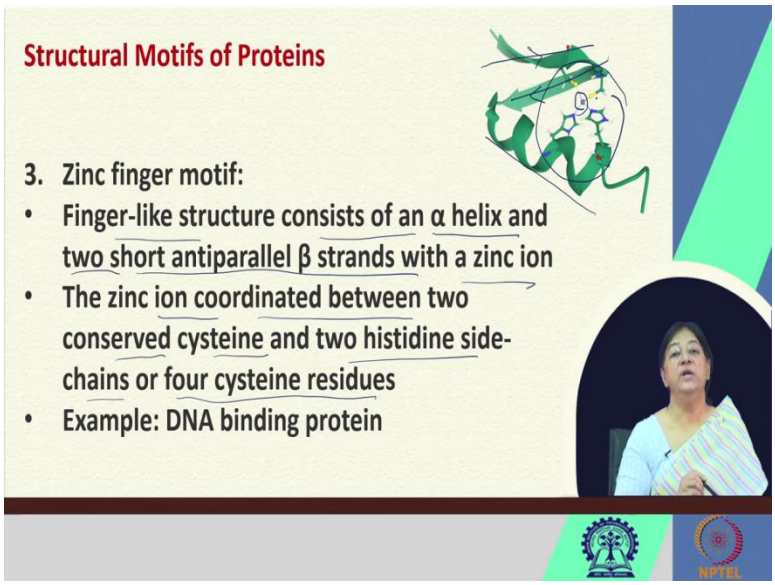
So the overlap that we see here [refer to slide], where we have the chains interact with each other, have hydrophobic interactions that allow this specific interaction to occur between the two chains.

(Refer Slide Time: 08:15)



When we look at a helix-turn-helix motif, these are found most commonly in DNA binding motifs and the C-terminus helix fits into the major groove of the DNA structure. And the interactions are via side-chains with the nucleotides and in a sequence specific manner.

(Refer Slide Time: 08:39)



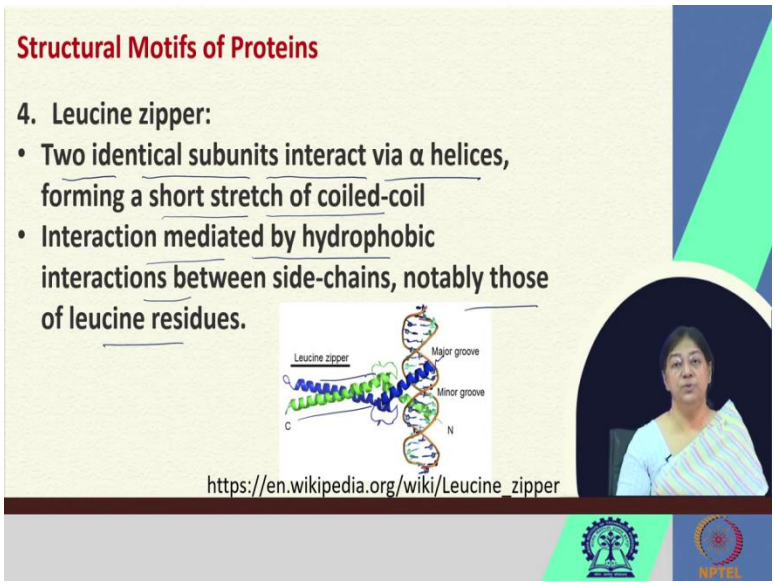
So there are specific ways in which these would interact with each other.

In a zinc finger motif, we have a finger-like structure that consists of an α-helix and two short antiparallel β-strands with a zinc ion. [Refer to slide] here is the helix and here we have the antiparallel β-strand.

The zinc ion is coordinated between two conserved cysteine and two histidine side chains or four cysteine residues. So this, as we can see is a very unique combination of residues that are required, as we can see this particular scaffold is required to hold the zinc ion in place. So this specific motif indicates that we have the α-helix; we have the β-strands containing these particular residues coming together in three-dimensional space, to form this scaffold that is going to hold the zinc in place.

This folding as we see is extremely important, because it is bringing the different parts of the polypeptide chain together in three-dimensional space, to form the specific pattern of residues that can hold the zinc ion in space.

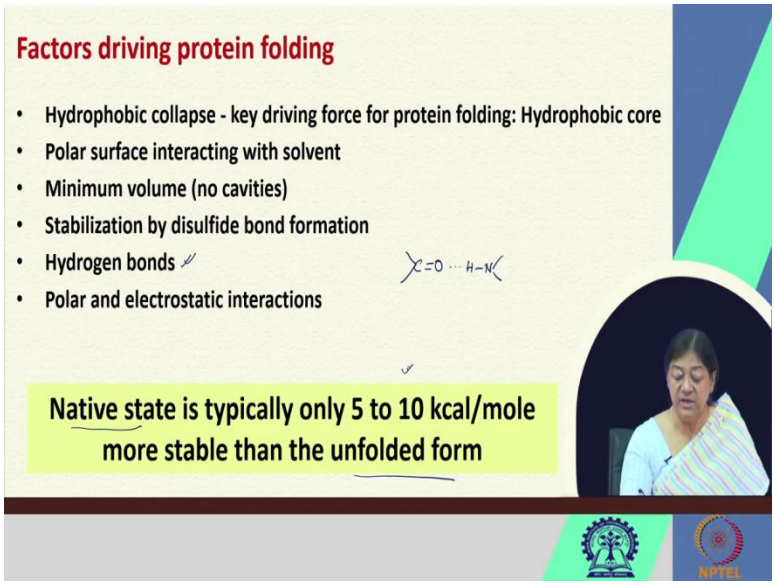
(Refer Slide Time: 10:13)



A leucine zipper is another such motif, where there are two identical subunits that interact via α-helices to form a short stretch of coiled-coil, similar to what we saw in α-keratin and myosin.

And here again there is an interaction mediated by hydrophobic interactions. But in these cases these are mostly leucine residues and is called the leucine zipper; because we have the leucine residue side by side interact with each other and form a hydrophobic sort of a zip. So, they interact in a manner where we have the hydrophobic interactions. This is extremely important in this leucine zipper and it is also another common structural motif that is observed in proteins.

(Refer Slide Time: 11:07)



Now what are the factors that actually drive this protein folding? Hydrophobic collapse - what does this mean? When we have a globular protein, we know that the hydrophobic amino acid residues would likely be occupying the central part of the protein.

We call this a hydrophobic collapse, where the hydrophobic residues will tend to very quickly come away from the surface of the globular protein, so that it interacts with other hydrophobic residues within the core of the protein.

The polar surface would preferably interact with the solvent and there would be a minimum volume (no cavities), so that we have a compact structure. And we have the stabilization which we saw by the disulfide bond formation; another covalent linkage that is going to bring different parts of the polypeptide chain together.

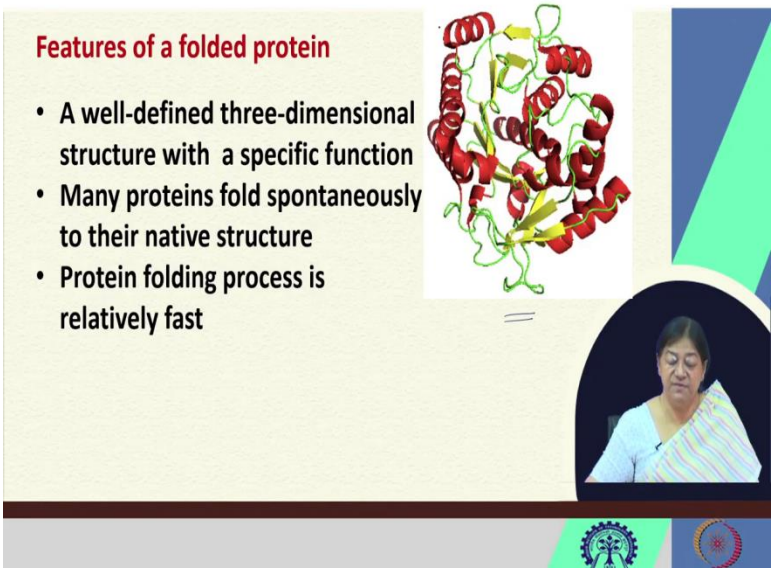
We know that the hydrogen bonds always form between the C = O and the NH, but in this case we can also definitely have hydrogen bonds between the side chain residues that have polar components in the sense they have oxygen or nitrogen in their side chain to also form hydrogen bonds that are going to stabilize the structure.

Polar and electrostatic interactions between ion pairs that can be associated with the acidic and the basic amino acid residues.

Now, if we look at the energetics associated with a native state and the unfolded form. We have an unfolded form of the protein, which is the primary sequence that finally folds into its three-dimensional native form, its three-dimensional functional form.

So, this stability is 5 to 10 kcal/mole.

(Refer Slide Time: 13:19)



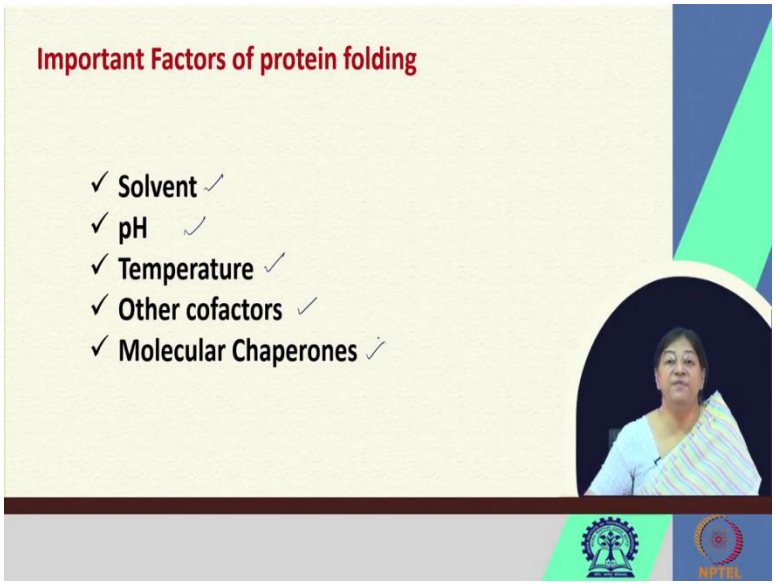
So what are the features of a folded protein. The features are that there is a well-defined three-dimensional structure with a specific function.

This [refer to slide] primary sequence that comprises this particular protein that is shown on the right here, is a long polypeptide chain with different amino acid residues linked by peptide

bonds, but it has folded into this particular structure. So, this well-defined three-dimensional structure has a specific function, which is why it has folded the way it resembles here.

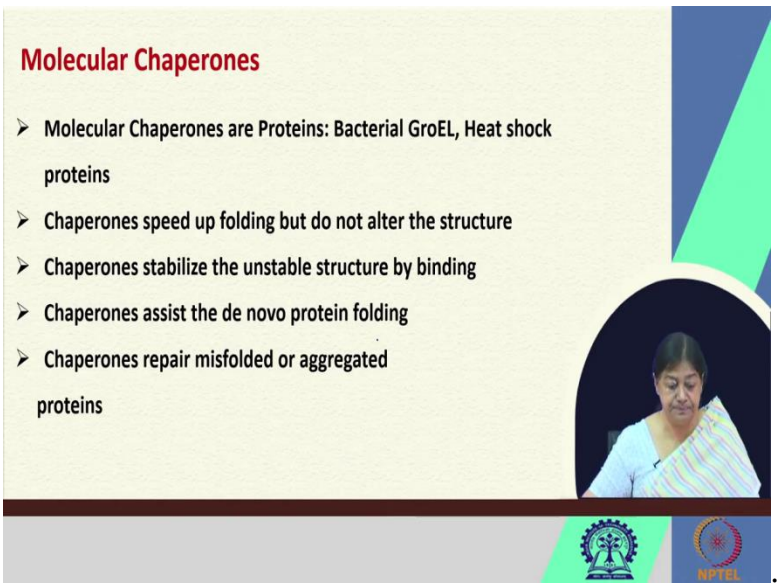
Now many proteins fold spontaneously to their native structure and the protein folding process is very fast.

(Refer Slide Time: 14:15)



What are the factors that drive protein folding? The solvent, the pH, temperature, other cofactors and molecular chaperones, are the factors that drive protein folding. And we will see how these would affect protein folding.

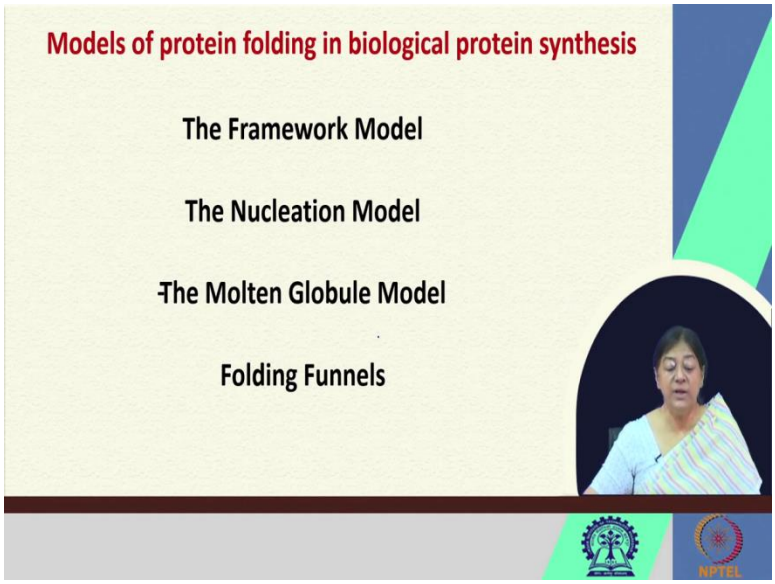
(Refer Slide Time: 14:39)



If we look at molecular chaperones. Molecular chaperones themselves are other proteins for example, the GroEL-GroES complex or other heat shock proteins. Now what these proteins do is, they assist the folding of other proteins, because as we know the unfolded polypeptide chain is this primary sequence. It has to fold into its three-dimensional structure.

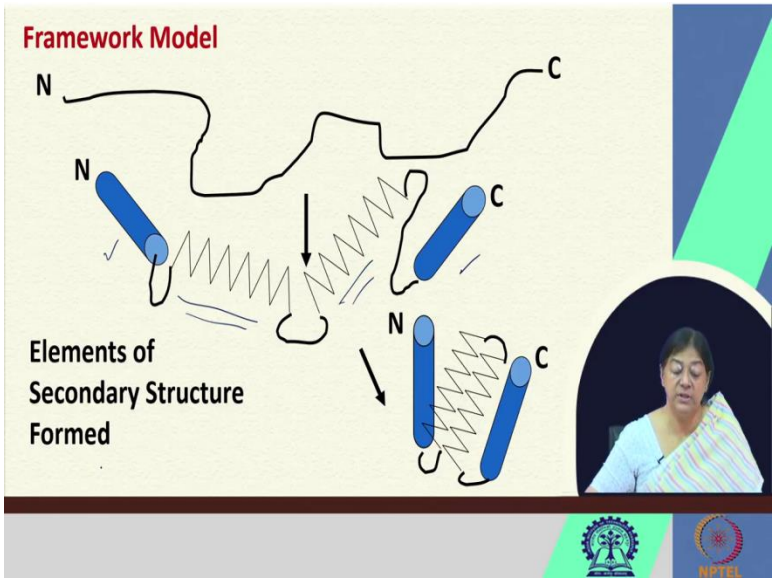
Chaperones help or they speed up the folding but they do not alter the structure of the protein. They stabilize the unstable structure by binding to it and they assist in the de novo protein folding. So, molecular chaperones are extremely important in the assistance of the folding, because they facilitate folding and they also can repair misfolded or aggregated proteins that actually are the cause for disease.

(Refer Slide Time: 15:37)



If we look at the models of protein folding that are known, we have the framework model, the nucleation model, the molten globule model and folding funnels. So we will visit these and find out what they mean.

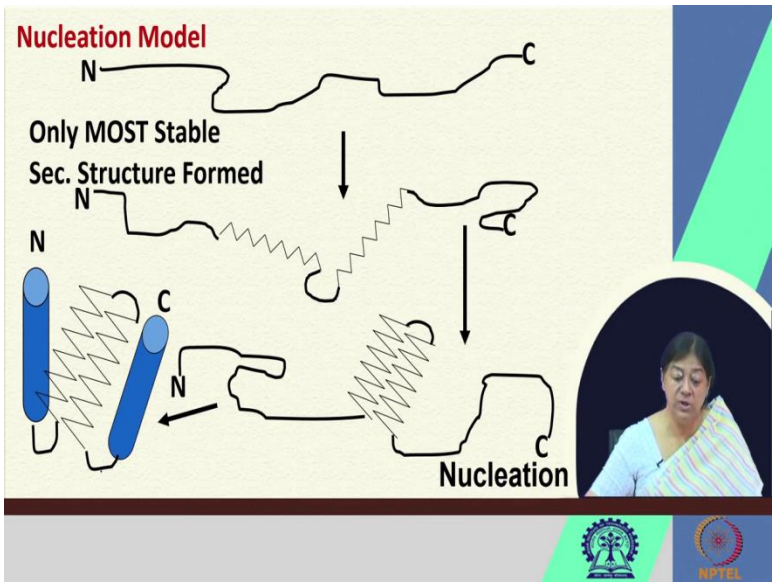
(Refer Slide Time: 15:55)



In the framework model, we have this [refer to slide] polypeptide chain. The framework model tells us that we would have the elements of secondary structure formed. So what we see is, we see a set of strands, we see a set of helices. These are helices and these are specific strands.

So these elements of secondary structure are formed first and then they form the folded structure. This is the framework model, where we have the elements of secondary structure formed in the beginning.

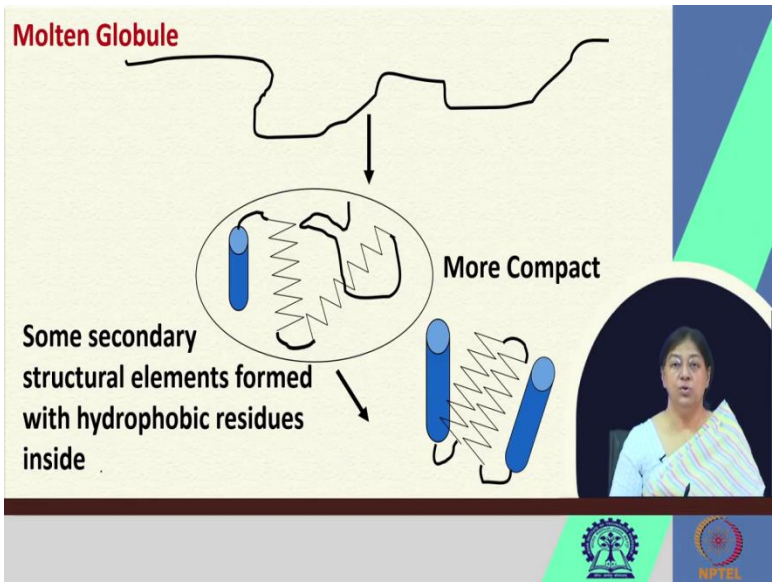
(Refer Slide Time: 16:31)



In the nucleation model, what happens is only the most stable secondary structure is first formed. Then this associates giving a nucleation to the protein folding process and this then finally will

form the folded protein. So we get the final folded structure either by a framework model or a nucleation model.

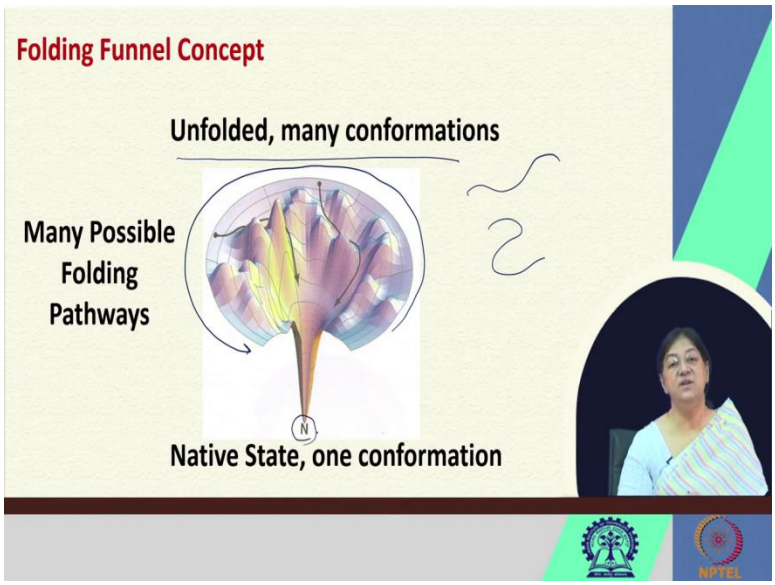
(Refer Slide Time: 16:57)



In the molten globule model, again we have the unfolded structure and we have a more compact structure. Just a folding of the polypeptide chain to form a unit, a molten global compact unit called the molten globule and some secondary structure elements are formed, some hydrophobic residues result in a collapse where the hydrophobic residues come together in the central part. And then this will form the overall folded protein.

Now each protein will have its own way of folding. In a sense that usually the molten globule way is where it is from. So, some secondary structural elements are formed, the hydrophobic residues tend to get into the central part of the protein-the hydrophobic core, and the rest we adjust to form the overall compact three-dimensional structure known as the protein.

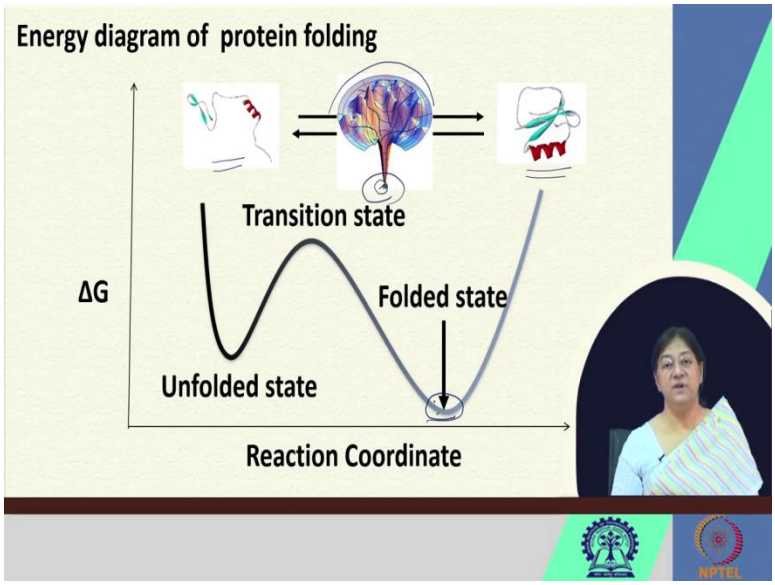
(Refer Slide Time: 17:55)



The folding funnel concept is what is in vogue. There are many conformations of the unfolded protein. We have these many conformations resulting in many possible foldings. So there is this folding funnel that says that there are many unfolded conformations present.

All of these are the same polypeptide chain, but folded in very different manners. What it does is, there is finally one native state, one conformation.

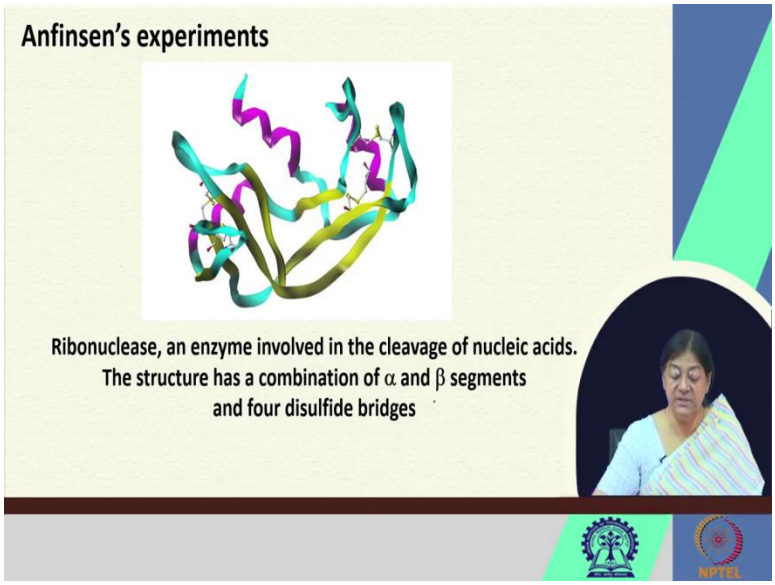
(Refer Slide Time: 18:33)



When we look at this [refer to slide], this is our unfolded structure, this is the final compact structure of the protein. It will follow several methods known as steepest descent methods, we will not go into the details of that, but it is enough to know that there are large number of unfolded conformations and there is one native conformation.

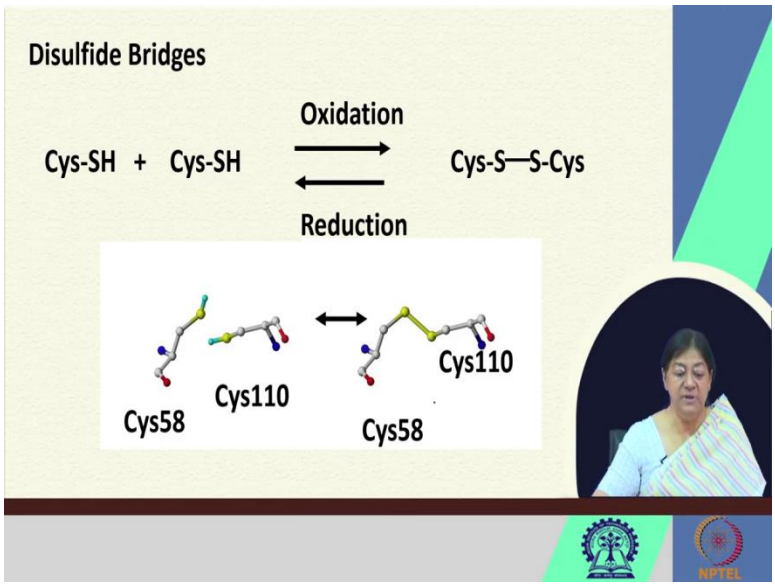
And if we focus on this part, this is the folded state of the protein that is the most stable one and this could be very many other unfolded states that are possible, that give us the resultant final native structure.

(Refer Slide Time: 19:13)



A very interesting experiment known as Anfinsen's experiment, showed us or gave us some important insight about protein folding. In this experiment a protein called ribonuclease, an enzyme that is involved in the cleavage of nucleic acid was used for the experiment. The structure has a combination of α and β segments and four disulfide bridges.

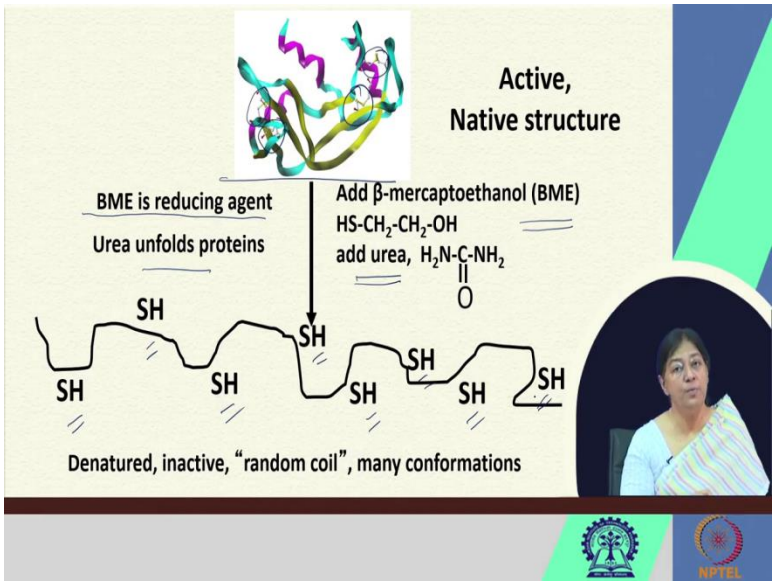
(Refer Slide Time: 19:39)



The experiment is such, if we remember what disulfide bridges were, there were two cysteine residues forming a cystine SS linkage through an oxidation process and we know that we can get back the two cysteine residues by a reduction process. So in this particular protein ribonuclease there are four such disulfide bonds. One of them is between cysteine 58 and cysteine 110 forming this disulfide linkage.

Which means that this [refer to slide] is a 124 amino acid protein, which means that residue number 58 and residue number 110 come close together in three-dimensional space in the formation of the disulfide linkage.

(Refer Slide Time: 20:29)



In the folded protein, there is an active native structure. Now how do we know that this [refer to slide] protein is active and this is the native structure. Because it has a particular function to cleave RNA.

So if we know that this folded protein that we have is cleaving RNA, we know that it is in its native form and is active. We can see the disulfide linkages marked here. So, there are four such disulfide linkages in this protein.

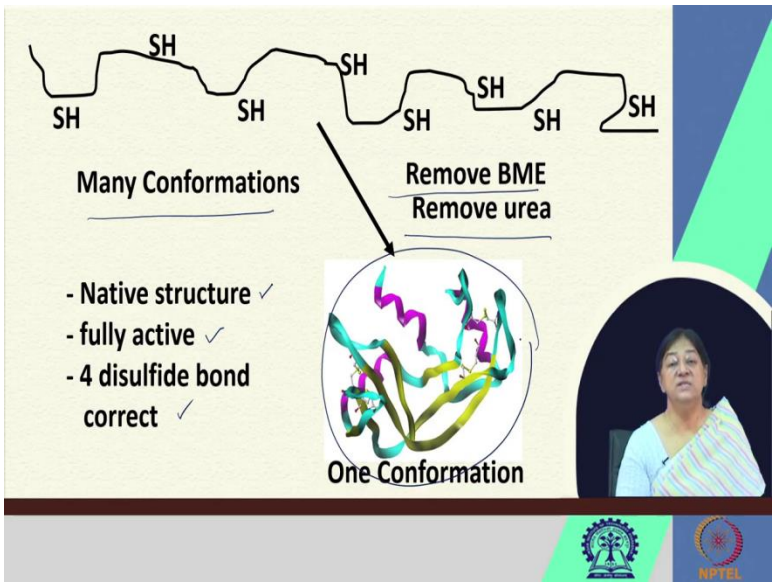
Now, the experiment went as such that there was BME added to the protein and there was urea added to the protein. Why? BME is a reducing agent. In the reduction process what is happening? The SS linkage is going back to forming its individual SH SH as we saw in the previous slide. And what urea does? It unfolds proteins. We will be doing this in this module, where we will be looking at what we mean by protein denaturation as to how we can unfold the folded protein.

So the experiment was to break down the disulfide linkages by adding BME and adding urea to unfold the protein. So what do we now have? We have a denatured, inactive, random coil with

many many possible conformations. So the 4 disulfide bonds have now been reduced. We have 8 of these SHs, these are the cysteine residues that we have.

The experiment was conducted in a manner that since now the SHs were free, they could interact with any other SH to form a different disulfide bond. The identification that we had 58 and 110 that formed a specific disulfide bond 58, now can pair with another cysteine to form a different disulfide bond.

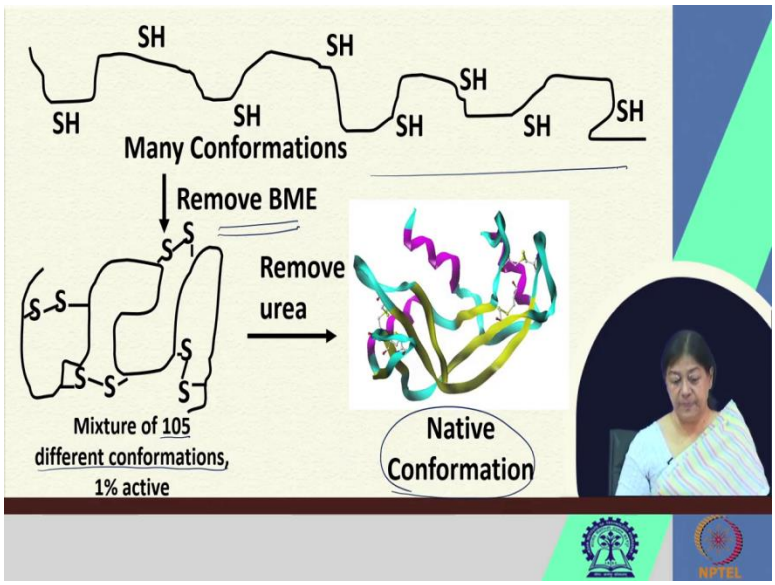
(Refer Slide Time: 23:01)



So what was the experiment? The experiment was about the many conformations that were possible. BME was removed allowing an oxidation process to occur forming the disulfide linkages again. Urea was removed to allow the protein to refold. The native structure was returned, it was fully active and the 4 disulfide bonds that were there in the original protein the native conformation, gave this one conformation indicating that the disulfide linkages that were reduced to form the SHs, reformed to form the native structure.

Now how was it known that this is the native structure? From the activity.

(Refer Slide Time: 24:03)

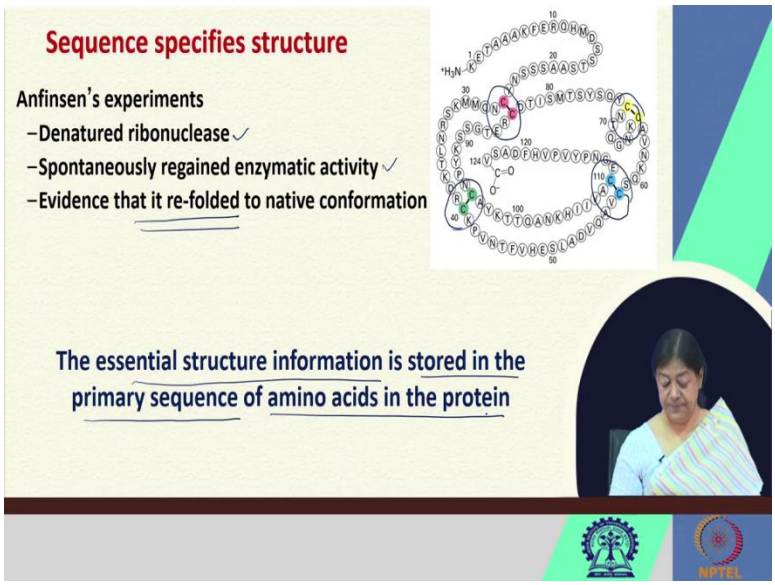


So of the many conformations, if we remove the BME the disulfide linkages can be formed. And this actually gives us 105 different conformations that is 1 percent active. How do we know that? Because the protein is not in its correctly folded form. Then urea is removed and we get the native conformation as was mentioned.

So our problem here is, that we had the native conformation with 4 disulfide linkages. The disulfide linkages were reduced to give us this particular polypeptide chain and the protein was allowed to unfold.

Then the process was reversed. The beta mercaptoethanol, the BME was removed so that the disulfide linkages could reform, the urea was removed so the protein could refold and it was found that the same disulfide linkages formed and the protein refolded back to its native conformation.

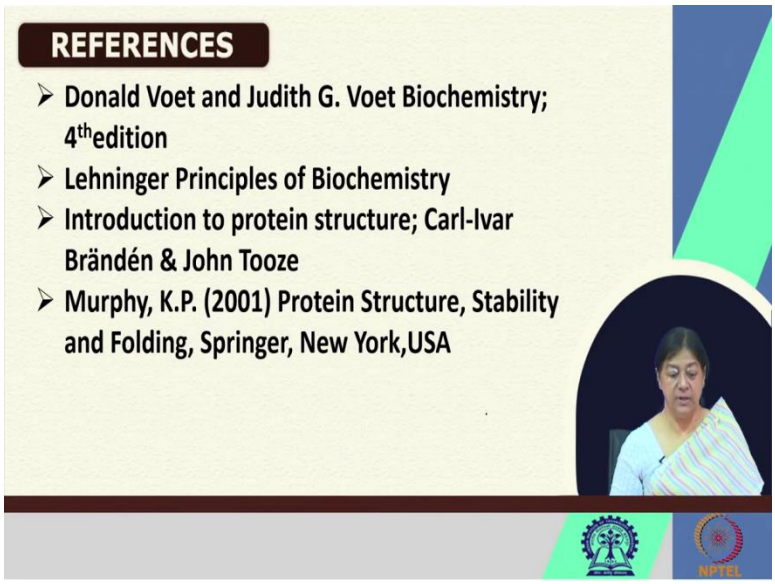
(Refer Slide Time: 25:21)



So this is a beautiful experiment that actually indicated that the sequence specifies the structure of the protein. What do we have? We had denatured ribonuclease, it spontaneously regained its enzymatic activity and this was evidence that it refolded to its native conformation. The cysteines that we have marked here [refer to slide] are the green, the pink, the yellow and the blue. When these bonds were broken and the protein was unfolded the green recognized the green, the blue recognized the blue and so on and so forth.

The conclusion is that the essential structure information is stored in the primary sequence of amino acids in the protein. So, the protein folding problem means that we have our primary structure, the sequence of the amino acids and the information that is going to make it fold to get to its final native three-dimensional structure; all the information is present in that polypeptide chain.

(Refer Slide Time: 26:49)



[Refer to slide] these are the references.

Thank you.