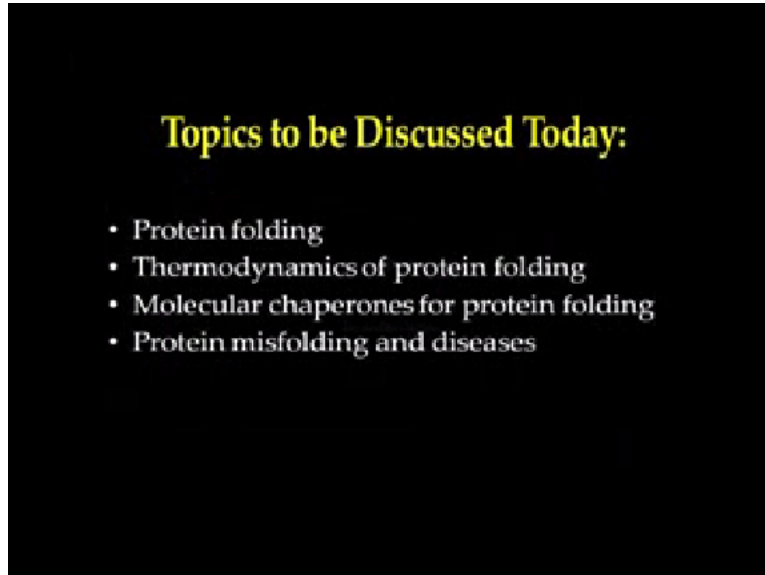


**Introduction to Proteomics**  
**Dr. Sanjeeva Srivastava**  
**Department of Biosciences and Bioengineering**  
**Indian Institute of Technology - Bombay**

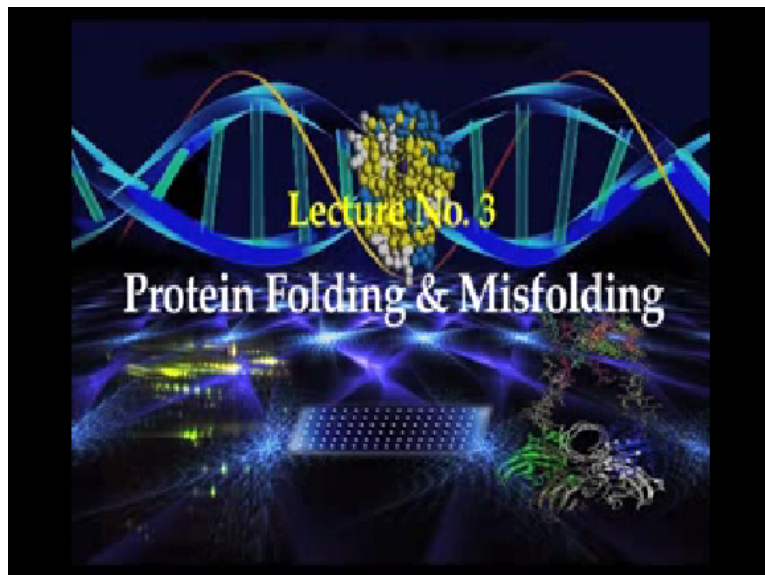
**Lecture – 03**  
**Protein folding & misfolding**

(Refer Slide Time: 00:14)



Today we will talk about proteins, its folding and misfolding.

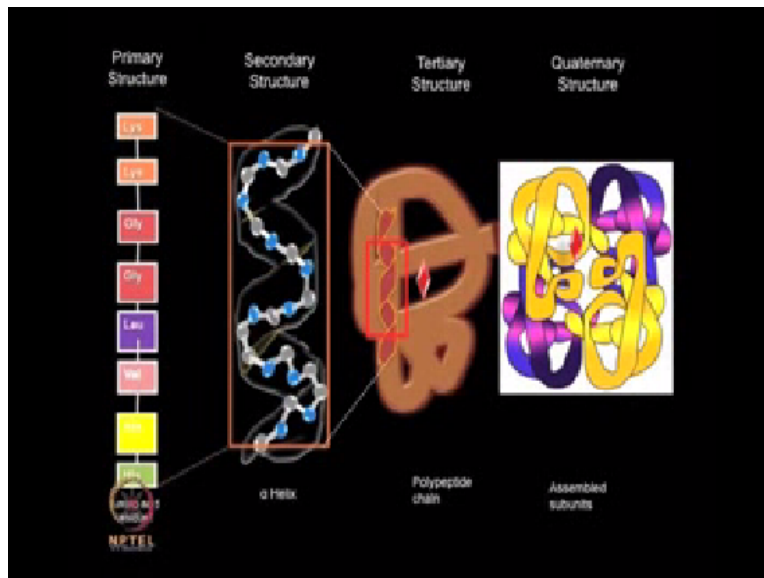
(Refer Slide Time: 00:20)



Understanding the processes of protein folding and misfolding have been a major research area

from last several decades in biology, chemistry and physics. Understanding protein folding and misfolding remains challenging and continues to motivate researchers to work both experimentally and theoretically in this area. In today's lecture, I will present and discuss the basics of protein folding, how this process works and how much folding may result in to various manifestation of diseases.

**(Refer Slide Time: 01:16)**



As we talked in the last class, proteins play a very crucial role in essential characteristics of living systems, how they function, how they replicate themselves through the intricate molecular interactions. Proteins are most important classes of molecules which are involved in promoting and regulating essentially all the reactions which takes place in living systems. We discussed previously the globular proteins, they can fold into conformations of ordered secondary and tertiary structures.

The interactions which govern the formation of secondary, tertiary and quaternary structures involve different forces and interactions. The cumulative effect of all of these interaction and forces are such that the folded proteins, the magnitude of the favourable reactions or interactions will be outweighing the sum of unfavourable interactions and the result, it governs the protein folding. So in today's lecture, we will talk about how protein folding occurs.

Some of the basic thermodynamics of protein folding concepts, how molecular chaperones

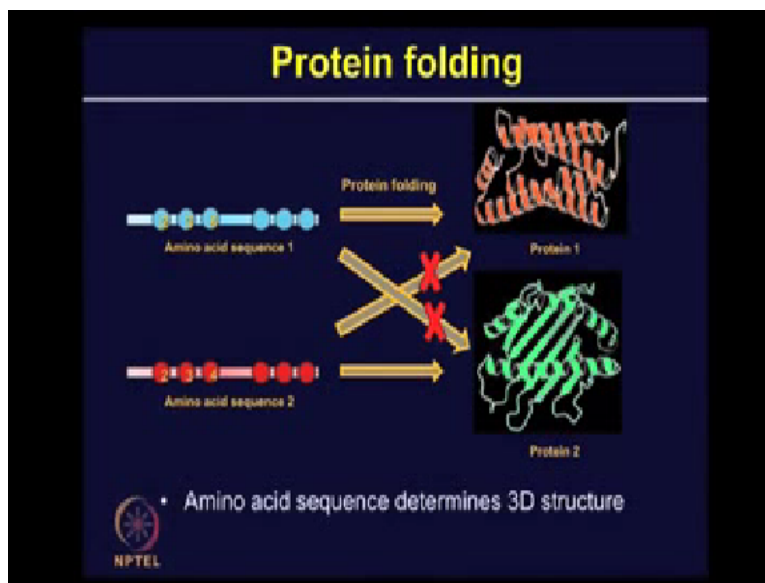
govern the protein folding process and how protein misfolding may result into various diseases.  
(Refer Slide Time: 03:10)

# Section I

## Protein folding

Let us talk about protein folding. Understanding the mechanism by which protein folding takes place still remains challenging for the scientific community. Protein folding provides an elegant example of biological self-assembly and understanding such complex machinery provides very critical information not only for the understanding of protein folding but also the evolutionary aspects of proteins and various biomolecules.

(Refer Slide Time: 03:56)



In protein folding, the amino acid sequence determines the 3-dimensional structure. Now as you can see here, the proteins having very much specificity. If you have an amino acid 1 shown in

blue colour, that will form protein 1 shown in the right side. If you have amino acid sequence 2 shown in red, that will form protein 2. Now if you take the amino acid sequence 1, protein 2 cannot be generated. Similarly, if you take amino acid sequence 2, protein 1 cannot be generated.

So there is very highly specificity of amino acid sequence which can determine the 3-dimensional structure of proteins. The protein folding process is governed by distribution of polar and non-polar amino acids. If you remember last class, we have talked about various amino acids. The polar side chains, they tend to arrange themselves near outside of the molecules. You take for example arginine, glutamine, histidine.

Similarly, on the non-polar side chains, they have tendency to cluster in the interior of molecules. For example, phenylalanine, leucine, valine and tryptophan.

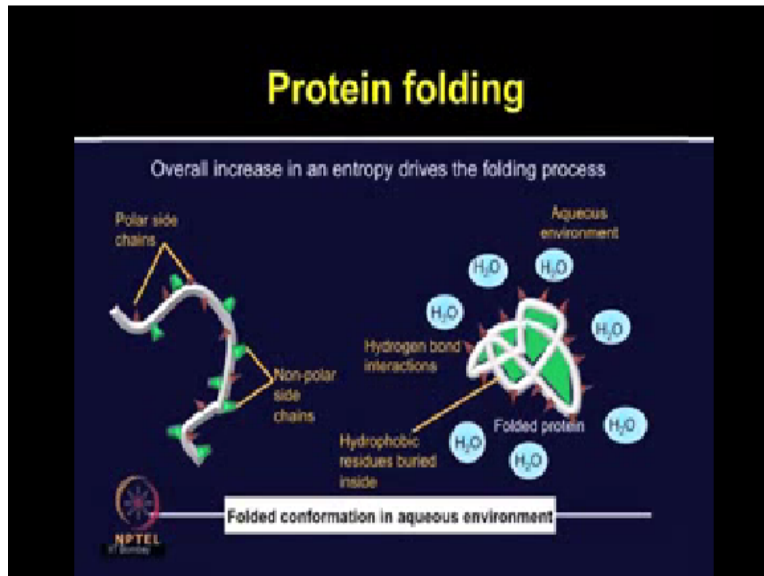
**(Refer Slide Time: 05:42)**

**Polar and non-polar amino acids**

Amino Acid	Abbreviation	Symbol	Hydrophobicity/ Charge
Aspartic acid	Asp	D	Negative
Glutamic acid	Glu	E	Negative
Arginine	Arg	R	Positive
Lysine	Lys	K	Positive
Histidine	His	H	Positive
Asparagine	Asn	N	uncharged polar
Glutamine	Gln	Q	uncharged polar
Serine	Ser	S	uncharged polar
Threonine	Thr	T	uncharged polar
Tyrosine	Tyr	Y	uncharged polar
Cysteine	Cys	C	non-polar
Glycine	Gly	G	non-polar
Isoleucine	Ile	I	non-polar
Leucine	Leu	L	non-polar
Methionine	Met	M	non-polar
Phenylalanine	Phe	F	non-polar
Proline	Pro	P	non-polar
Tryptophan	Trp	W	non-polar
Valine	Val	V	non-polar
Alanine	Ala	A	non-polar

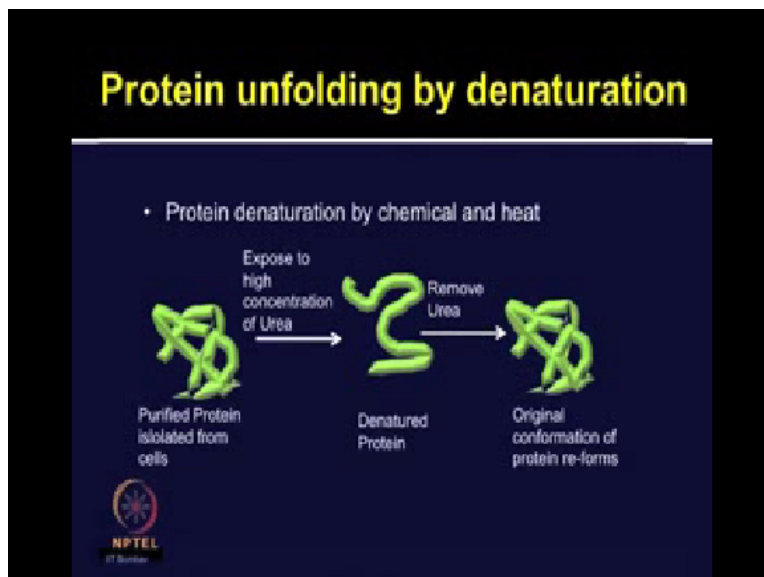
This chart is only for your information which shows the various amino acids which belong to polar and non-polar category and then you can think of how they are going to govern the protein folding process. So continuing on protein folding, the hydrophobic amino acids, they are driven to associate the hydrophobic collapse.

**(Refer Slide Time: 06:11)**



So when these amino acids come together as you can see on the right-hand side, the loss of water surrounding these amino acids increases entropy of the system; therefore, overall increase in an entropy drives the folding process.

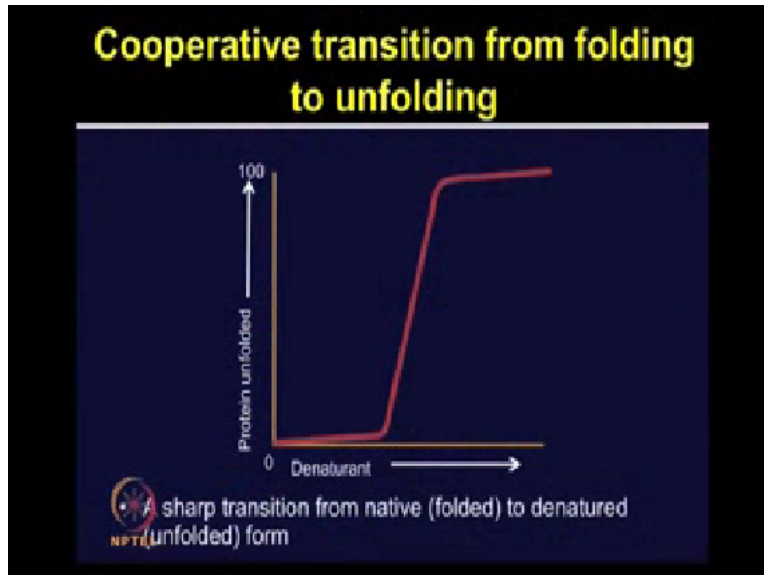
**(Refer Slide Time: 06:32)**



Now as you have seen in the classical experiment of Anfinsen, the protein unfolding can be done by using denaturants. Say if you take denaturants whether it is chemical like urea and guanidinium chloride or you heat treat it. So as you can see here, if you have a purified protein isolate taken from the cells and you expose it to the high concentration of denaturants whether it is chemical or heat, that will result into the denatured protein shown in the centre.

If you remove the denaturing condition, it will again form the proper folding. Protein conformation will be restored in its original form.

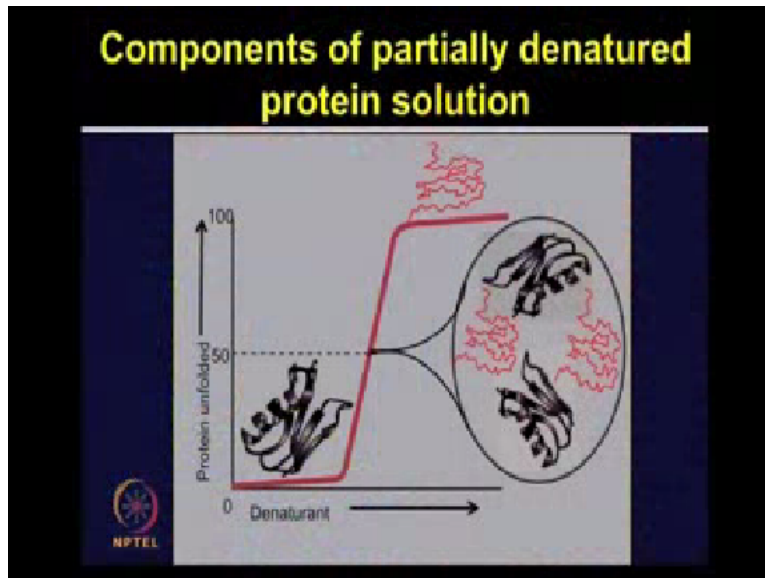
**(Refer Slide Time: 07:25)**



So how this process of folding to unfolding works. The various hypothesis and mechanism have been proposed. Let us talk about cooperative transition here from folding to the unfolding form. As you can see in this graph, on the Y axis, the protein in the unfolded form from 0-100 and on the x-axis, the presence of denaturant, a sharp transition from the native or the folded to the denatured or unfolded forms occur.

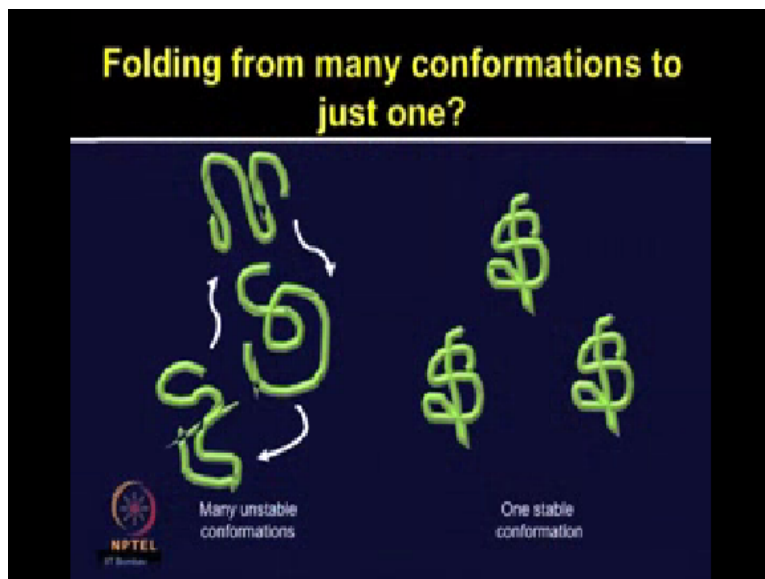
So only 2 conformational states are present significantly whether it is folded form or unfolded form. If denaturants are removed from the unfolded protein, it allows protein to make folded forms.

**(Refer Slide Time: 08:24)**



So what are the components of partially denatured protein. If you look at this graph, in the transition state at 50%, it will be 50% fully folded and 50% unfolded form of the protein; however, existence of only 2 states, the folded and unfolded or possibility of unstable transient intermediates between the folded and unfolded states still remains a topic of research in protein folding area. So how folding occurs from many conformations to only 1 form. A particular sequences along the polypeptide backbone, they impose key restrictions.

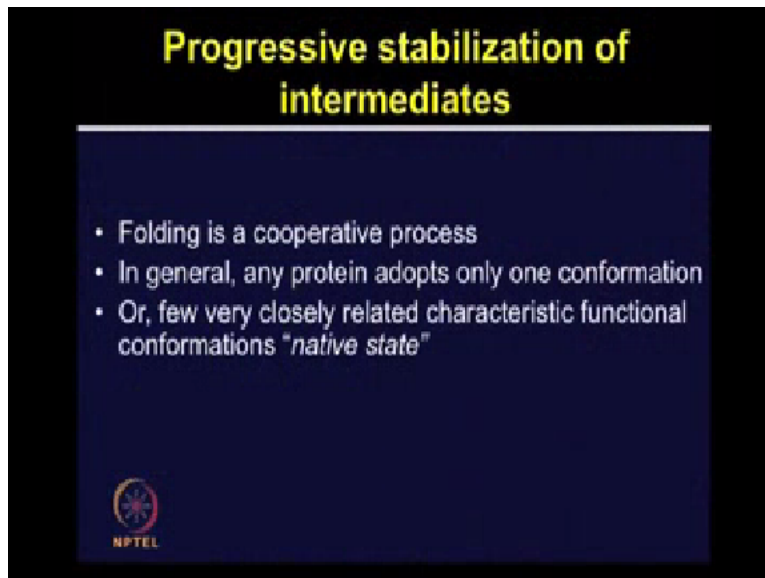
**(Refer Slide Time: 09:27)**



The various properties of the side chains which we have talked in the previous lecture including size, hydrophobicity, ability to form hydrogen and ionic bond, all of these govern this process. Let us take example of arginine. A side chain with positive charge might attract a segment of the

polypeptide which has complementary negative charge. You take for example aspartic acid. So these type of side chains and various type of backbone properties are going to impose key restrictions; therefore, various type of folded conformations will be selected and it can result only one which is going to govern the folding process.

**(Refer Slide Time: 10:20)**

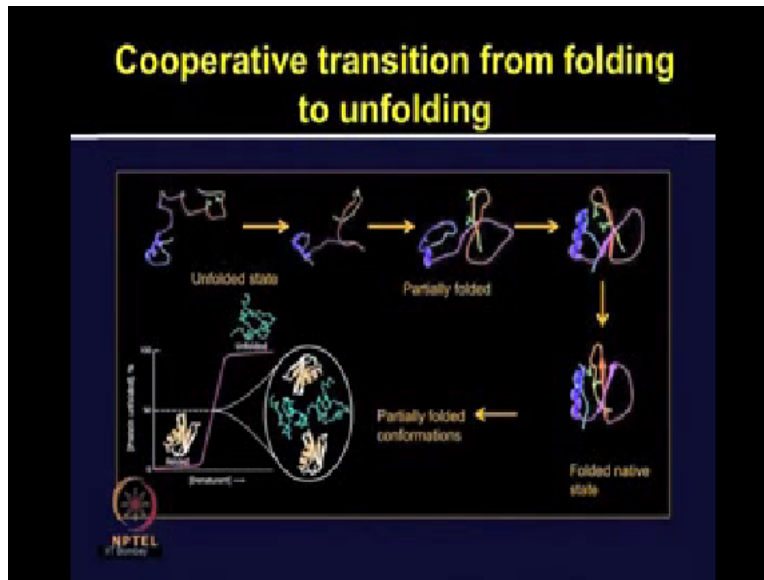


So there are various progressive stabilisation of intermediates occur in the folding process. As we talk, folding is a cooperative process which involves progressive stabilisation of various intermediates. In general, any protein adopts only one conformation which we just talked in the last slide or few very closely related characteristic functional conformations may occur which will give rise to the native state.

Native state in this context here will be the conformation which has the lowest free entropy or the most stable folded form for majority of the proteins.

**(Refer Slide Time: 11:16)**





So let us see in the slide how cooperative transition occurs from the folding to unfolding. Folding is a cooperative process which arises from simultaneous formation of multiple interactions within a polypeptide chain. If you take individually, so each interaction is weak but their cooperative formation drives polypeptide chains towards the folded state.

**(Refer Slide Time: 11:49)**

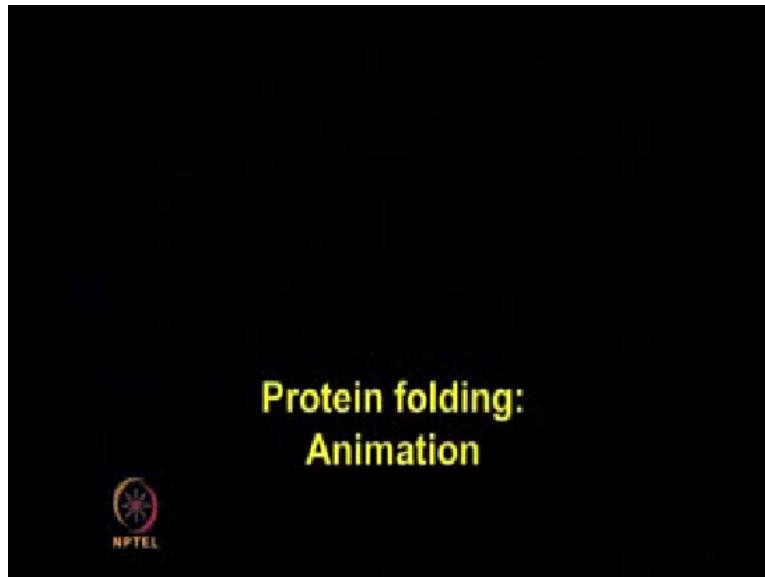
### Prediction of protein structure

- Amino acid sequence dictates protein structure
- knowledge-based and *Ab initio* "from the beginning" prediction to predict protein structure

So how to do structure prediction of proteins. As we have seen in the previous experiment, the amino acid sequence dictates the protein structure. So theoretically, the prediction of final folded structure is possible from its sequence; however, the long range of interactions and vast number of possible conformations which are possible and therefore, it limits these type of predictions. However, knowledge-based and every ab initio "from the beginning" prediction do take place to

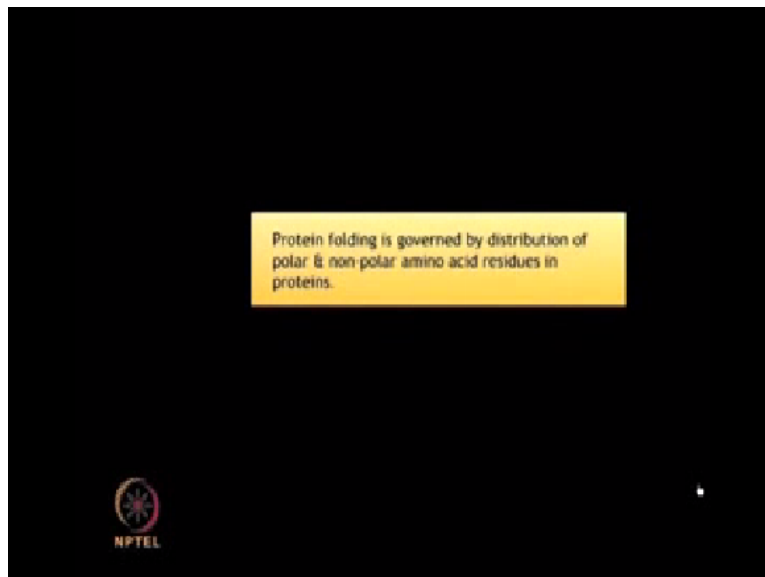
predict the protein structure.

**(Refer Slide Time: 12:36)**



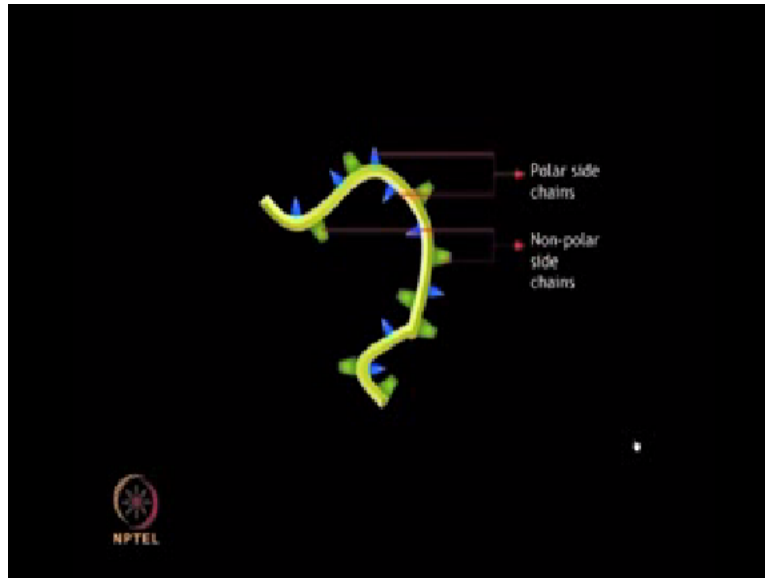
So let me show you the protein folding process in following animation.

**(Refer Slide Time: 12:42)**



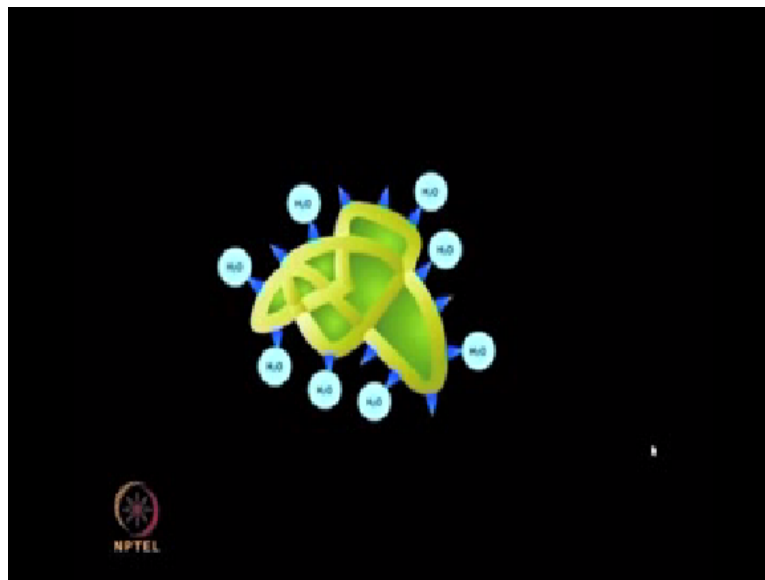
The process of protein folding is dictated by the distribution of polar and non-polar amino acids residues in the protein.

**(Refer Slide Time: 12:53)**



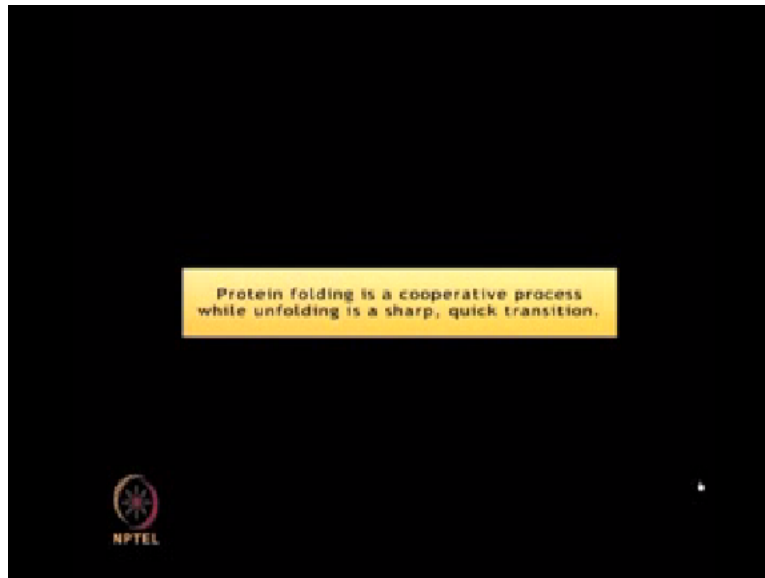
The hydrophobic amino acids are driven to interact with one another by a process known as hydrophobic collapse.

**(Refer Slide Time: 13:03)**

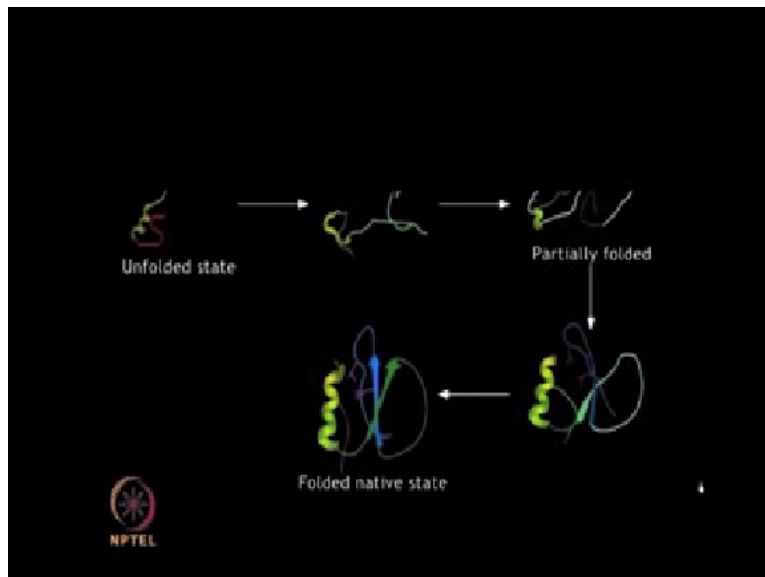


They come together and during this process, eliminate water molecules surrounding them. The polar residues remain on the surface and form hydrogen bonds with water molecules while the hydrophobic residues get buried within the core of protein.

**(Refer Slide Time: 13:29)**

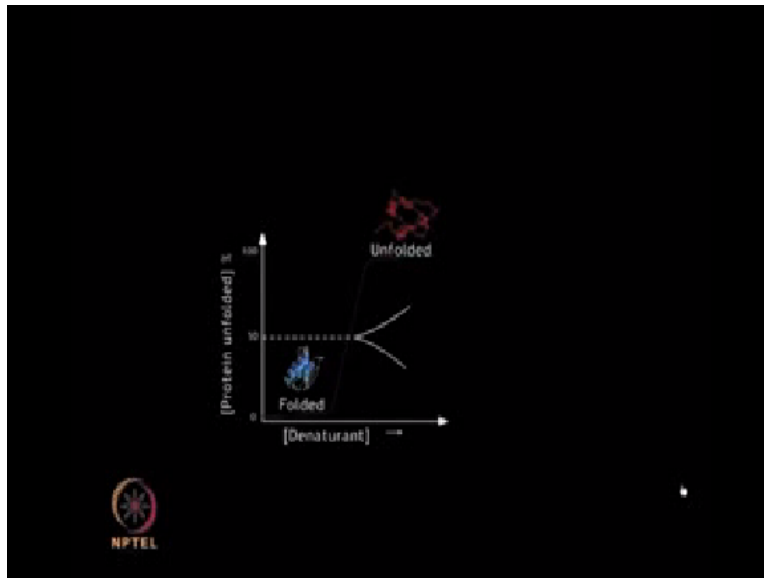


Protein folding is a cooperative process whereas the unfolding is a sharp and quick transition.  
**(Refer Slide Time: 13:38)**



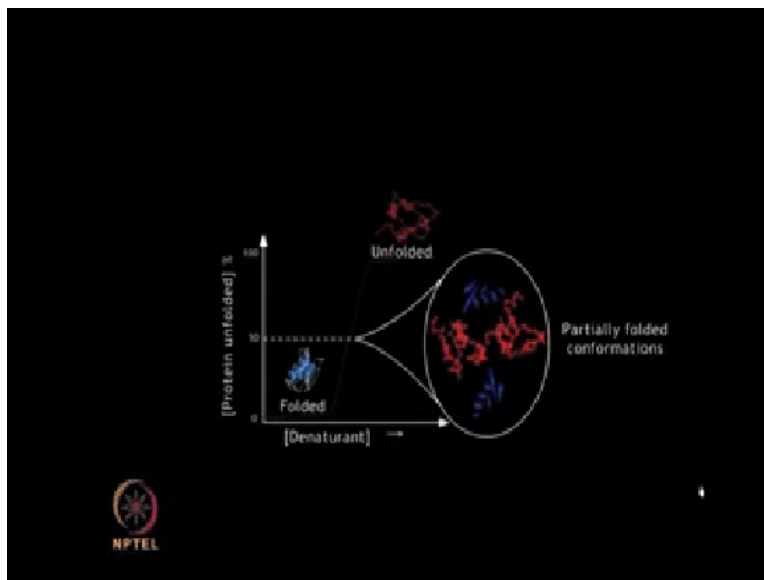
Proteins typically adopt only one characteristic functional native state conformation which has lowest free energy and it is more stable. Folding is limited to one conformation due to properties of the amino acid side chain such as hydrophobicity, size, shape, etc. Folding is highly cooperative process wherein there is progressive stabilization of the intermediates as you can see here. Although it is theoretically possible to predict the protein structure from the amino acid sequence.

**(Refer Slide Time: 14:26)**



Several long-range interactions can often limit such predictions. Here on x-axis, denaturants are plotted and y-axis, the percentage protein unfolded is plotted.

**(Refer Slide Time: 14:45)**



On 0, you can see that is totally folded form of the protein. On 100%, it is unfolded form, but if you take a mixture from 50%, that is either unfolded or folded form, which shows that protein can assume either folded form or unfolded form of the proteins.

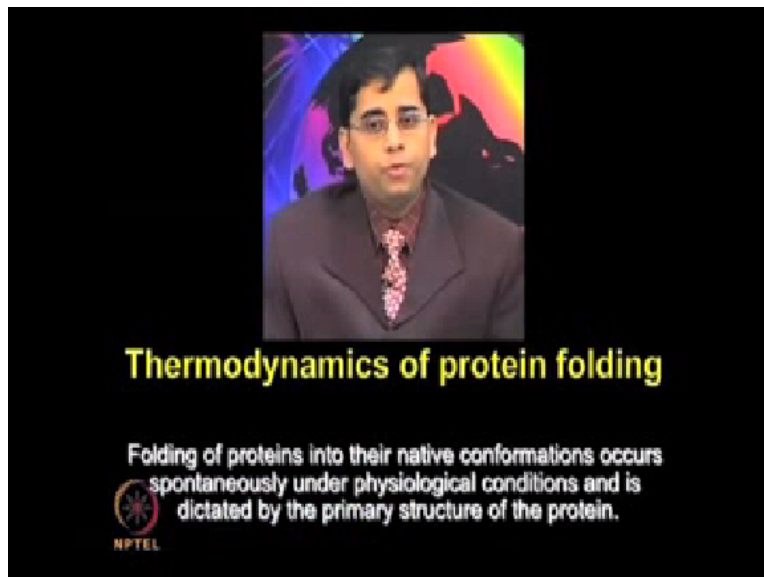
**(Refer Slide Time: 15:12)**

## Section II

### Thermodynamics of protein folding

Let us now talk about thermodynamics of protein folding.

**(Refer Slide Time: 15:22)**



**Thermodynamics of protein folding**

Folding of proteins into their native conformations occurs spontaneously under physiological conditions and is dictated by the primary structure of the protein.

NPTL

The folding of proteins into their native conformations occur spontaneously under physiological conditions and is dictated by the primary structure of the protein. Protein folding is thermodynamically favourable process where decrease in free energy from unfolded to folded state occurs.

**(Refer Slide Time: 15:52)**

## Thermodynamics of protein folding

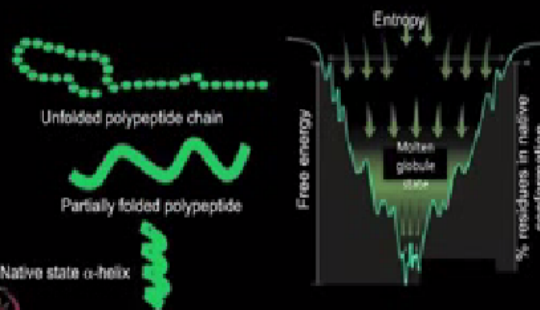
- Hydrophobic amino acids are driven to associate-hydrophobic collapse
- Thus, overall increase in an entropy drives the folding process



First talk about some of the basics of thermodynamics for protein folding. As we have seen earlier, the hydrophobic amino acids, they are driven to associate hydrophobic collapse; therefore, overall increase in an entropy drives the folding process.

**(Refer Slide Time: 16:17)**

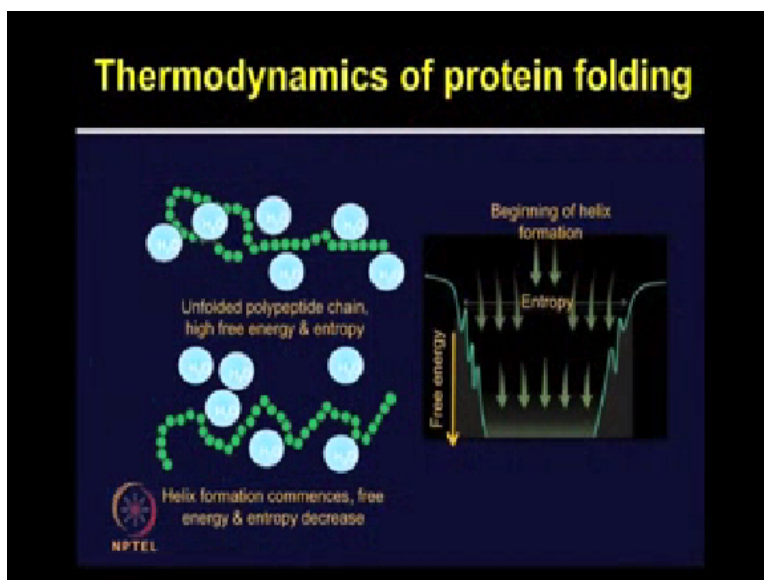
## Thermodynamics of protein folding



As you can see in this complex picture here, the folding process can be explained as free energy funnel thermodynamically. If you look at the right-hand side, the open mouth of funnel represents the wide range of structures which are accessible to the ensemble of denatured proteins. The initial collapse state of protein with very little thermodynamic stability is known as molten globule.

The amino acid side chains are extremely disordered in this state and several fluctuations can be observed as you can see from these arrows. As free energy of protein molecules decreases, the protein molecules move down to the narrower part of the funnel, look at the bottom part here and only few conformations can be accessible here. So at the bottom of the final, well-defined and folded confirmation states are present.

**(Refer Slide Time: 17:30)**



So if you look at unfolded polypeptide chain, so the amino acids that have been joined together by the peptide bonds but they have not yet formed their secondary or tertiary structure. So this confirmation has highest free energy and entropy. The amino acids in the polypeptide chain start interacting by means of hydrogen bonds across the polypeptide backbone in order to initiate the folding process.

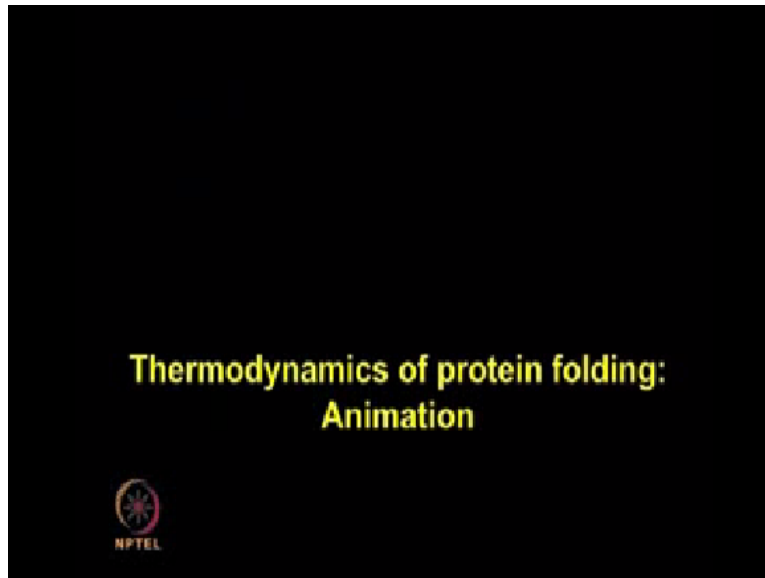
The free energy and entropy of the system gradually decreases as folding takes place. The entropy of the polypeptide chain decreases during this process. So in the thermodynamic terms, the lowering of entropy is favoured by a corresponding increase in entropy in the surroundings composed of the water molecules. Now if you look at the funnel again, as the protein continues to fold in order to assume a stable low-energy native state conformations, the entropy also decreases.

While it may appear unfavourable for the system; however, entropy of the surrounding water



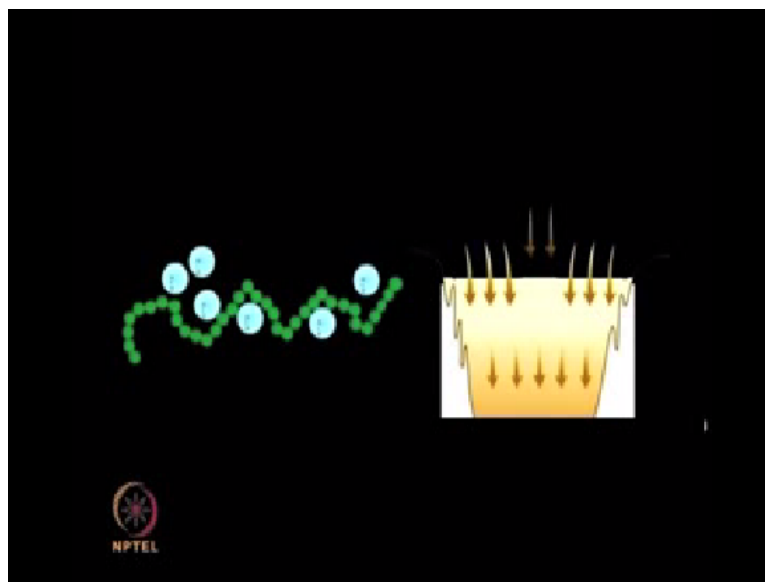
molecules increases in this process and it increases overall entropy and makes it favourable and spontaneous.

**(Refer Slide Time: 19:15)**



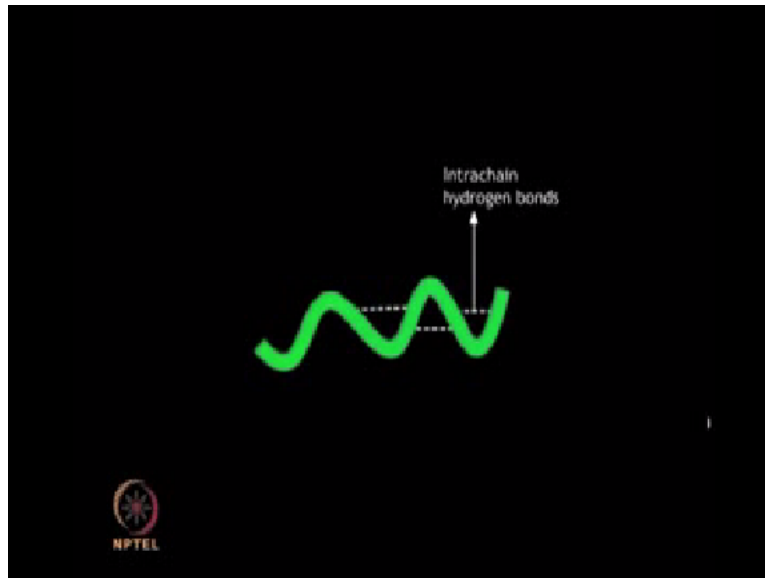
So let me show you how protein folding works and how it can be described in the thermodynamics terms in following animation.

**(Refer Slide Time: 19:26)**



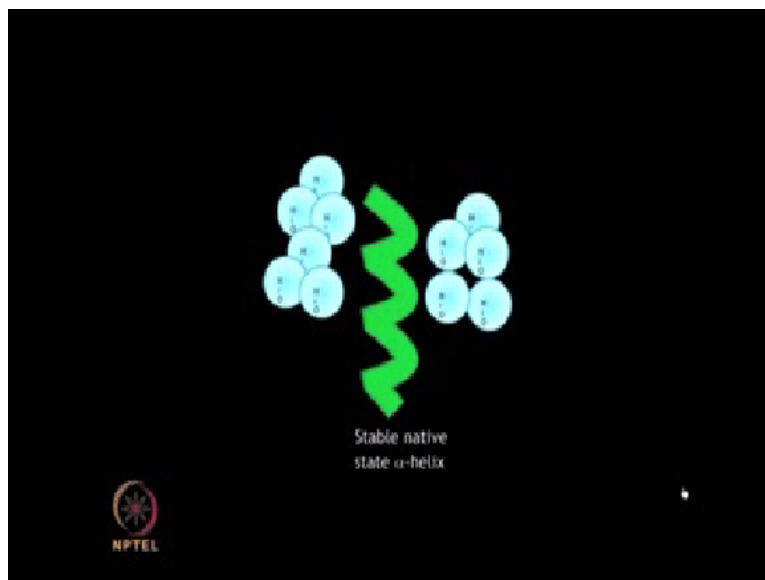
An unfolded polypeptide chain has very high energy and entropy. The protein folding acts to decrease the free energy of the system by forming favourable re-interactions and assuming a more stable state. The entropy of the polypeptide chain decreases during this process.

**(Refer Slide Time: 19:53)**



As the protein continues to fold in order to assume a stable low energy native state conformation, the entropy also decreases.

**(Refer Slide Time: 20:03)**



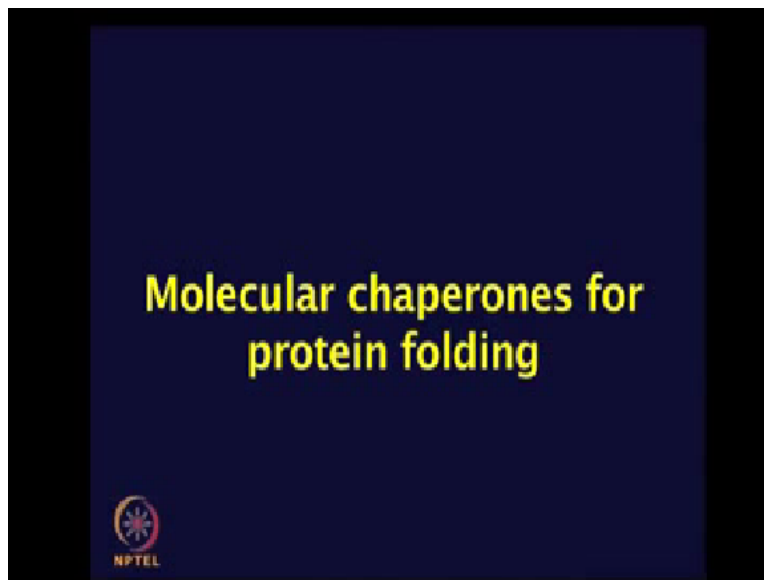
As the protein continues to fold in order to assume a stable low-energy native state conformation, the entropy also decreases. While this should seem unfavourable for the system, it must be recalled that the entropy of the surrounding water molecule increases during the process thereby increasing overall entropy and making it favourable and spontaneous.

**(Refer Slide Time: 20:37)**

## Section III

### Molecular chaperones for protein folding

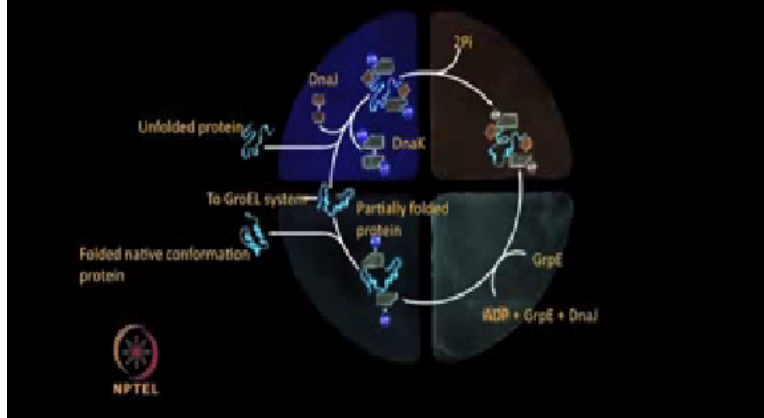
(Refer Slide Time: 20:43)



Let us now talk about molecular chaperones for protein folding. The molecular chaperones are a class of heat inducible proteins which provides kinetic assistance in protein folding process. They prevent protein aggregation and promote protein folding by binding to the hydrophobic surfaces which are exposed in non-native protein conformations.

(Refer Slide Time: 21:17)

## Molecular chaperone systems



So let us talk about various molecular chaperone systems. Mainly newly synthesised proteins form folded structures in vivo spontaneously and without any assistance; however, folding efficiency could be limited by various processes such as protein aggregation which are promoted by the transiently exposed hydrophobic surfaces.

In response to the heat-shock, the cells produce significant amount of unfolded proteins by synthesising new systems which are known as molecular chaperones which are designed to promote the protein folding process. The several molecular chaperones systems which have been exemplified in *E. coli* includes GroEL/GroES system, DnaK, DnaJ, GrpE and ClpB.

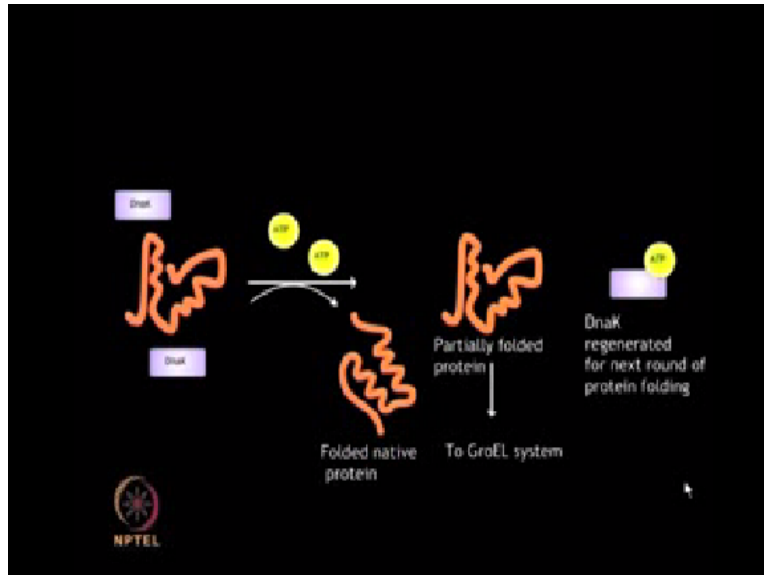
**(Refer Slide Time: 22:27)**

Molecular chaperone for protein  
folding:  
Animation



The bacterial chaperone in GroEL, it binds proteins in non-native state and allows enzymes to be recovered quantitatively in native form by binding which requires co-chaperone GroES and ATP. Let me show you how this chaperone works and govern the protein folding process in following animation.

**(Refer Slide Time: 23:00)**



The unfolded protein is bound by DnaJ and then by DnaK which is an ATP-bound protein. The hydrolysis of ATP into ADP and Pi by DnaK is stimulated by the energy DnaJ. This resulting DnaK, ADP remains tightly bound to the unfolded protein. The nucleotide exchange factor GrpE present in bacteria facilitates release of ADP along with DnaJ. This leaves the DnaK bound to the partially folded protein which continues to undergo folding to a more favourable low energy conformation.

Once the protein gets completely folded, it gets detached from DnaK which then binds ATP again and completes the cycle and prepare it for next round of protein folding. Any protein which may not have been folded completely, is then taken over by the GroEL chaperone system which completes the folding.

**(Refer Slide Time: 24:31)**

## Section IV

### Protein misfolding and diseases

After talking about how protein folding works, let us discuss about protein misfolding and how misfolding may result into various diseases. So protein misfolding results into large number of human diseases which arise as a consequence of protein misfolding. In protein folding, mutations cause defective folding, aberrant assembly and incomplete processing which results into altered folding properties.

**(Refer Slide Time: 25:16)**

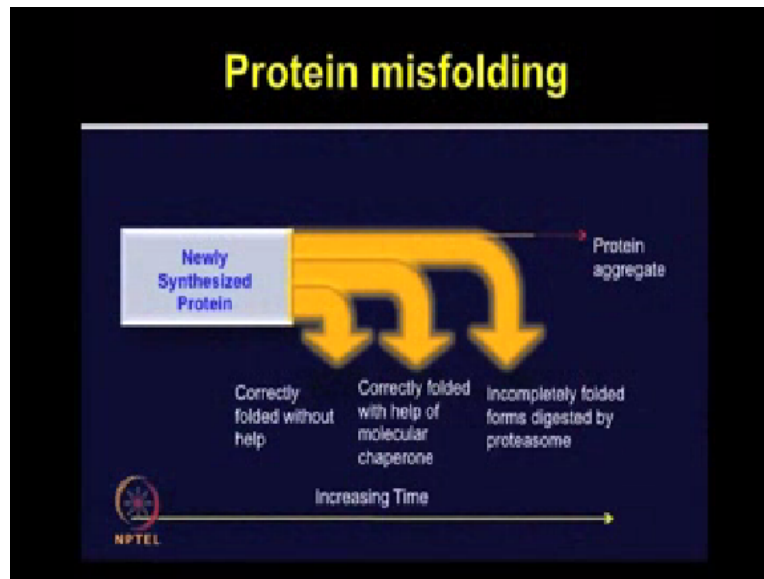
### Protein misfolding

- Protein folds into a single, energetically favorable conformation, specified by its amino acid sequence
- A protein may fold into alternative 3D structure due to mutations, inappropriate covalent modifications



Proteins fold into a single, energetically most favourable confirmation which is specified by its amino acid sequence. A protein may fold into alternative 3-dimensional structures because of mutations or inappropriate covalent modifications; therefore, protein misfolding may lead to loss of normal protein function.

(Refer Slide Time: 25:57)



So in the slide, we can see from a newly synthesised protein, it could have various state. It may form the proper folded form without any assistance or it can give correct folded forms in pretence of molecular chaperones or in completely folded forms can be digested by proteasome machinery or it may result into the protein aggregation. So a newly synthesised protein may give rise to any one on these forms depending on various factors which are going to govern the protein folding process.

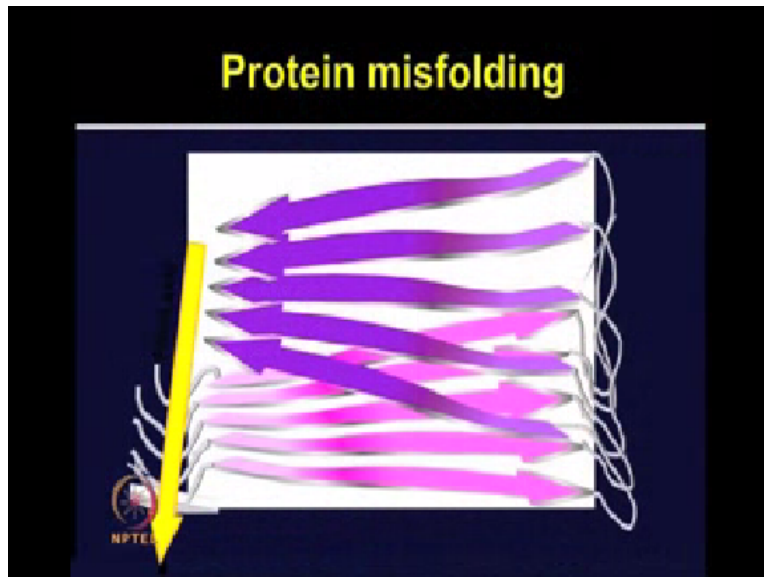
(Refer Slide Time: 26:36)



So accumulation of these misfolded proteins or proteolytic fragments can result into few degenerative diseases. These degenerative diseases are characterised by the presence of insoluble

protein plaques in organs such as brain and liver. For example, Alzheimer disease in human as well as Parkinson disease. Bovine spongiform encephalopathy also known as mad cow disease in cows and scrapie disease in sheep.

**(Refer Slide Time: 27:17)**



So in this slide, it is shown that amyloid fibers are involved in neurodegenerative diseases and the protein aggregation due to the large beta sheets are deduced from solid-state NMR. In Alzheimer disease, the presence of beta-amyloid containing plaques is associated with neurodegeneration and dementia. In other neurodegenerative diseases, it has also been shown that it involves protein aggregation.

Prion diseases such as Jakob disease and BSE or bovine spongiform encephalopathy are associated with amyloid deposit of PRP proteins. So how insoluble protein aggregates can result into various diseases.

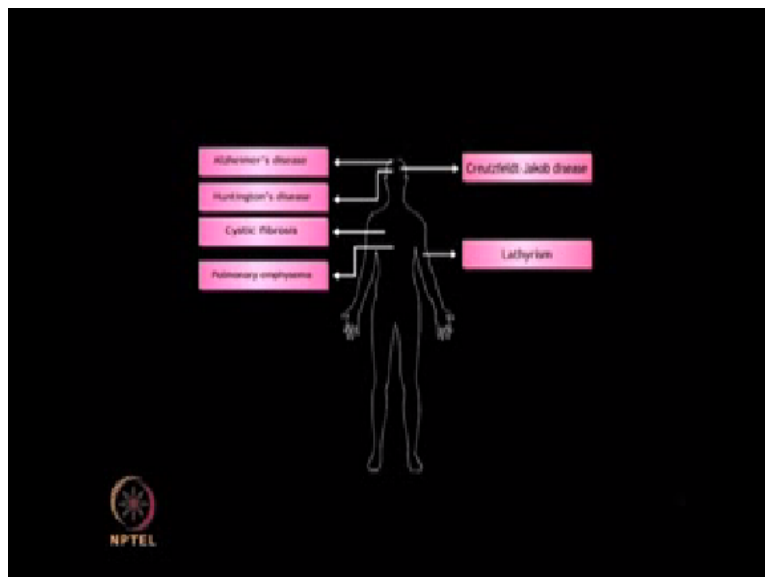
**(Refer Slide Time: 28:16)**





Let us discuss some of protein misfolding related disease in following animation.

**(Refer Slide Time: 28:28)**



Protein misfolding results into various diseases such as Alzheimer disease. In Alzheimer's disease, structure of certain normal soluble cellular proteins which are normally rich in alpha helical regions are converted into beta strand conformations which further link with each other to form beta sheet aggregates known as amyloids. The insoluble amyloid plaques are essentially made up of single polypeptide chain or fibrils known as amyloid b-protein.

It is observed in the brain of patients with Alzheimer where dead or drying neurons are surrounded with plaques. A neurotoxicity is believed to be caused by the amyloid fibrils before

they get deposited as amyloid plaques. This disease presents various symptoms such as memory loss, decreased neuromuscular coordination, confusion and dementia. Jakob disease, it was initially believed to be caused by viruses or bacteria; however, later it was discovered to be transmitted by small proteins known as prions.

The prion proteins are composed of beta sheet structures that have been modified from previously existing alpha helices. The protein aggregates of one abnormal protein is sufficient to function as a nuclei for other normal proteins to attach. Huntington's disease, it is a neurodegenerative disorder of genetic origin which affects muscular coordination. It is caused by increased number of trinucleotide repeats, CAG in Huntington gene leading to increased number of glutamine residues incorporated in corresponding protein.

This alters the folding of Huntington protein which has highest concentration in brain and testis. The exact function of this protein is unclear but it is known to interact with several other proteins. The mutated protein has also been found to have effects on chaperone proteins which in turn helps to fold several other proteins. Cystic fibrosis, this is an autosomal recessive disorder caused by mutation in gene for the protein cystic fibrosis transmembrane conductance regulator or CFTR.

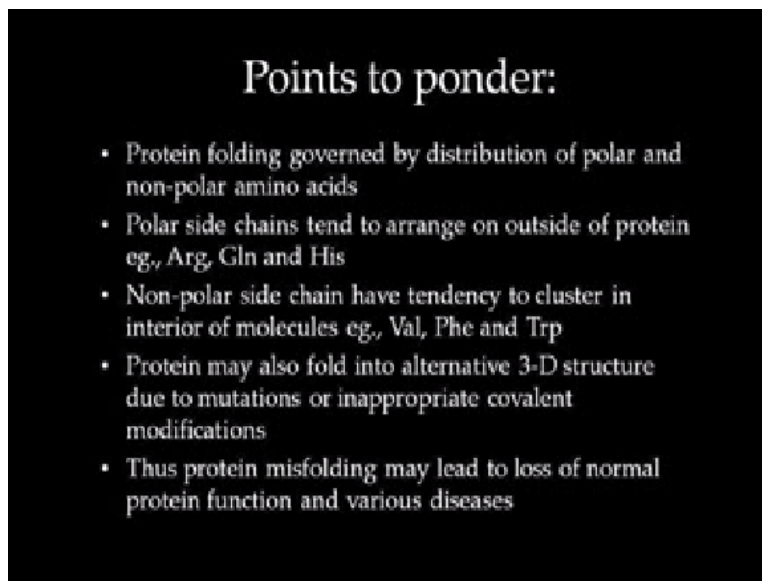
The CFTR gene regulates components of sweat, digestive juices and mucus. It is caused by a deletion of 3 nucleotides leading to the elimination of the phenylalanine residue from the protein and therefore results into abnormal folding. The dysfunctional protein gets degraded by the cell. Pulmonary emphysema, it is a progressive disease of lung which causes shortness of breath. It can be caused by deficiency of the protein alpha 1 antitrypsin or A1AT.

The A1AT gene is responsible for protection of lung tissue from damage by enzyme neutrophil elastase. Abnormally secreted A1AT gets accumulated in liver thereby allows lung tissue damage. The disease results in to wheezing, shortness of breath and asthma-like symptoms. Lathyrism, it is regular ingestion of seeds from sweet pea, *Lathyrus odoratus*, which causes disruption of cross-linking in muscle protein, collagen.

Collagen is very important structural protein which has triple helical structure. The cross links formed are due to the oxidation of lysine residues by the enzyme lysyl oxidase to form allysine. These are essential for proper folding of collagen and giving it the required strength. B- aminopropionitrile present in abundance in sweet pea deactivates this enzyme by binding to its active site.

This prevents cross-linking and proper folding of the protein. It may also result in muscle rigidity and weakness.

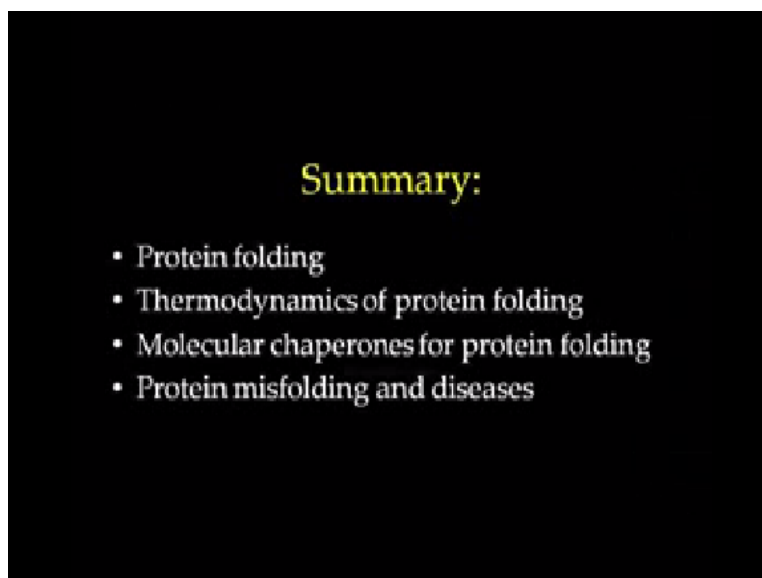
**(Refer Slide Time: 34:50)**



**Points to ponder:**

- Protein folding governed by distribution of polar and non-polar amino acids
- Polar side chains tend to arrange on outside of protein eg., Arg, Gln and His
- Non-polar side chain have tendency to cluster in interior of molecules eg., Val, Phe and Trp
- Protein may also fold into alternative 3-D structure due to mutations or inappropriate covalent modifications
- Thus protein misfolding may lead to loss of normal protein function and various diseases

**(Refer Slide Time: 35:10)**



**Summary:**

- Protein folding
- Thermodynamics of protein folding
- Molecular chaperones for protein folding
- Protein misfolding and diseases

So in summary, was talked about protein folding and how various polar and non-polar side chain restrict and govern the process of protein folding. We then looked at the thermodynamics of protein folding albeit very briefly. We talked about entropy and how it governs the protein folding. The molecular chaperones, we talked about some classical example in animations and then we discussed about protein misfolding and described some of the diseases which may result due to the protein misfolding.

We will continue our discussion about basics of protein structure and function in next class as well. Thank you.

**(Refer Slide Time: 36:07)**



**(Refer Slide Time: 36:13)**

## REFERENCES

- Horwich A. Protein aggregation in disease: a role for folding intermediates forming specific multimeric interactions. Volume 110, Issue 9 (November 1, 2002). J Clin Invest. 2002;110(9): 1221–1232. doi:10.1172/JCI16781.
- Harold A. Scheraga. Protein structure and function, from a colloidal to a molecular view. Carlsberg Research Communications. January 1984, Volume 49, Issue 1, pp 1-55.
- Berg J., Tymoczko J. & Stryer L., Biochemistry fifth ed., W. H. Freeman & company, 2002. ISBN: 0716746840.
- Nelson D. & Cox M., Lehninger, Principles of Biochemistry fourth ed., 2004. W. H. Freeman and company. ISBN: 023022699X.
- Creighton T. E., Protein folding first ed., W. H. Freeman and company, 1992. ISBN: 071677027X.
- Voet D. & Voet J., Biochemistry fourth ed., Wiley, 2000. ISBN: 047158651X.