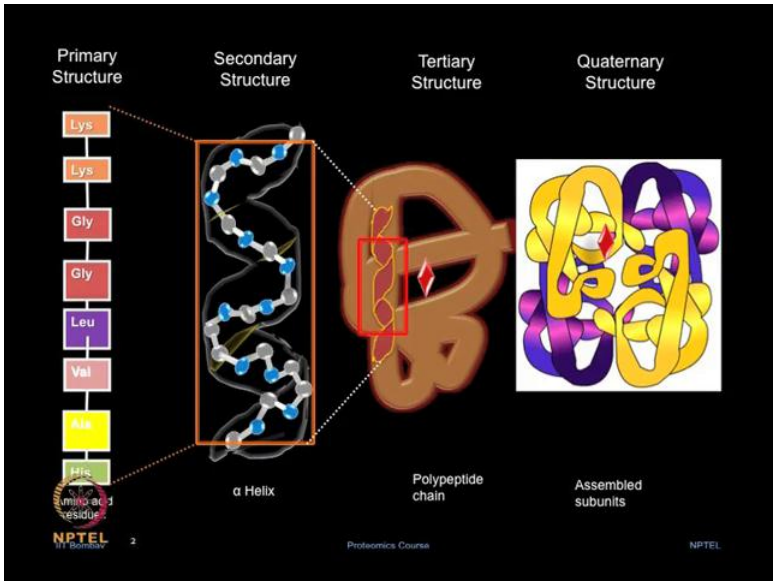
**Proteomics: Principles and Techniques**
**Prof. Sanjeeva Srivastava**
**Department of Biosciences and Bioengineering**
**Indian Institute of Technology, Bombay**

**Lecture No. # 05**
**Proteins: Folding and Misfolding**

Welcome to the proteomics course. Today, we will talk about proteins, it is folding and misfolding. Understanding the processes of protein folding and misfolding have been a major research area from several decades in biology, chemistry, and physics. Understanding protein folding and misfolding remains challenging, and continuous 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 misfolding may result into various manifestation of diseases.

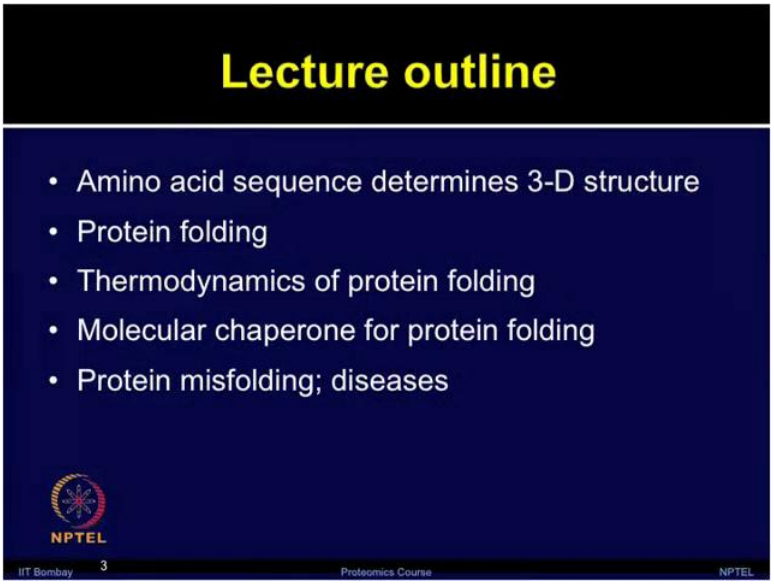
(Refer Slide Time: 01:23)




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 integrations. 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 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 involved different forces and interactions. The cumulated effect of all of these interactions and forces are

such that the folded proteins, the magnitude of the favorable reactions or interactions will be out wing the sum of unfavorable interactions, and the result it governs the protein folding.

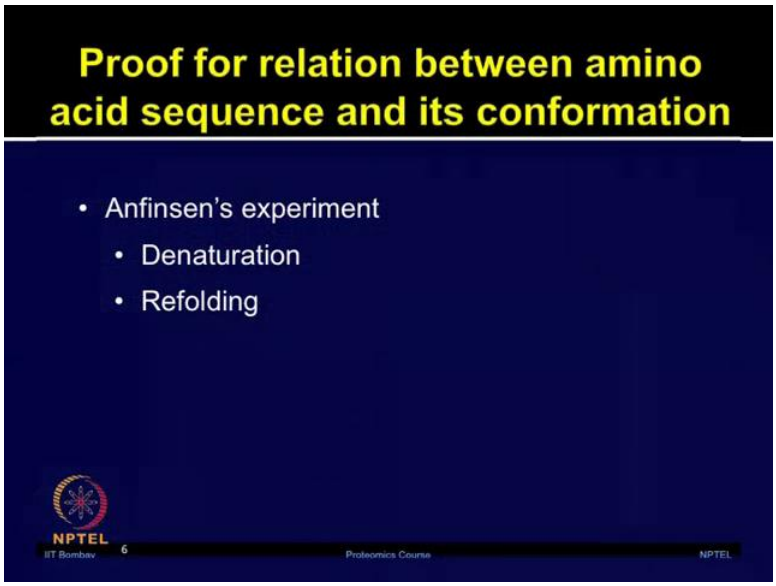
(Refer Slide Time: 03:02)




So, on today's lecture, we will talk about how amino acid sequence determines three-dimensional protein structures. I will explain you a classical experiment done by Anfinsen. We will talk about how protein folding occurs, some of basic thermodynamics of protein folding concepts, molecular chaperones govern the protein folding process, and how problem misfolding may result into various diseases. So, let us start our first topic amino acid sequences determines three-dimensional structures of proteins. So, there is very intricate sequence structure relationship, the amino acid sequence dictates the confirmations which are adopted by the polypeptide chains at secondary and tertiary levels.

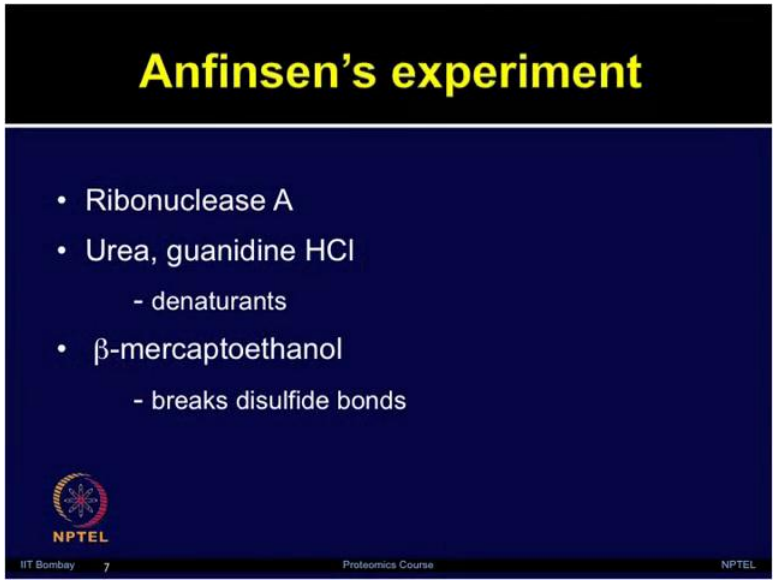
Scientist Anfinsen he did a classical experiment where he tested the ability of reduced and unfolded proteins to spontaneously fold into native it is state by using a protein rib nuclease A. The experiment established that the primary amino acid sequence of a protein contains all the information which is required for the proper protein folding into its native form.

(Refer Slide Time: 05:07)



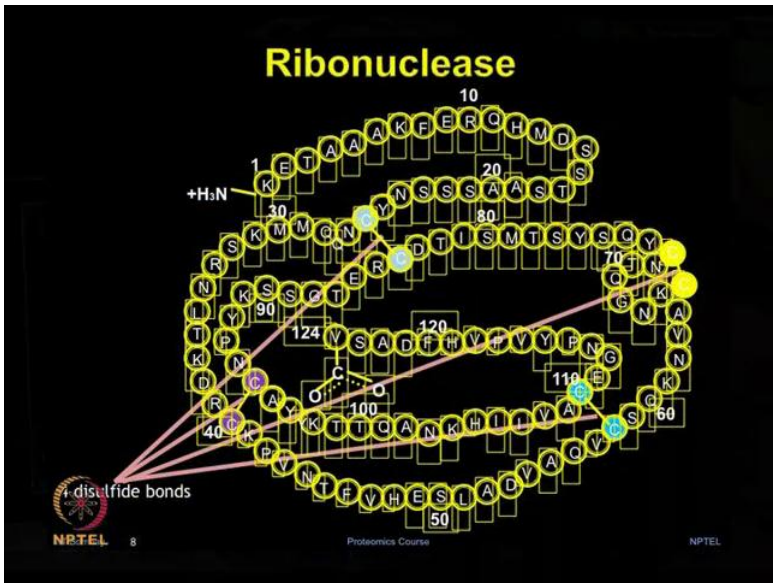

The fundamental discovery of Anfinsen let him to receive the Nobel Prize in chemistry in 1972. So, let me explain you how these experiment work to establish the proof for relationship between amino acid sequence and it is conformation Christian Anfinsen in 1950's performed an experiment where he performed two steps denaturing and refolding. So, how denaturation and refolding works?

(Refer Slide Time: 05:31)



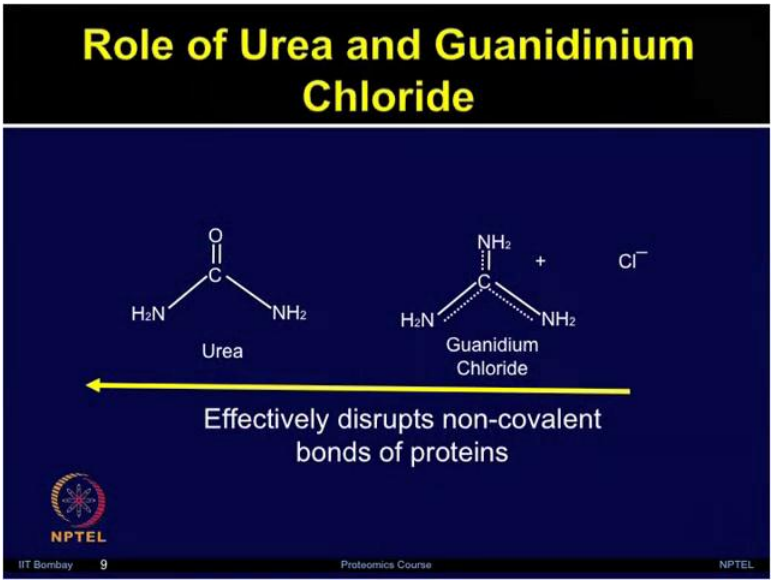
So, in this classical experiment Anfinsen used protein Rib nuclease A he used flow denaturants such as urea or guanidine hydrochloride and beta mercaptoethanol which breaks the disulfide bonds.

(Refer Slide Time: 05:55)



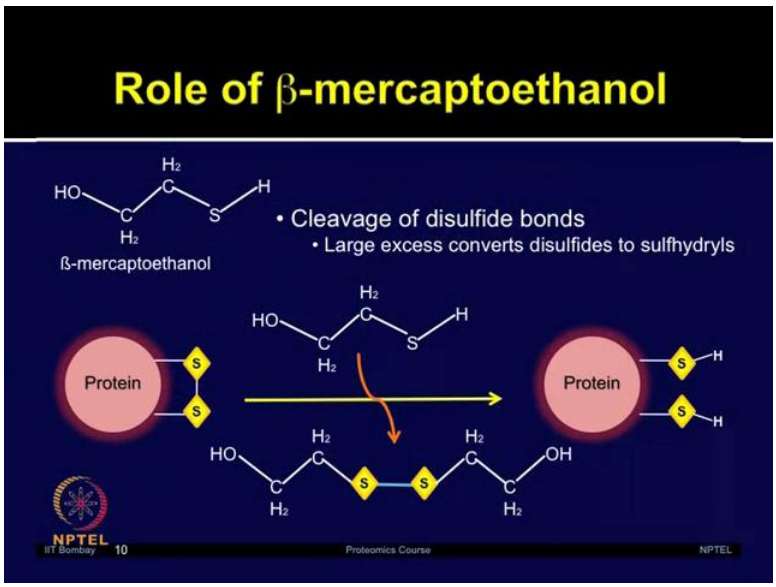

So, let us look at each one of these component a little bit more detail first talk about Rib nuclease a protein. So, this protein has contributed greatly to our understanding of protein folding in vitro for the landmark experiment of Anfinsen. As you can see in the structure Rib nuclease has 124 amino acid residues and it forms 4 disulphide bridges which are located between the 16 residues of 20 sic and 84, 40 and 95, 58 and 110 and 65 and 72 this protein catalysis the hydrolysis of RNA.

(Refer Slide Time: 06:45)

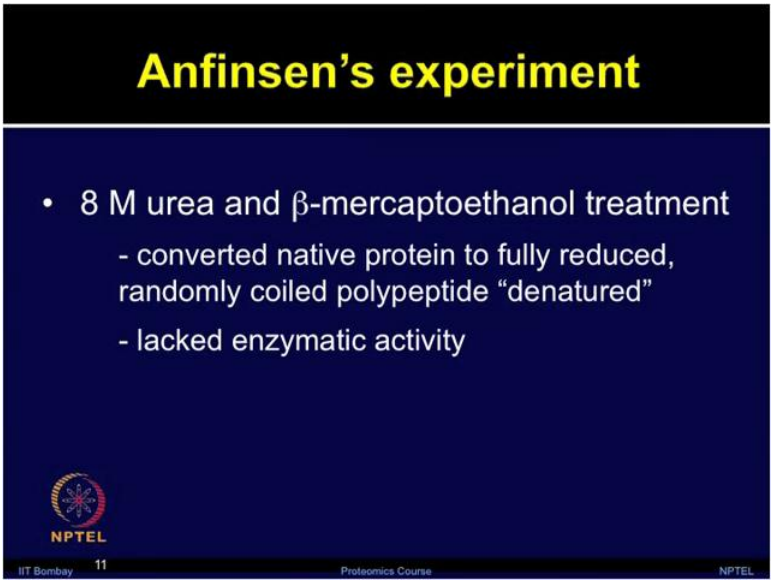



So, what is the role of urea and guanidinium chloride? Urea is an organic compound which has 2 amino groups joined by a carbonyl group and used at a concentration of around 6 molar for denaturing the proteins by breaking the non-covalent interactions, both urea and guanidinium chloride can effectively disrupts the proteins non-covalent interactions.

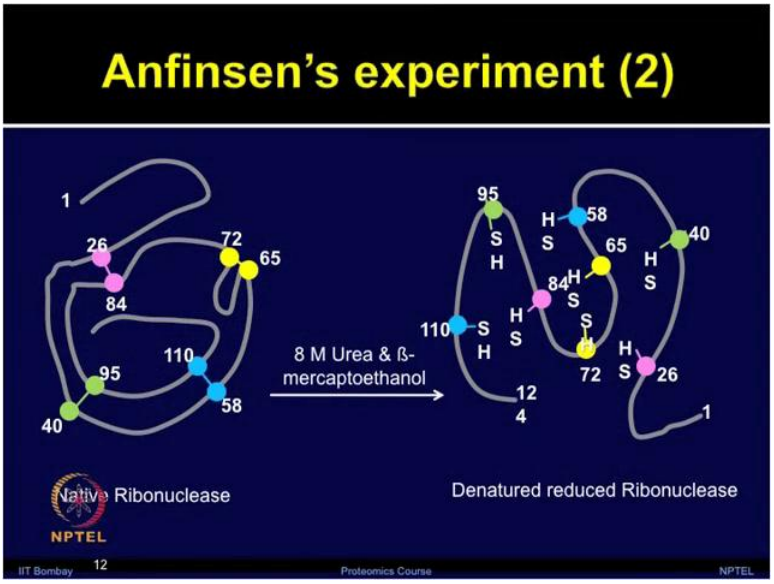
(Refer Slide Time: 07:23)



(Refer Slide Time: 08:07)




(Refer Slide Time: 08:30)

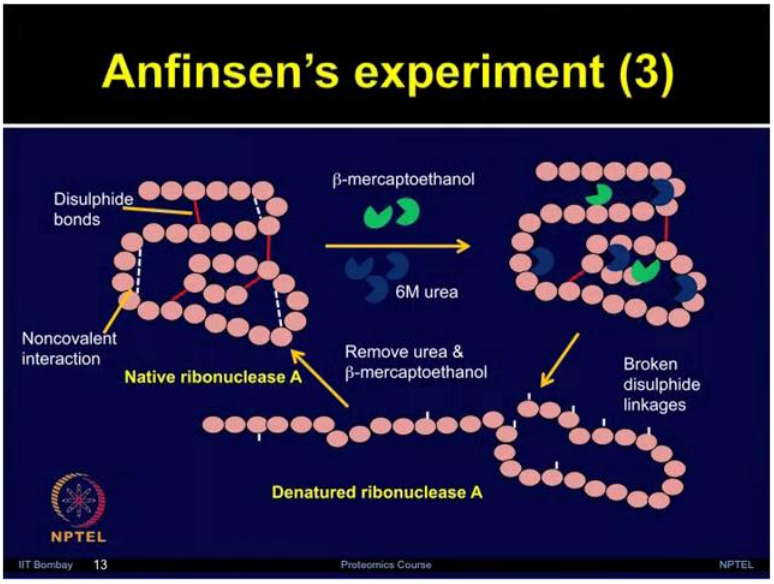



What is a role of beta mercaptoethanol? The beta mercaptoethanol is commonly used to reduce the disulphide linkages in proteins and thereby it disrupts the tertiary and quaternary structures. As you can see in the structure here, in the slide in presence of excess of beta mercaptoethanol the disulphide or cysteines can be fully converted into sulfhydryls or cysteines. It wills Anfinsen's experiment he used 8 molar of urea and beta mercaptoethanol treatment which converted the native proteins to fully reduce a state into the randomly coiled polypeptides known as the denatured structure. The denatured polypeptide lacked enzymatic activity, such we have discussed the ribonuclease protein

it contains 124 amino acid residues and forms 4 disulphide linkages, these linkages are formed between the cysteines as shown here of 26, 84, 40 and 95, 58 and 110 and 65 and 72.

The rib nuclease native conformation is lost when it was treated with eight molar urea and beta mercaptoethanol. As you can see here the native ribonuclease has formed denatured reduced rib nuclease, due to the breaking of disulphide and non covalent interactions. So, on treatment of urea and beta mercaptoethanol ribonuclease a protein lost it is native conformation, because of breaking of disulphide and non-covalent linkages. Anfinsen noticed that when the ribonuclease was oxidized in air and urea removed by the process of dialysis the enzyme activity slowly recovered, and as a result of the protein folding.

(Refer Slide Time: 10:05)

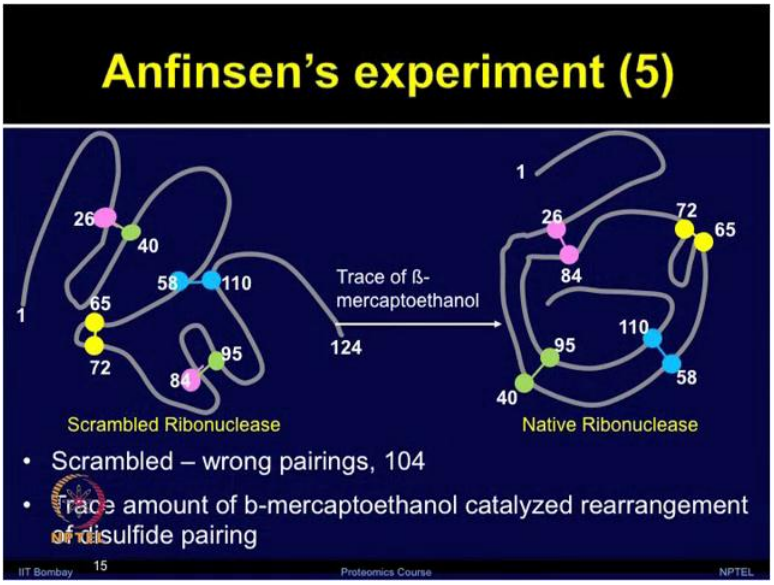



As you can see here, if you have used beta mercaptoethanol at 6 molar urea all the disulphide and covalent bonds are breaking. Once urea is removed then slowly protein folding occurs; it results into the reformation of tertiary structure and active site.

What Anfinsen repeated this experiment in presence of denaturant urea that led to regeneration of less than 1 percent of enzyme activity? So, what could be the reason? In fact, urea prevented the correct disulphide pairing which resulted into the scrambled form scrambled Rib nuclease. Now if you mathematically calculate due to the presence of 4 disulphide bonds here and presence of 8 system residues it can actually give rise to

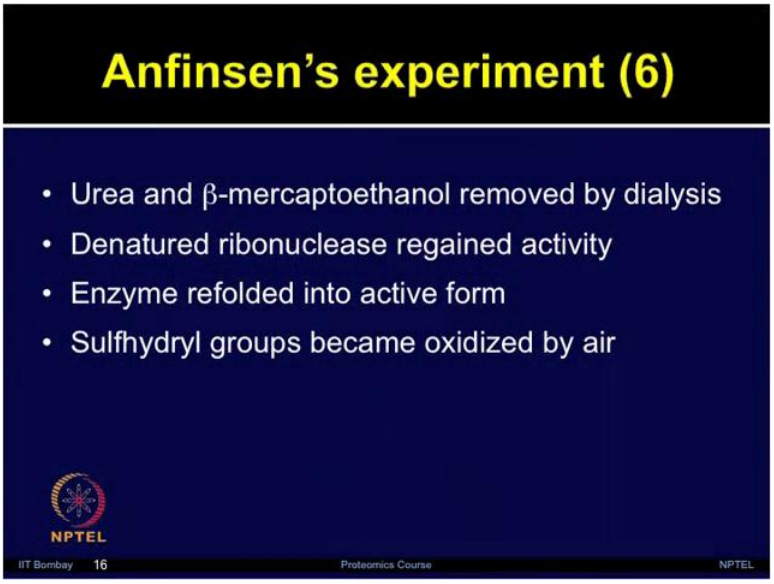
105 different forms in which these for disulphide bonds can be formed. So, in the absence of urea the correct disulphide bridge formation occurred and it allowed folded and thermodynamically it is stable state to be reached in rib nuclease protein.

(Refer Slide Time: 11:42)




Now, this figure shows here that in presence of trace amount of beta mercaptoethanol and complete removal of denaturant urea. The refolding of rib nuclease was accurate, and for intrachain disulphide bonds were reformed in the same positions where they expected in the native rib nuclease. The random distribution of disulphide bonds was obtained when denaturants were used, as you can see in the scrambled state which indicates that weak bonding interactions were required for the correct positioning of disulphide bonds and achieve the native conformation.

(Refer Slide Time: 12:32)



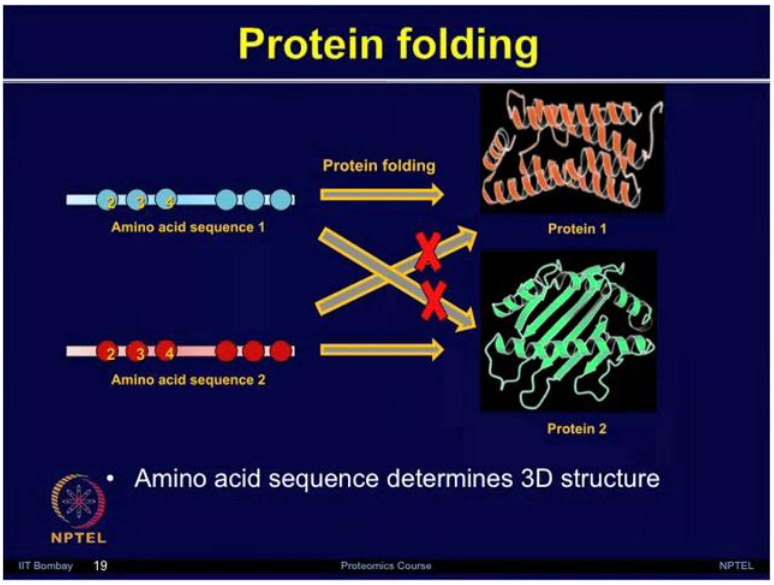

So, in Anfinsen's experiment he removed urea and beta mercaptoethanol by dialysis process. The denatured rib nuclease regained it is enzyme activity. The enzyme was refolded into the active form, and a sulfhydryl groups oxidized in presence of air. The experiments prove that information required for specific catalytic active structure of ribonuclease is contained in it is amino acid sequence.

(Refer Slide Time: 13:15)


The classical study of Anfinsen proved that all the information which is crucial for protein folding resides in it is primary sequence. Let me explain you this experiment in

following animation. In Anfinsen's experiment rib nuclease a in its native state, has 4 disulphide bonds between it is 16 residues. And take at that beta mercaptoethanol and 6 molar urea, the protein undergoes denaturation and the disulphide linkages are broken. The enzyme activity is lost nature state it was observed by Anfinsen that removal of urea and beta mercaptoethanol led to the refolding of enzyme to assume its native state with more than 90 percent enzyme activity being intact. However if only beta mercaptoethanol was removed in presence of urea, the formation of disulphide bonds was random which led to enzyme with only around 1 percent activity. So, after studying the classical experiment of Anfinsen 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.

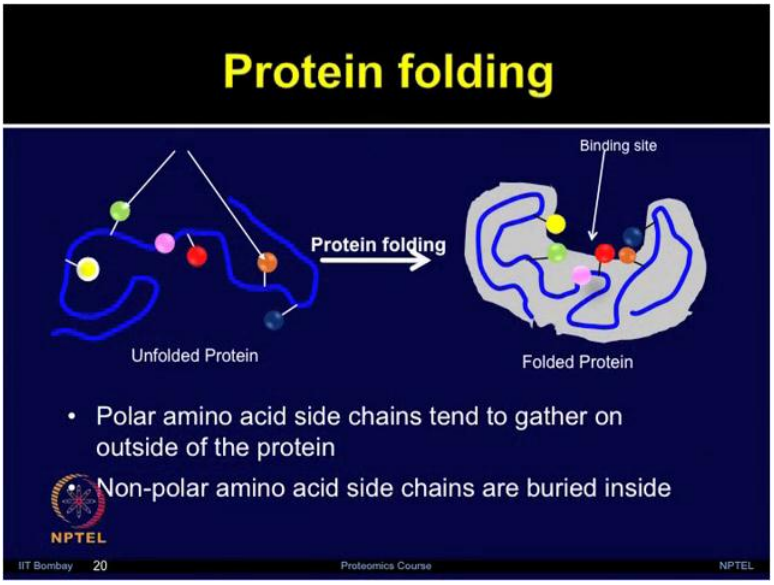
(Refer Slide Time: 15:35)




In protein folding the amino acid sequence determine the three-dimensional structure. Now, as you can see here the proteins having very much specificity if you have an amino acid sequence 1 shown in blue color that will form protein one, shown in the right side. If you have amino acid sequence two shown in red that will form protein two. Now if you take the amino acid sequence one protein two cannot be generated similarly, if you take amino acid sequence two protein one cannot be generated. So, there is very high

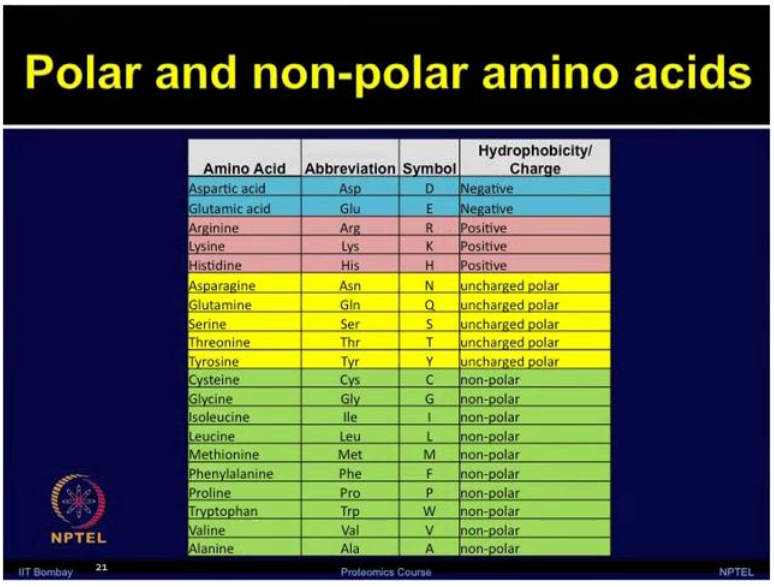
specificity of amino acid sequence which can determine the three-dimensional structures of proteins. The protein folding process is governed by distribution of polar and non-polar amino acids. If you remember last class, we have talked about various amino acids.

(Refer Slide Time: 16:49)




The polar side chains they tend to arrange themselves near outside of the molecules you take for example, argentine, glutamine, histerine similarly, on the non-polar side chains they have tendency to cluster in the interior of molecules for example, phenyl, elerine, leucine, valine and tryptophan.

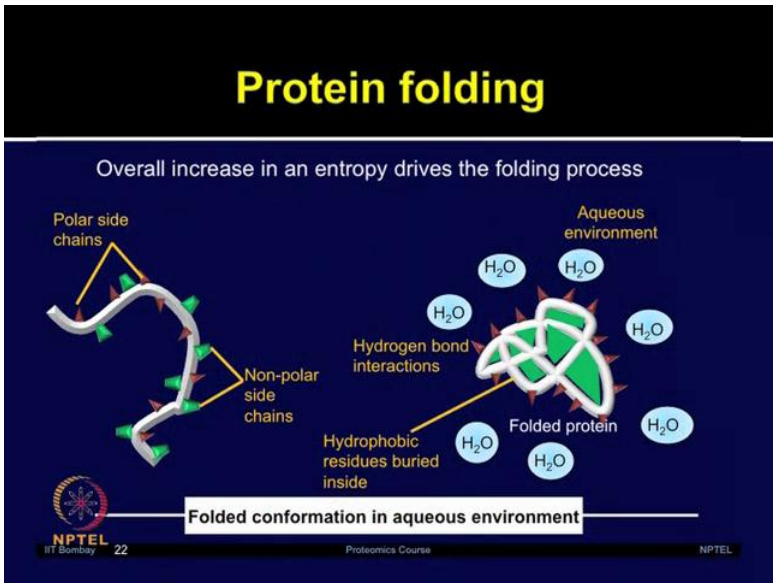
(Refer Slide Time: 17:20)



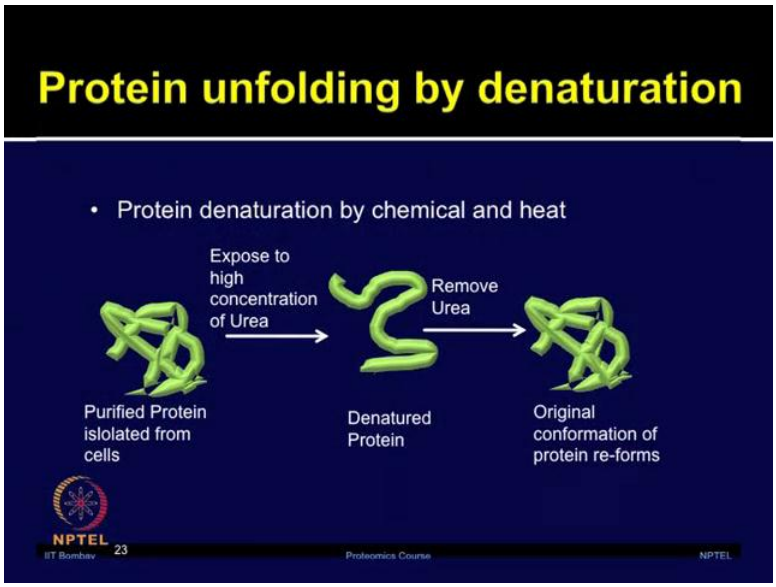

| Amino Acid | Abbreviation | Symbol | Hydrophobicity/ Charge |
|---|---|---|---|
| Aspartic acid | Asp | D | Negative |
| Glutamic acid | Glu | E | Negative |
| Arginine | Arg | R | Positive |
| Lysine | Lys | K | Positive |
| Histidine | His | H | Positive |
| Asparagine | Asn | N | uncharged polar |
| 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 their 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.

(Refer Slide Time: 17:50)



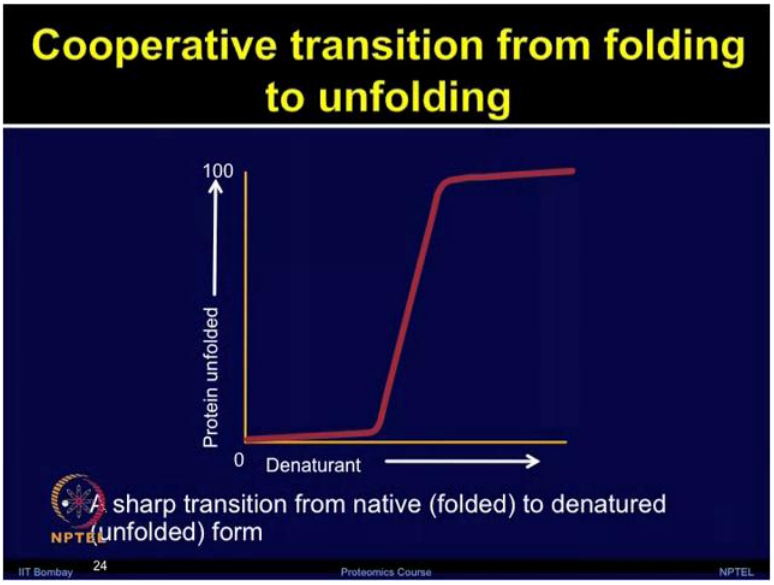

(Refer Slide Time: 18:09)




So, continuing on protein folding the hydrophobic amino acids they are driven to associate the hydrophobic collapse. So, when these amino acids come together, as you can see on the right hand side the loss of water surrounding these amino acid increases entropy of the system. Therefore, overall increase in entropy drives the folding process.

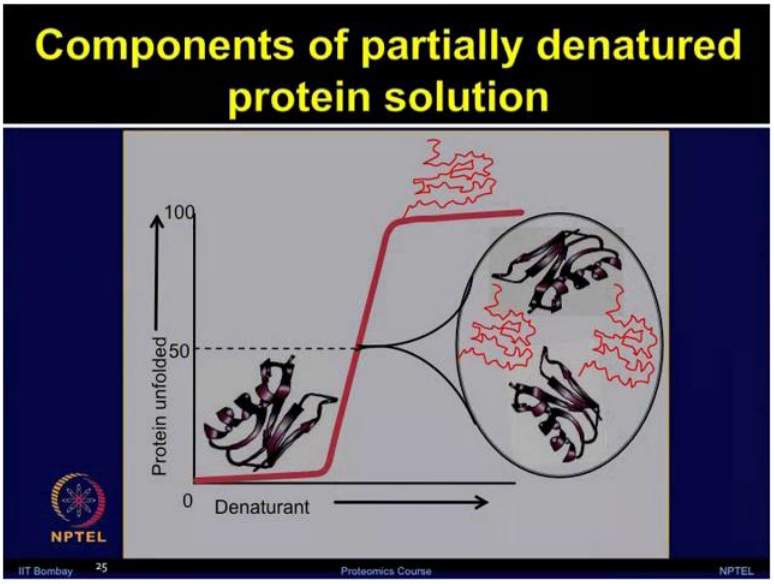
Now as we have seen in the classical experiment of Anfinsen the protein unfolding can be done by using denaturants. So, if you take denaturants whether it is chemical like urea and gouradanium chloride or you heat treated. So, as you can see here if you have a purified protein isolate taken from the cells and you expose to the high concentration denaturants whether it is chemical or heat that will result into the denatured protein shown in the center. If you remove the denaturing condition it will again form the proper folding protein conformation will be restored in it is original form.

(Refer Slide Time: 19:03)



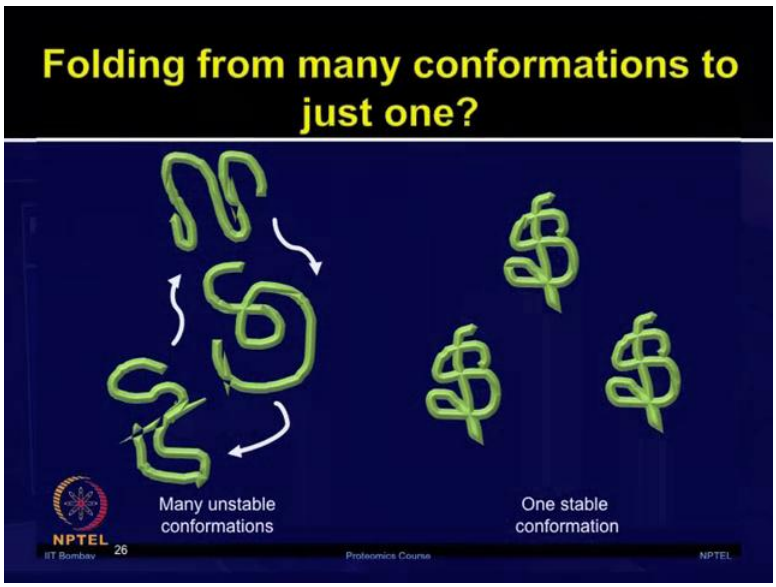

So, how that process of folding to unfolding holds the various hypothesis and mechanism have been proposed. Let us talk about cooperative transition here, from folding to the unfolding form. As you can see in this graph, on the y axis the protein in the unfolded form from 0 to 100 and on the x axis the presence of denaturant a sharp transition from the native or the folded to the denatured or unfolded forms occur. So, only two 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: 20:03)



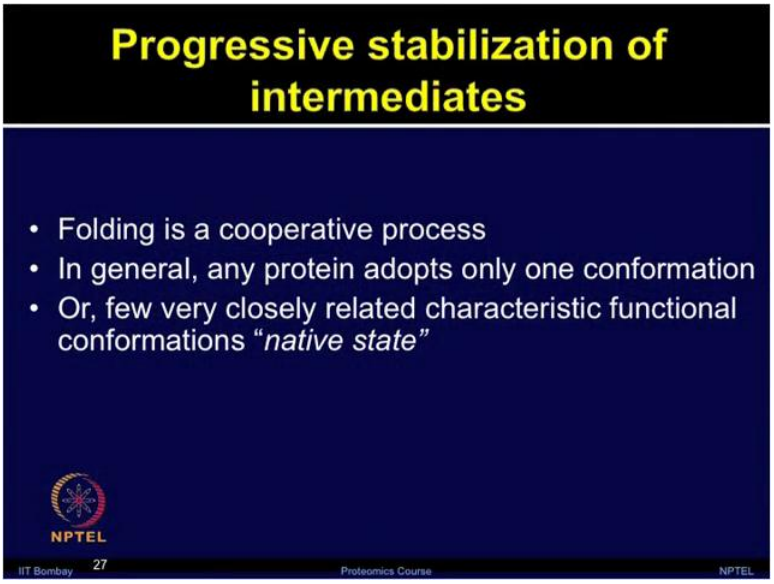

So, what are the components of partially denatured protein? If you look at this graph in the transition state at 50 percent, it will be 50 percent fully folded and 50 percent unfolded form of the protein; however, existence of only two states the folded and unfolded or possibility of unstable transient 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 one form? The particular sequences along polypeptide backbone they impose key restrictions.

(Refer Slide Time: 21:05)



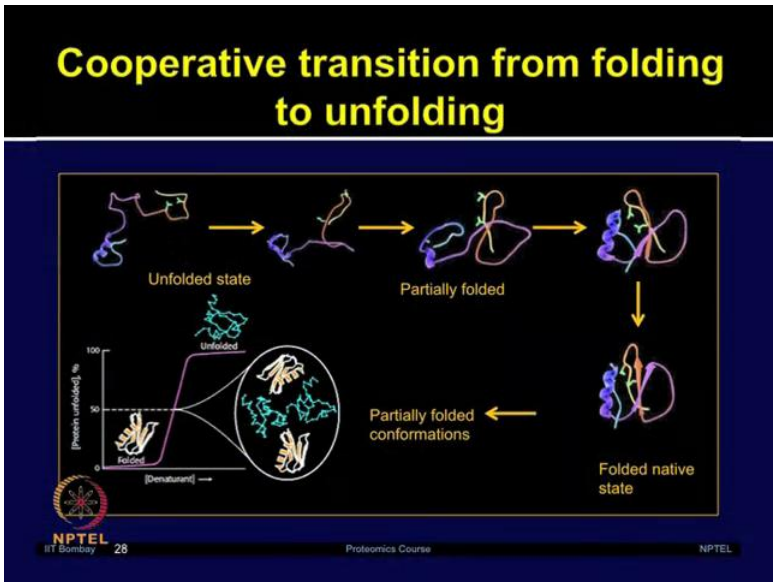
The various properties of the side chains which we have talked in the previous lecture, including size hydrophobicity ability to form hydrogen and on the bond all of these governs this process. Let us take example, of arginine a side chain with positive charge might attract a segment of the polypeptide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

(Refer Slide Time: 21:58)



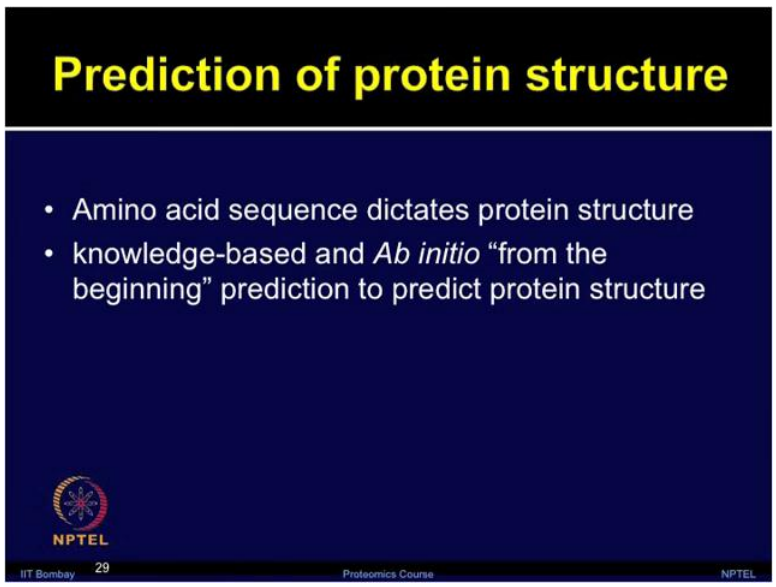

So, there are various progressive stabilization of intermediates occur in the folding process. As we talked folding is a cooperative process, which involves progressive stabilization of various intermediates. In general any protein adopts only one conformation which we just talked in the last slide or few very closely related characteristic functional conformations may occur, which will give rise to the native state. Native state in structure context here, will be the conformation which has the lowest free entropy or the stable folded form for majority of the proteins.

(Refer Slide Time: 22:54)




So, let us see in the slide cooperative transition occurs form the folding to unfolding. Folding in the 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: 23:28)




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 predication

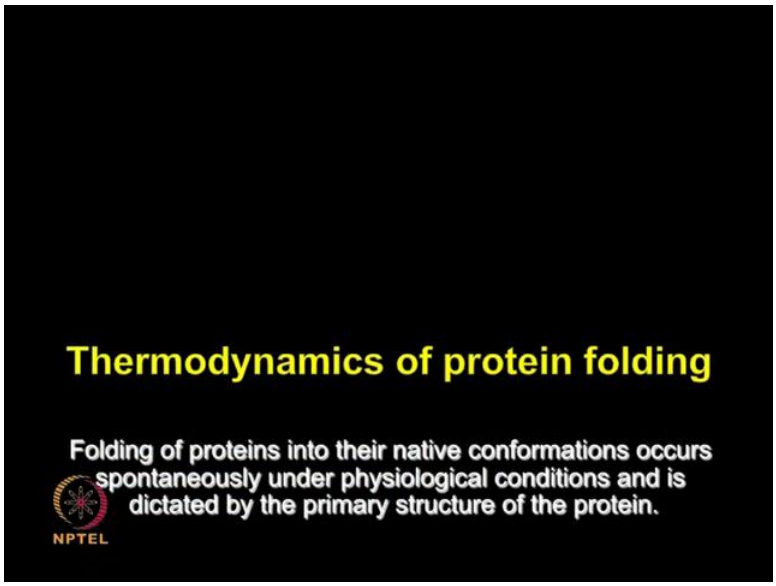
final folded structure is possible from it is sequence; however, there are long range of interactions and vast number of possible conformations which are possible. And therefore, it limits these type of predications; however, knowledge-based and a b initio from the beginning prediction do take place to predict the protein structure. So, let me show you the protein folding process in following animation.

The process of protein folding is dictated by the distribution of polar and non-polar amino acid residues in the protein. The hydrophobic amino acids are driven to interact with one another by a process known as hydrophobic collapse. They come together and during that process eliminate water molecule surrounding them, the polar residues remain on the surface and form hydrogen bonds with water molecules, while the hydrophobic the hydrophobic residual get buried within the core of protein.

Protein folding is a cooperative process whereas; the unfolding is a sharp and quick transition. 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, and shape etcetera. 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. Several long range interactions can often limit such predictions.

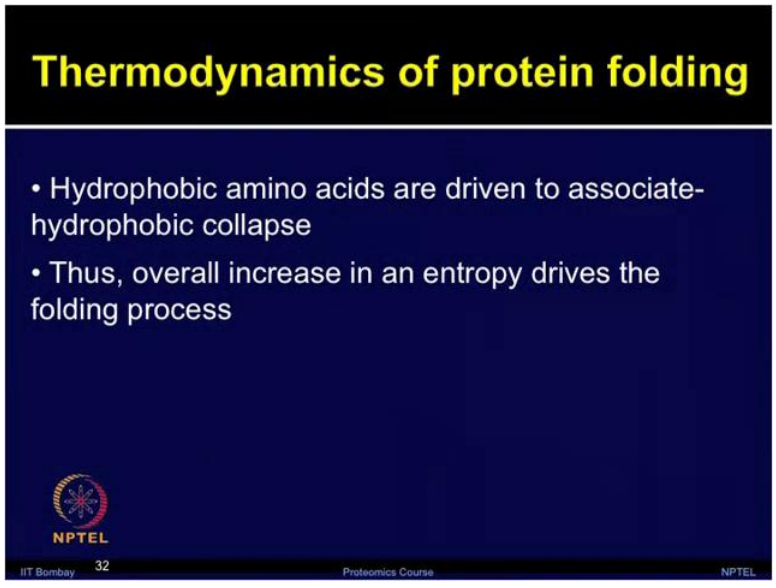
Here on x axis denaturants are plotted and y axis the percentage protein unfolded is plotted. On 0 you can see that is totally folded form of the protein on 100 percent it is unfolded form, but if you take a mixture from 50 percent that is either unfolded or folded form. This shows that protein can assume either, folded form or unfolded form of the proteins. Let us now talk about thermodynamics of protein folding.

(Refer Slide Time: 26:54)



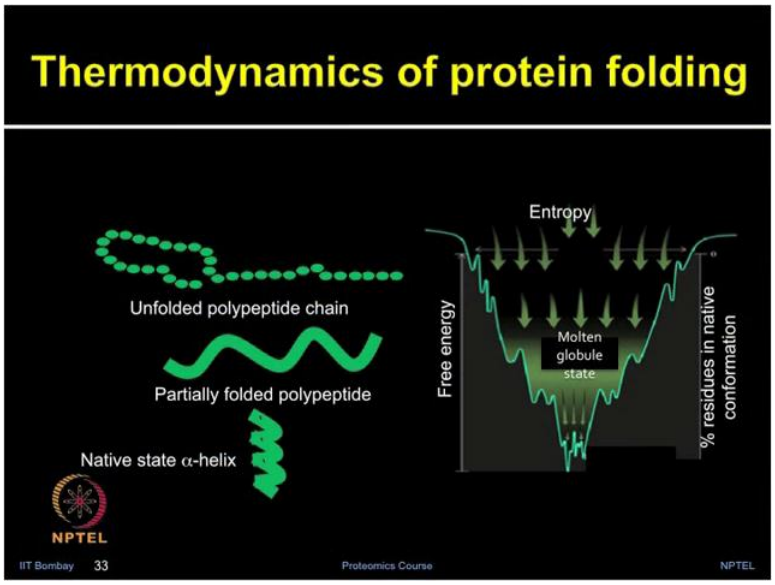

The folding proteins into their native conformation occur spontaneously under physiological conditions, and are dictated by the primary structure of the protein. Protein folding is thermodynamically favorable process where decrease free energy from unfolded to folded state occurs.

(Refer Slide Time: 27:26)




Let us 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 entropy drives the folding process.

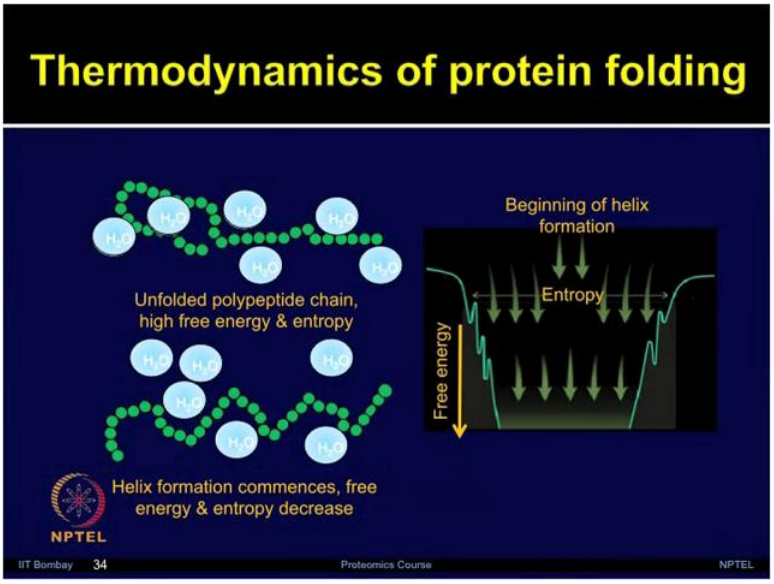
(Refer Slide Time: 27:49)




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.

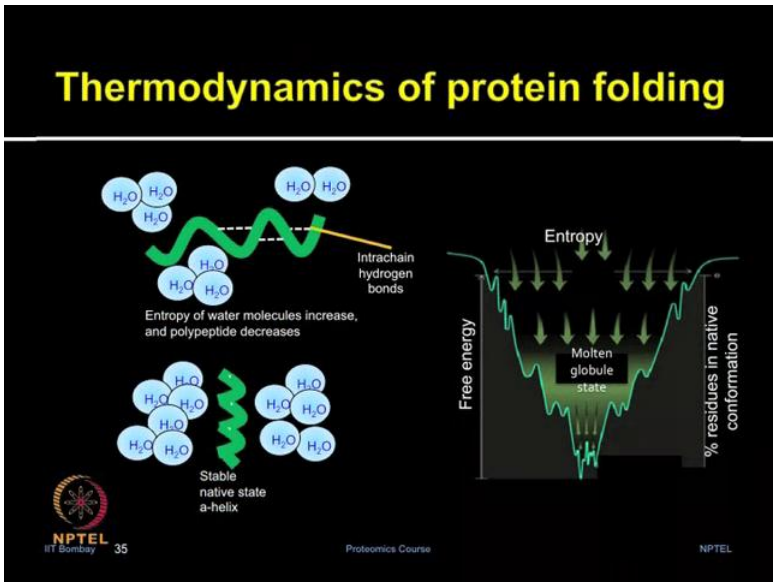
The initial collapse state of protein with very little thermodynamic stability is known as molten globule. The amino acid side chains are extremely disordered in this state and several fluctuations can be observed. As you can see from these arrows, as free energy of protein molecules decreases, the protein 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 funnel well defined and folded conformation states are present.

(Refer Slide Time: 29:03)



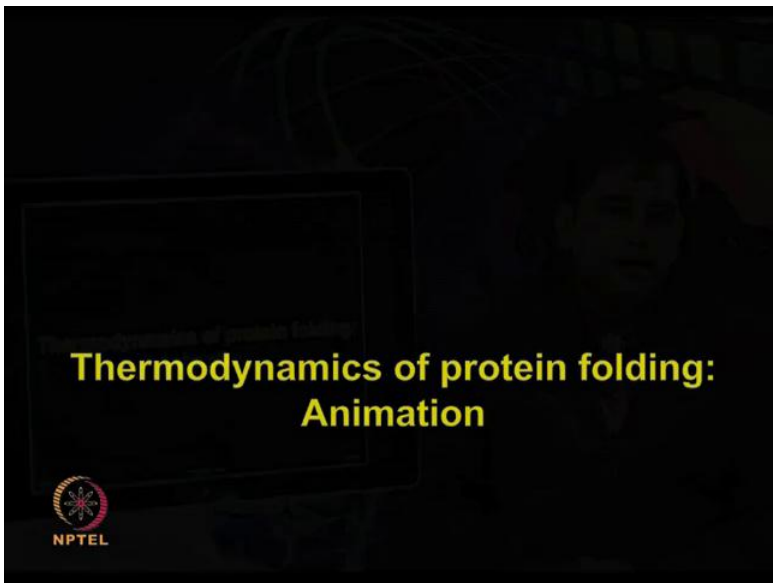

So, if you look at unfolded polypeptide chain. So, the amino acids that have been joined together by the peptide bonds, but they have not yet formed their secondary or tertiary structure. So, this conformation has highest free energy and entropy. The amino acids in the polypeptide chain start interacting 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 thermodynamic terms the lowering of entropy is favored by a corresponding increase in entropy in the surroundings composed of the water molecules.

(Refer Slide Time: 30:14)




Now if you look at the funnel again, as the protein continues to fold in order to assume it is stable low energy native state conformations, the entropy also decreases while it may appear unfavorable for the system; however, entropy of the surrounding water molecule increases in this process, and it increases overall entropy and makes it favorable and spontaneous.

(Refer Slide Time: 30:55)



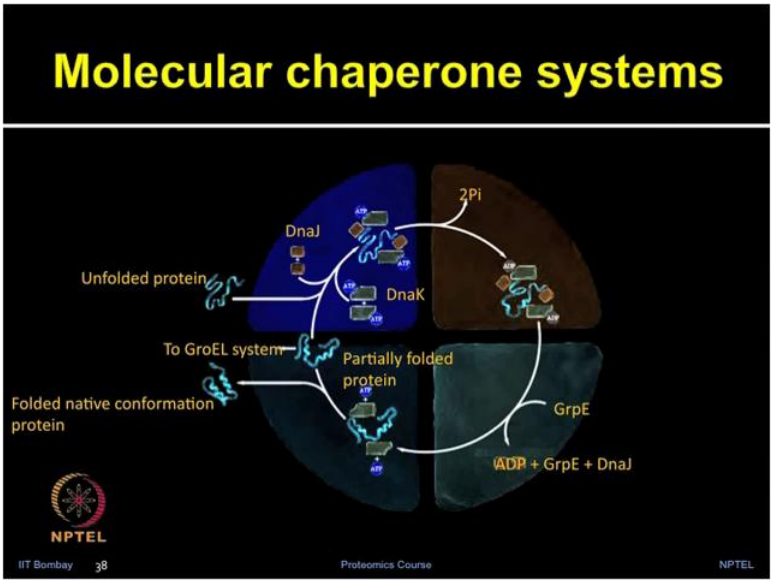

So, let me show you how protein folding works, and how it can be described in the thermodynamics terms in following animations. An unfolded polypeptide chain has very

high energy and entropy. The protein folding acts to decrease the free energy of the system, by forming favorable interactions and assuming a more stable state. The entropy of the polypeptide chain decreases during this process, as the protein continues to fold in order to assume stable low energy native state conformation the entropy also decreases, while they should seem unfavorable 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 favorable and spontaneous.

(Refer Slide Time: 32:11)


Let us now talk about molecular chaperones for protein folding. The molecular chaperones are 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.

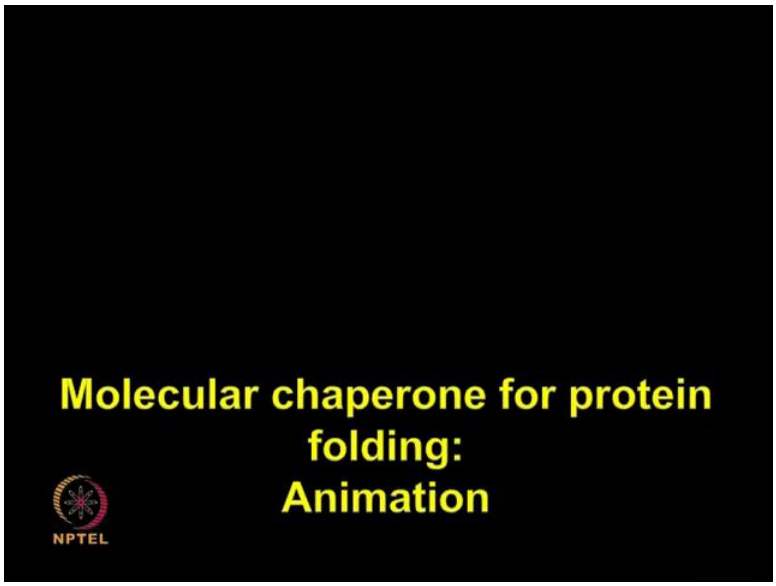
(Refer Slide Time: 32:44)




So, let us talk about various molecular chaperone systems. Many newly synthesized 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 this is produced significant amount of unfolded proteins by synthesizing new systems which are known as molecular chaperones, which are designed to promote the protein folding process. The several molecular chaperone systems which have been example, you find in e coli includes gro e l system d n a k, d n a j, g r p e and c l p b

(Refer Slide Time: 33:54)



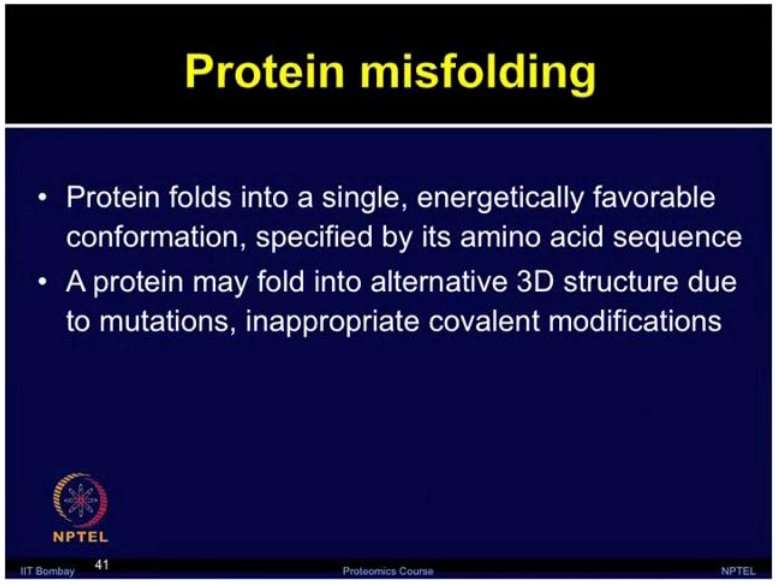

The bacterial chaperon in gro e l it binds proteins in non-native states, and allows enzymes to be quantitatively in native form by binding which requires cochaperanin gro e l and a t p. Let me show you how this chaperon in works and govern the protein folding process in following animations. The unfolded protein is bound by d n j and then by d n a k which is an a t p bound protein. The hydrolysis of a t p into a d p and p I by d n a k is stimulated by d n a j.

This resulting d n a k a d p remains tightly bound to the unfolded protein. The nucleotide exchange factor g r p e present in bacteria facilitates release of a d p along with d n a j. This leaves the d n a k bound to the partially folded protein, which continues to undergo folding to a more favorable low energy conformation. Once a protein gets completely folded, it gets detached d n a k which then binds a t p again, and completes the cycle and prepares it for next round of protein folding. Any protein which may not have been folded completely is then taken over by the gro e l chaperon in systems which complete the folding.

(Refer Slide Time: 36:02)


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 in complete processing which results into altered folding properties.

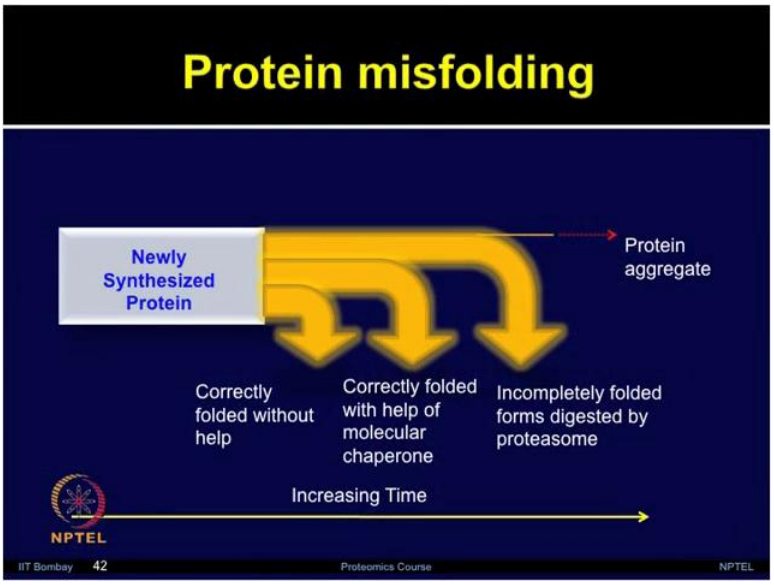
(Refer Slide Time: 36:39)




Proteins fold into a single energetically most favorable conformation which is specified by it is amino acid sequence. A protein may fold into alternative 3 dimensional

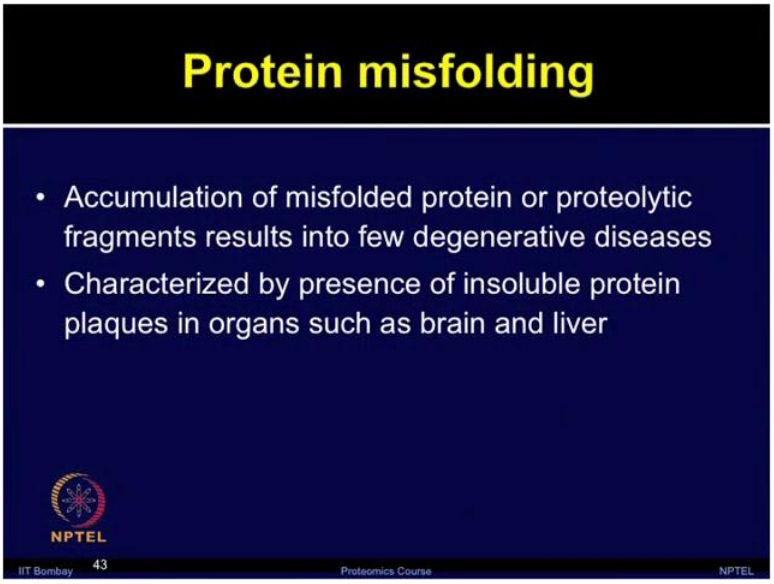
structures, because of mutations or inappropriate covalent modifications. Therefore, protein misfolding may lead to loss of normal protein function.

(Refer Slide Time: 37:19)



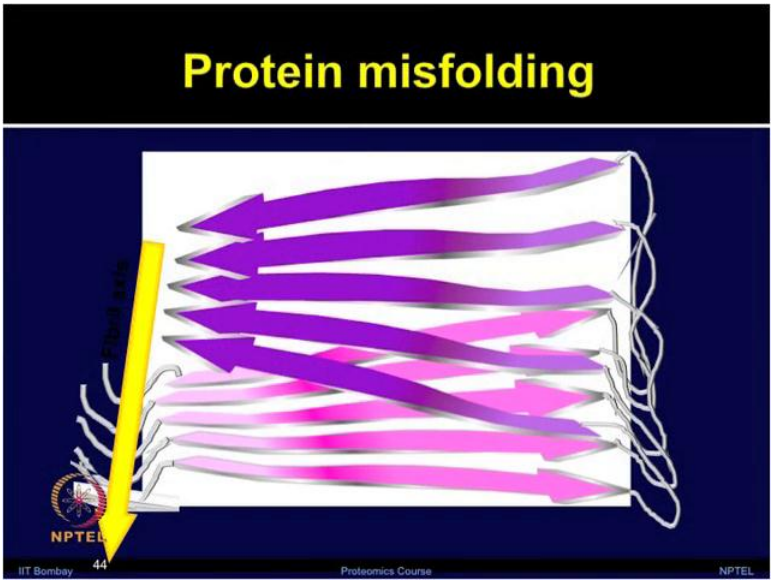

So, in this slide we can see from a newly synthesized protein it could have various phase it may form the proper folded form without any assistance or it can give correct folded forms in presence of molecular chaperones or incompletely folded forms can be digested by protiozome machinery or it may result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

(Refer Slide Time: 37:45)




So, accumulation of these misfolded proteins or proteolytic fragments can result into few degenerative diseases. These degenerative diseases are characterized by the presence of insoluble protein plates in organs, such as brain and liver for example, Alzheimer's disease in human as well as Parkinson disease. Bovine spongiform encephalopathy also known as mad cow disease in cows and a creepy disease in sheep.

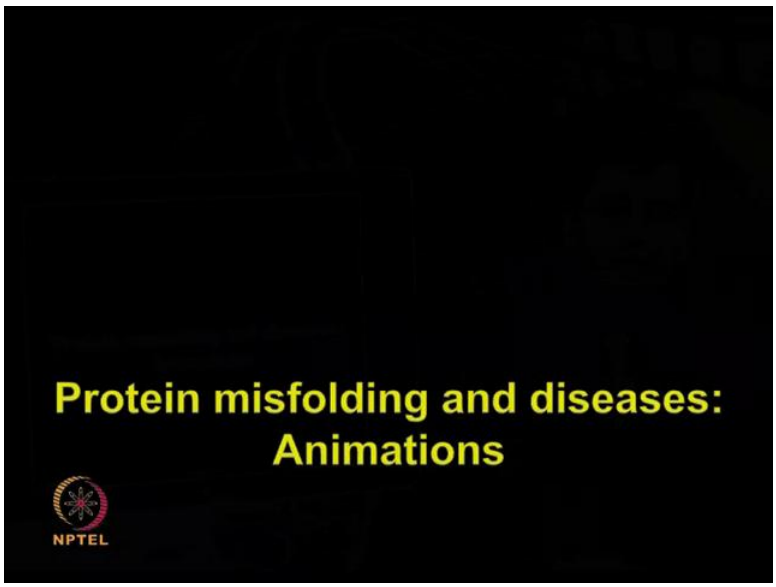
(Refer Slide Time: 38:40)




So, in this slide at least shown that amyloidal fibrils are involved in neurodegenerative diseases; and the protein aggregations due to the large beta sheets are reduced from solid

state n m r. In Alzheimer's disease the presence of beta amyloidal containing plaques is associated with neuro degeneration and dementia. In other neurodegenerative diseases it has also been shown that it involves protein aggregation preon diseases such as jakob disease and b a c or bovine spongiform encephalopathy are associated with amyloidal deposit of p r p proteins.

(Refer Slide Time: 39:37)




So, how insoluble protein aggregates can result into various diseases. Let us discuss some of protein misfolding related disease in following animation. Protein misfolding results into various diseases, such as Alzheimer's disease. In Alzheimer's disease structure of certain normal soluble silver 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 amyloidal plaques are essentially made up of single polypeptide chain or fibrils known as amyloidal b protein it is observed in the brain of patients with Alzheimer's where dead or drying neurons are surrounded with plaques.

The neuroplasticity is believed to be caused by the amyloidal fibrils before they get deposited as amyloidal 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

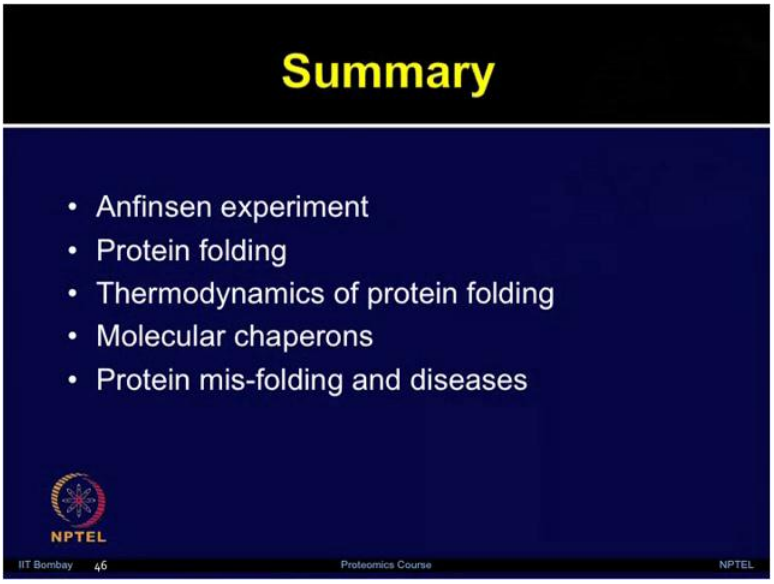
composed of beta sheet structures that have been modified from previously existing alpha helices the protein aggregates of one abnormal protein it is sufficient to function as nuclei for other normal proteins to attach.

Huntington's disease it is a neurodegenerative disorder of genetic origin which affects muscular coordination. It is caused by increased number of trinucleotide repeats c a g 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 help 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 c f t r. The c f t r gene regulates components of sweat digestive juices and mucous, it is caused by a deletion three nucleotides leading to the elimination of a 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 one entire trypsin or a one a t. The a one a t gene is responsible for protection of lung tissue from damage by enzyme neutrophil elastase abnormally secreted a one a t gets accumulated in liver thereby allows lung tissue damage results into wheezing shortness of breath and asthma like symptoms.

Lathyrism it is regular ingestion of seeds from sweet pea lathyrus odoratus which causes description of cross linking a muscle protein collision collision is very important structure protein which has triple helical structure. The cross links formed are due to the oxidation of lysine residues by the enzyme lysil oxiadase to form a lysine these are essential for proper folding of collision 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 fragility and weakness.

(Refer Slide Time: 46:17)




So, in summary today we talked about a classical experiment of Anfinsen which has proved that all the information which is crucial for protein folding resides within the primary amino acid sequence. We then 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. I will beat 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: 47:30)