LECTURE-05

PROTEINS: FOLDING AND MISFOLDING

TRANSCRIPT

Welcome to the proteomics course, today we will talk about proteins, its folding and misfolding. Understanding the processes of protein folding and misfolding has been a major research area from several decades in biology chemistry and physics. Understating 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 and how does this process works? And how misfolding may result into various manifestation of diseases.

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 that globular protein 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 forces and interactions is such that the folded proteins the magnitude of favorable interactions or reactions will be outweighing the sum of unfavorable interactions and the result, it governs the protein folding.

In today's lecture we will talk about how amino acid sequence determines the three-dimensional protein structure? I will explain you a classical experiment done by Anfinsen. We will talk about how protein folding occurs? Some of the basic thermodynamics of protein folding concepts and how molecular chaperone govern the protein folding process, and how protein misfolding may result into various diseases?

So let's start out first topic amino acid sequences determines 3-D structure of proteins. There is very intricate sequence-structure relationship. The Amino acid sequence dictates conformations, 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 state by using protein ribonuclease 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. The fundamental discovery of Anfinsen led him to win Nobel prize in chemistry in 1972. So let me explain you how this experiment worked?

To establish the proof for relationship between amino acid sequence and its conformation, Anfinsen in 1972, performed an experiment where he performed two steps denaturation and refolding. So how denaturation and refolding works?

In this classical experiment Anfinsen used protein ribonuclease A. He used few denaturants such as Urea or Guanidine HCL and β-mercaptoethanol which breaks the disulphide bonds. let's look at each one of these components in little more detail. First talk about ribonuclease A protein. This protein has contributed greatly to our understanding of protein folding *in vitro* from the landmark experiment of Anfinsen. As you can see in the structure ribonuclease has 124 amino acid residues and it forms four disulphide bridges, which are located between cysteine residues of 26 and 84, 40 and 95, 58 and 110 and 65 and 72. This protein catalyzes the hydrolysis of RNA.

What is the role of urea and guanidinium chloride? Urea is an organic molecule, which contains two amino groups joined by a carbonyl group and used at a concentration of around 6M for denaturating the protein by breaking the non-covalent interaction. Both urea and guanidinium chloride can effectively disrupt the protein's non-covalent interactions.

What is the role of β-mercaptoethanol? The β-mercaptoethanol is commonly used to reduce disulphide linkages of proteins and thereby it disrupts tertiary and quaternary structure. As you can see here in the slide, in the excess of β-mercaptoethanol the disulphide or cystines can be fully converted into sulphahydral or cyteines.

In Anfinsen's experiment, he used 8M urea and β-mercaptoethanol treatment which converted native protein to fully reduced, randomly coiled polypeptide known as the "denatured" structure. The denatured polypeptide lacked enzymatic activity. So as we have discussed the ribonuclease protein that it contains 124 amino acid residues and forms four disulphide linkages. These linkages are formed between the cysteines (I have shown here) 26 and 84, 40 and 95, 58 and 110 and 65 and 72. The ribonuclease native conformation is lost when it was treated with 8 M urea and β-mercaptoethanol. As you can see here, the native ribonuclease has formed the denatured reduced ribonuclease due to the breaking of disulphide and non-covalent interactions.

On treatment of urea and β-mercaptoethanol ribonuclease A protein lost its native conformation because of breaking of disulphide and non-covalent linkages. Anfinsen noticed that when ribonuclease was oxidised in air and urea was removed by the process of dialysis. The enzyme activity slowly recovered as a result of the protein

folding. As you can see here if you have used 6 M urea and β-mercaptoethanol, all the disulphide and non-covalent interactions are breaking, once urea is removed then slowly protein refolding occurs it results into reformation of tertiary structure and active site. When Anfinsen repeated this experiment in the presence of denaturant urea that led to regeneration of less than 1% of enzyme activity. So what could be the reason? In fact urea prevented the correct disulphide pairing which resulted into the scrambled form. Scrambled ribonuclease, now if you mathematically calculate due to the presence of four disulphide bonds here and presence of 8 cysteine residues, it can actually give rise to 105 different forms in which these four disulphide bonds can be formed. In the absence of urea correct disulphide bond formation occurred and it allowed folded and thermodynamically stable state to be reached in ribonuclease protein.

Now this figure shows here that in the presence of trace amount of β-mercaptoethanol and complete removal of denaturant urea the refolding of ribonulease was accurate and four intra-chain disulphide bonds were reformed in the same position where they were expected in native ribonuclease. 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 bond and achieve the native conformation.

In Anfinsen's experiment, he removed urea and β-mercaptoethanol by dialysis process. The denatured ribonuclease regained its enzyme activity, the enzyme was refolded into active form and sulfhydral groups became oxidized by air. The experiment proved that the information required for specific catalytic activity of ribonuclease is contained in its amino acid sequence.

The classical study of Anfinsen proved that all the information required for protein folding resides in its primary sequence. Let me explain you this experiment in following animation. In Anfinsen's experiment ribonuclease A is its native state has four disulphide bonds between its cysteine residues. When treat this protein with β-mercaptoethanol and 6 M urea the protein undergoes denaturation and the disulphide linkages are broken. The enzymes activity is lost in its denaturated state. It was observed by Anfinsen that removal of urea and β-mercaptoethanol led to refolding of enzyme to its native state with more than 90% enzyme activity being intact. However if only β-mercaptoethanol was removed in presence of urea, the formation of disulphide bonds was random which led to enzyme with only 1% activity.

After discussing classical experiment of Anfinsen, let's 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.

In protein folding the amino acid sequence determines the 3-D structure. Now as you can see here 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 2 shown in red that will form protein 2. Now if you take amino acid sequence 1, protein 2 cannot be generated, similarly if you take sequence 2 protein 1 cannot be generated. So there is very high specificity of amino acid sequence, which can determine 3-D 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 on outside of the protein. You take for example arginine, glutamine and histidine. Similarly, on the non-polar side chain, they have tendency to cluster in the interior of molecules. For example- phenylalanine, leucine, valine and tryptophan.

This chart is only for your information which shows there are 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.

Continuing on protein folding, the hydrophobic amino acids they are driven to associate 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 acids increases entropy of the system.

Now as you have seen in the classical experiment of Anfinsen, protein unfolding can be done by using denaturants. If you take denaturants like urea and guanidium chloride or you heat treat it, so as you can see here if you have a purified protein isolate from cells and you expose it to high concentration of denaturants, whether it's chemical or heat that will result into 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.

How this process of folding to unfolding works? Various hypothesis and mechanisms have been proposed; let's talk about cooperative transition from folding to 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 native (folded) to denatured (unfolded) forms occur. So only two conformational states are

present significantly, whether it's folded or unfolded form. If denaturants are removed from the unfolded protein, it allows protein to make folded forms.

What are the components of partially denatured/unfolded protein? If you look at this graph in the transition state at 50%, it will 50% fully folded and 50% unfolded form of proteins. However, existence of only two state the folded and unfolded or possibility of unstable transient intermediates between folded and unfolded states still remains a topic of research in protein folding area.

How does folding occur from many conformations to just one? The particular sequences along the polypeptide backbone, they impose key restrictions. 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. For example, if you take tryptophan it has large side chains which sterically block the one region of chain from partially closely against another region. So this type of hindrance can be governing factor. Similarly let's take an example of Arginine, a side chain with positive charge might attract a segment of the polypeptide, which has complementary negative charge. So these type of side chains and various type backbone properties are going to impose key restrictions, therefore various type of folded conformations will be selected and it can result in only one which is going to govern the folding process.

There is various progressive stabilization of intermediates occurring in the folding process. As we talked, folding is a cooperative progress, which involves progressive stabilization of 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 lowest free entropy or the most stable folded form for majority of the proteins.

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

How to do structural prediction of proteins? As we have seen in the previous experiment the amino acid sequence dictates protein structure. So theoretically prediction of final folded structure is possible from its sequence. However, there are long range of interactions and vast number of conformations which are possible and therefore it limits

these type of predictions. However knowledge based and *Ab initio* "from the beginning" prediction do take place to predict the protein structure.

Let me show you protein folding process in following animation. The process of protein folding is governed by distribution of polar and non-polar amino acid residues in proteins. The hydrophobic amino acids are driven to interact with one another by a process known as hydrophobic collapse. 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 amino acid residues get buried in the protein core. Protein folding is a cooperative process, whereas unfolding is a sharp and quick transition. Proteins typically adopt only one characteristic functional native state conformation, which has lowest free energy and it's most stable. Folding is limited to one conformation due to properties of amino acid side chains such as hydrophobicity, size, shape etc. Folding is highly cooperative process, wherein there is progressive stabilization of intermediates. 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 on Y-axis the percentage protein unfolded is plotted. On 0, you can see there is totally folded form of the protein, on 100, it is unfolded form. But if you take a mixture from 50% that is either folded or unfolded form, which shows protein can assume either folded or unfolded form.

Let's now talk about thermodynamics of protein folding. Folding of proteins into their native conformation occurs spontaneously under physiological conditions and is dictated by the primary structure of the protein. Protein folding is thermodynamically favorable process where decrease in free energy from unfolded to folded state occurs.

Let's talk about some of the basics of thermodynamics for protein folding. As we have seen earlier, the hydrophobic amino acids are driven to associate hydrophobic collapse. When these amino acids come together, the loss of water surrounding these amino acids increases the free energy of the system. Therefore, overall increase in entropy drives the folding process.

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 the 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 narrower part of the funnel, look at the bottom of the funnel and only few conformations can be accessible here. So at the bottom of the funnel well defined and folded conformation states are present.

If you look at the 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 chains 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. In the thermodynamic terms the lowering of entropy is favored by a corresponding increase in entropy in the surrounding composed of the water molecules.

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

Let me show you how protein folding works and how it can be described in thermodynamic terms in following animation. An unfolded polypeptide chain has very high entropy and energy. The protein folding acts to decrease free energy of the system by forming favorable interaction and assuming 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 this should seem unfavorable for the system, it must be recalled that the entropy of the surrounding water molecules increases during the process. Thereby increasing overall entropy of the system and making it favorable and spontaneous.

Let's now talk about molecular chaperones for protein folding. The molecular chaperones are the class of heat inducible proteins, which provide 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.

Now let's 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 transiently exposed hydrophobic surfaces. In response to the heat shock the cells produce significant amount of unfolded proteins by synthesizing new systems, which are known as molecular chaperone which are designed to promote the protein folding process.

There are several molecular chaperone systems which have been exemplified in *E. coli* include GroEL, GroES, DnaK, DnaJ and GrpE. The bacterial chaperonin GroEL it binds proteins in non-native state and allows enzyme to be recovered quantitatively in native form by binding which requires co-chaperonin GroES and ATP. Let me show you how this chaperonin works and govern the protein folding process in following aninamtion.

Molecular chaperone for protein folding. The unfolded protein is bound by DnaJ and then byDnaK which is an ATP bound protein. The hydrolysis of ATP into ADP and Pi by DnaK is stimulated by 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 DnaK bound to partially folded protein which continues to undergo folding to a more favourable low energy conformation. Once a protein gets completely folded it gets detached from DnaK which then binds ATP again and completes cycle and prepare for next round of protein folding. Any protein may not have been folded completely is then taken over GroEL chaperonin system which completes the cycle.

After talking about how protein works let's talk 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 result into altered folding properties.

Proteins folds into single energetically favorable conformation, which is specified by its amino acid sequence. A protein may fold into alternative 3-D structure because of mutations or inappropriate covalent modifications. Therefore, protein misfolding may lead to loss of normal protein function.

In this slide we can see, from a newly synthesized protein it could have various fate, 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 proteasome machinery or it may result into the protein aggregation. So a

newly synthesized protein may give rise to anyone of these forms depending on various factors, which are going to govern protein folding process.

Accumulation of these misfolded or proteolytic fragments can results into few degenerative diseases. These degenerative diseases are characterized by 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, Scrapei disease in sheep.

In this slide it is shown that amyloid fibrils are involved in neurodenerative diseases and the protein aggregation due to large β sheets are deduce form solid state NMR. In Alzheimer disease the presence of β amyloid containing plaque is associated with neurodegeneration and dementia. In other neurodegenerative diseases it has also been shown that it involves protein aggregation. Prion diseases such as Jacobs disease, BSE are associated with amyloid deposit of Prp protein.

How insoluble protein aggregates can result into various diseases? Let's discuss some of protein misfolding related diseases in following animation.

Alzheimer's disease - In Alzheimer's disease, structure of certain normal soluble cellular proteins which are normally rich in α-helical regions are converted into β-strands which further link with each other to form β-sheet aggregates known as amyloids. The insoluble amyloid plaques are essentially made up of single polypeptide chains or fibrils known as amyloid-B protein. It is observed in the brain of patients with Alzheimer's, where dead or dying neurons are surrounded with plaques. The 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 neuro-muscular coordination, confusion and dementia.

Jacobs' 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 β-sheet structures that have been modified from previously existing α-helices. The protein aggregate of one abnormal protein is sufficient to function as nuclei for other normal protein to attach. It is characterized by muscular dystrophy, loss of muscle control and memory loss.

Huntington disease - It is a neurodegenerative disorder of genetic origin, which affects muscular coordination. It is caused by increased number of trinucleotide gene repeats (CAG) 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 trans-membrane conductance regulator or CFTR. The CFTR gene regulates components of sweat, digestive juices and mucus. It is caused by a deletion of 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 α1-antitrypsin or A1-AT. The A1-AT gene is responsible for protection of lung tissue from damage by enzyme neutrophil elastase. Abnormally secreted A1-AT gets accumulated in liver thereby allows lung tissue damage. The decease results in 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 crossings are formed due to the oxidation lysine residues by the enzyme lysyl oxidase to form allysine. These are essential for proper folding of collagen and giving it 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.

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 chains restrict and govern the process of protein folding. We then looked at the thermodynamics of protein folding, albeit very briefly, we talked entropy and how it governs protein folding.

The molecular chaperones- we talked about some classical examples in animation 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.