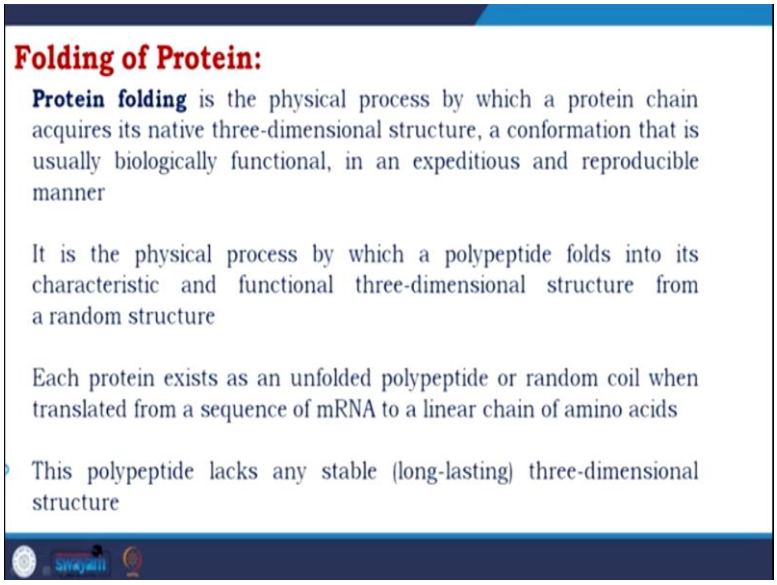
**Structural Biology**
**Prof. Saugata Hazra**
**Department of Biotechnology**
**Indian Institute of Technology – Roorkee**

**Lecture -10**
**Protein: Folding of Protein, Thermodynamics and Kinetics of protein folding, Characterization of Proteins**

Hi everyone, welcome again. So, today we are in the last class of the protein module. And here, we will discuss protein folding, Thermodynamics, and Kinetic parameters of protein folding.

**(Refer Slide Time: 00:53)**



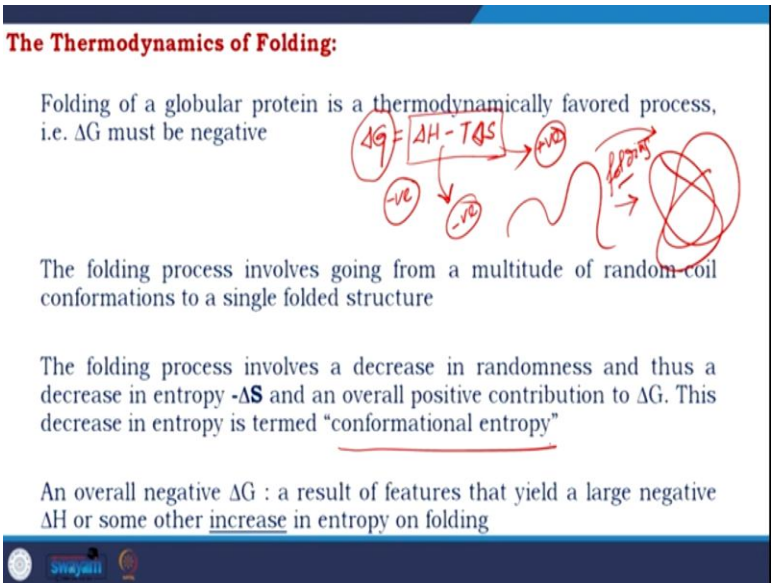

Protein folding is the physical process by which a protein chain acquires its native 3-dimensional structure, a conformation that is usually biologically functional in an expeditious and reproducible manner. Protein folding is a process by which protein sequences make trial and error to reach a form where the protein is compact and functional.

It is the physical process by which a polypeptide folds into a characteristic and functional 3-dimensional structure from a random structure. Each protein exists as an unfolded polypeptide or random coil when translated from an mRNA sequence to a linear chain of amino acids.

Nature allows that primary sequence to become a protein good enough to make those non-covalent interactions with the disulfide bond. So get a meaningful structure shape so that either

they could work as a structural protein like a fibrous protein or a functional protein like a globular protein.

**(Refer Slide Time: 05:15)**



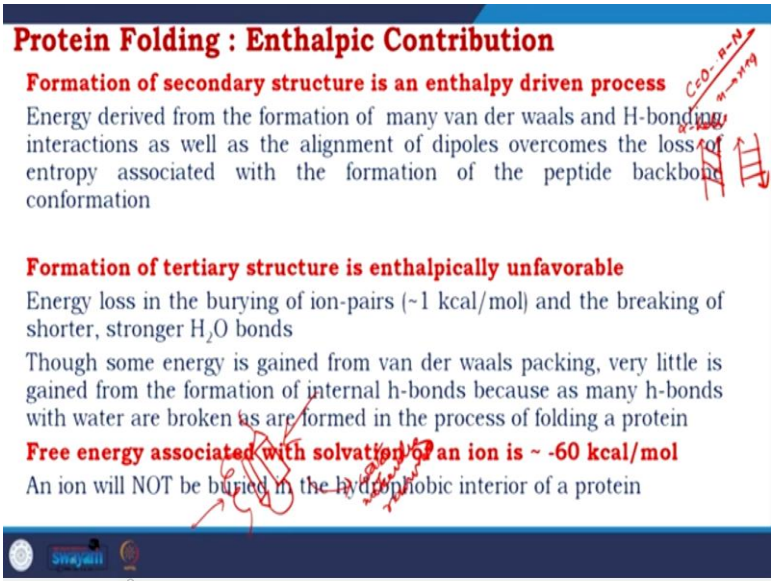

The Folding of a globular protein is a thermodynamically favored process, that is ΔG must be negative,

$\Delta G = \Delta H - T \Delta S$

where ΔG is the free energy change, H is the enthalpy change, and ΔS is the entropy change.

The folding process involves going from a multitude of random coil conformation to a single folded structure from a random coil to a fold. The folding process involves a decrease in randomness and, thus, a decrease in entropy and an overall positive contribution to ΔG. These decreases in entropy are called conformational entropy. The folding process involves a decrease in randomness. This decrease in entropy is called conformational entropy. And overall, negative ΔG results from features that yield a large negative ΔH (enthalpy change) or some other increase in entropy on folding.

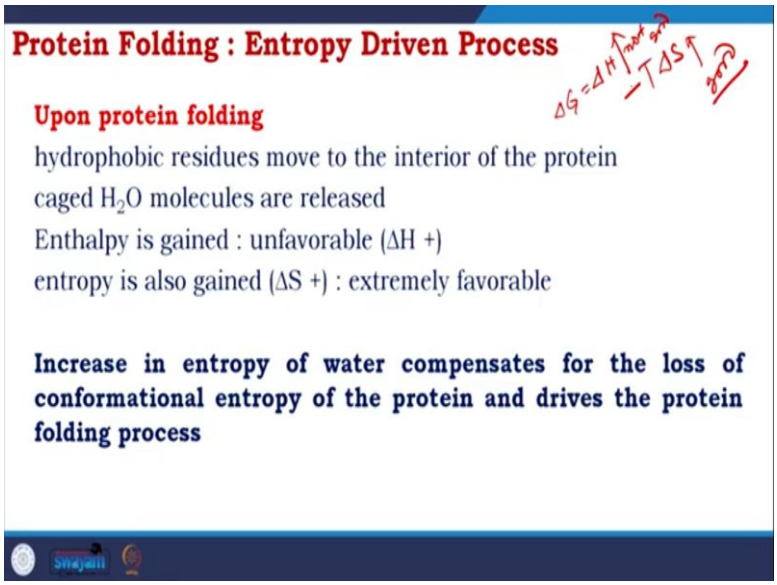
**(Refer Slide Time: 09:46)**




The formation of a secondary structure is an enthalpy-driven process. Why? If you remember, so many hydrogen bonds are developed mainly on the main chain. So, energy derived from the formation of many van der Waal and hydrogen bonding interactions and alignment of dipoles overcomes the loss of entropy associated with the formation of peptide backbone conformation.

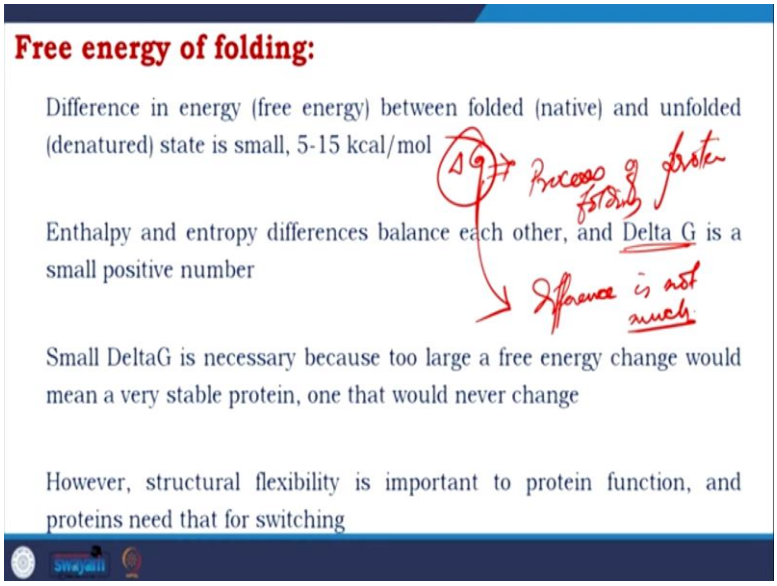
The alpha helix, and the dipole formation all favor the enthalpy. Formation of tertiary structure is enthalpically unfavorable energy loss in burying ion pairs and breaking up shorter, stronger $H_2O$ bonds. Some energies are gained from van der Waal and hydrophobic packing, and very little is gained from forming internal hydrogen bonds. Free energy associated with the solvation of an ion is -60 kcal/mol.

**(Refer Slide Time: 13:15)**



Protein folding is called an entropic-driven process. Upon protein folding, the hydrophobic residues move to the protein's interior. When caged water molecules are released, then enthalpies are gained. Entropy is also gained, which is extremely favorable. Increasing the entropy of water compensates for the loss of conformational entropy of the protein and drives the protein folding process.

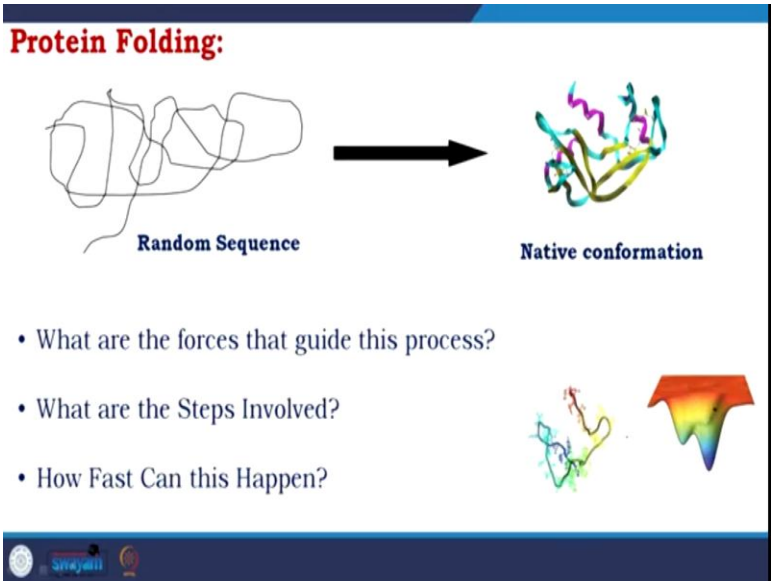
**(Refer Slide Time: 14:38)**




The difference in the free energy between the folded native and unfolded state is small, 5 to 15 kcal/mol. Enthalpy and entropy differences balanced each other, which we discussed, and ΔG is a small positive number. So, when we say that the delta G would be small, what that means? It is necessary because too large a free energy change would mean a very stable protein. If the protein is very stable, what will happen, is it would not be able to change its conformation frequently, and because of that, you know, as we discussed, when a protein has to be functional, it changes. Small non-coherent forces like hydrogen bond hydrophobic interactions are altered, and there is a clear scenario of change of protein conformation.

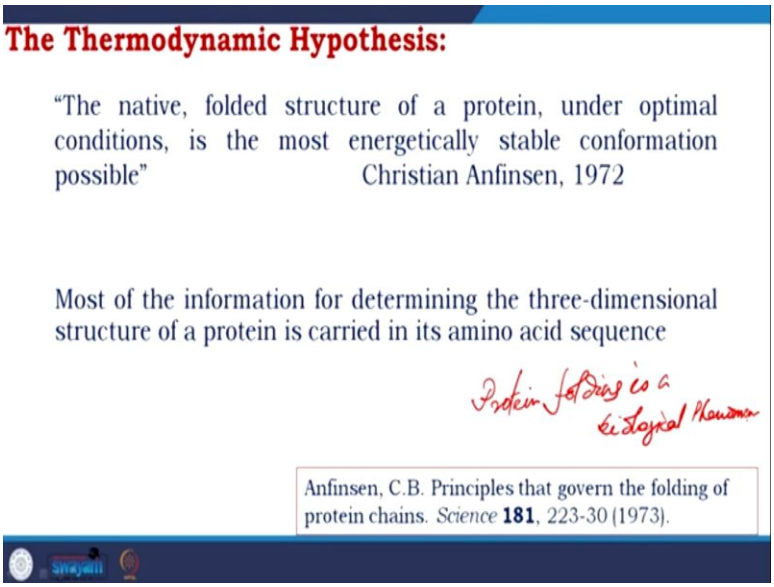
But, if one protein conformation is very stable, the alteration would not happen. Let us say a substrate is coming. The protein has to be opening up which is a change of conformation. If the protein is very stable, that conformation change will not happen. However, structural flexibility is important to protein function, and proteins need that for switching.

**(Refer Slide Time: 16:57)**




So, protein folding, as we discussed, from a random sequence to a native conformation, as you see the changes happening here, you see that the protein is attaining the minima. So, what forces guide this process, what are the steps involved, and how fast can this happen? These are questions for millions of years, especially lasting over 100 years.

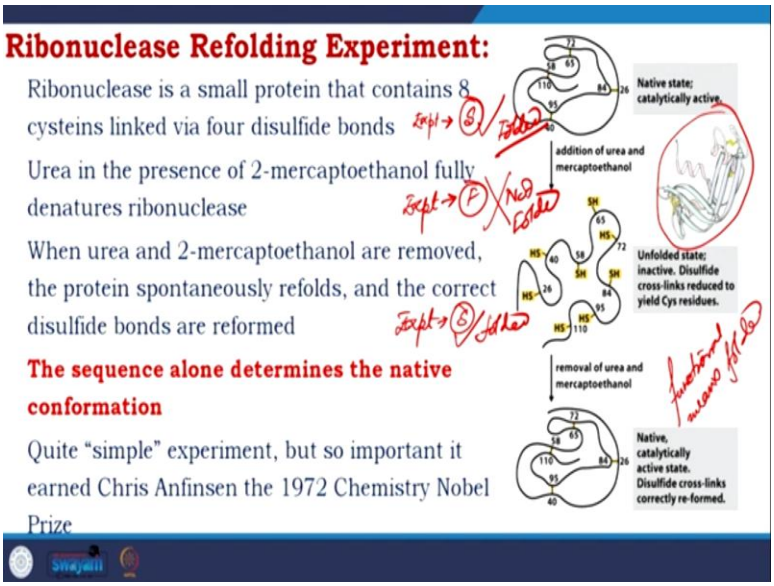
**(Refer Slide Time: 17:32)**




We are doing intense research about this. And one of the breaks I have to talk about is from Christian Anfinsen in 1972, who told that the native folded structure of a protein under optimal conditions is the most energetically stable conformation possible. And most of the information for determining the 3-dimensional structure of a protein is carried in its amino acid sequence.

Before this, the protein was thought to be a biological molecule, and the protein folding depends on the system.

**(Refer Slide Time: 18:59)**



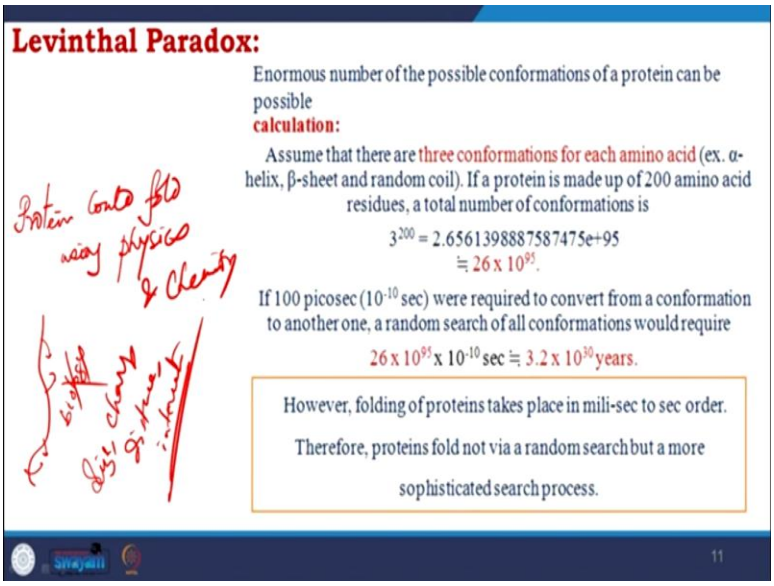

Now, he has performed some experiments called the Ribonuclease refolding experiment. So, ribonuclease is the protein with alpha helix and beta sheets. Now interestingly, the proteins have several cysteines, around 8 cysteines. So, Christian Anfinsen chose ribonuclease not because of these but because it is used to develop an assay where a colorless solution is converted to red color.

So, ribonuclease is a small protein containing 8 cysteines linked via 4 disulfide bonds. So, he purified the protein, did the assay, and showed that the protein could generate red color, which means the protein is folded.

Then he used urea in the presence of 2-mercaptoethanol, which fully denature the protein. So, from folded, he brings it to the unfolded state, proving that the protein is unfolded because it cannot convert the colorless solution to red. When urea and 2-mercaptoethanol are removed, the proteins spontaneously fold, and the correct disulfide bonds are reformed.

It is quite a simple experiment. But it was so crucial that he was awarded Nobel Prize.

**(Refer Slide Time: 24:51)**




Protein could fold using physics and chemistry. A sequence remembers where to fold, so it folds not by biology but by remembering the size, charge, distance, and interaction. I told, it is folding by physics and chemistry instead of biology. When Christian Anfinsen performed this experiment, critics told that this was successful for probable ribonuclease and this would not be successful for other proteins. Unfortunately, Christian Anfinsen has no other experimental way to do that because protein folding was impossible to instrumental or experimentally demonstrate display at that time. But Christian Anfinsen got a huge backup from Levinthal.

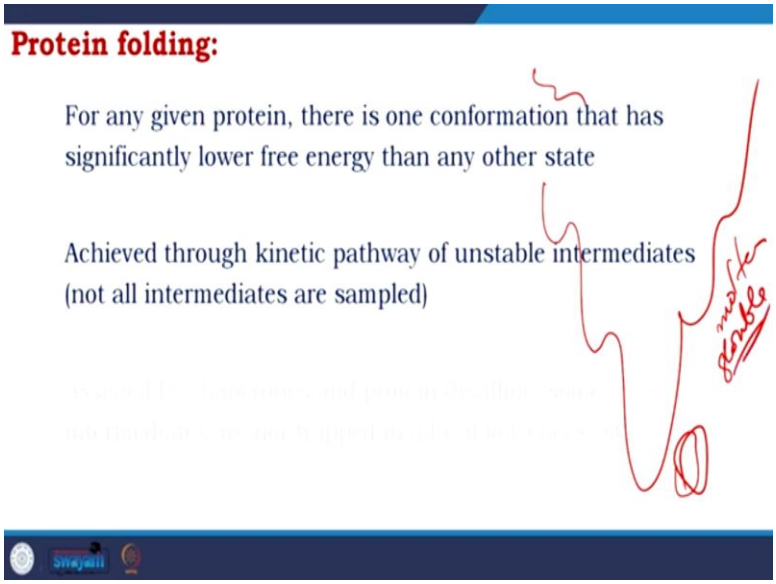
Levinthal told to assume that there are 3 confirmations of each amino acid. So, consider the minimum number of conformation for each amino acid. Now, if you take an average protein, its size range from 200 to 300 amino acids. So, he took the lowest one, a protein with 200 amino acids. So, the total number of conformation is $3^{200}$ is equal to 26 x $10^{95}$. Now, if consider 100 picoseconds which is $10^{-10}$ seconds is required to convert from one confirmation to another, a random search of all confirmation would require 26 x $10^{95}$ x 10 -$^{10}$ seconds, which is 3.2 x $10^{30}$ years. However, we all know that proteins take time to fold from millisecond to second order.

So what Levinthal proved is that protein folding is not a random phenomenon. The protein is not randomly searching for a stable interaction and then finding a stable interaction. Rather, it is a sophisticated search process. It is a thermodynamically driven process.

**(Refer Slide Time: 29:08)**

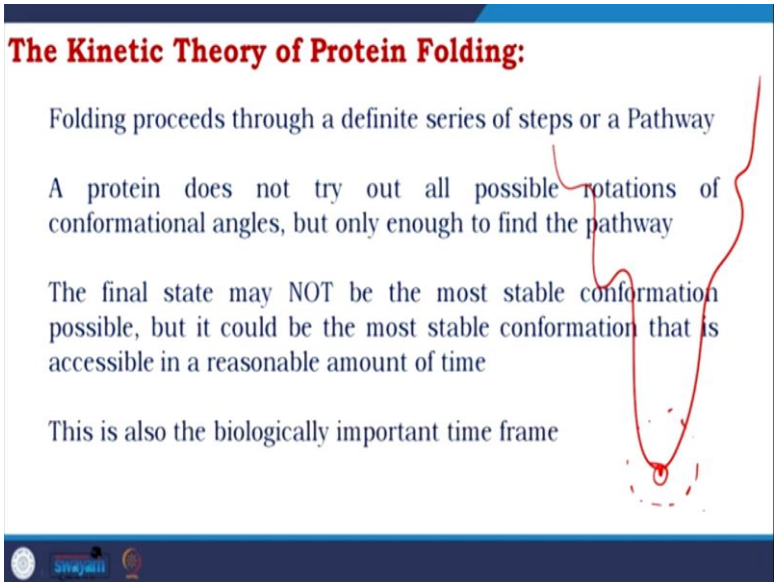
In the in-vitro situation, if you see the protein is in the much-diluted situation, maintaining proper temperature and buffering condition, you are making a dilute protein. Whereas when we look at an in-vivo situation, a protein does not have that much facility. Because, in the in-vivo situation in the presence of other proteins, proteolytic enzymes, and a load of other physiological problems, the protein must get a different folding mechanism

**(Refer Slide Time: 33:32)**



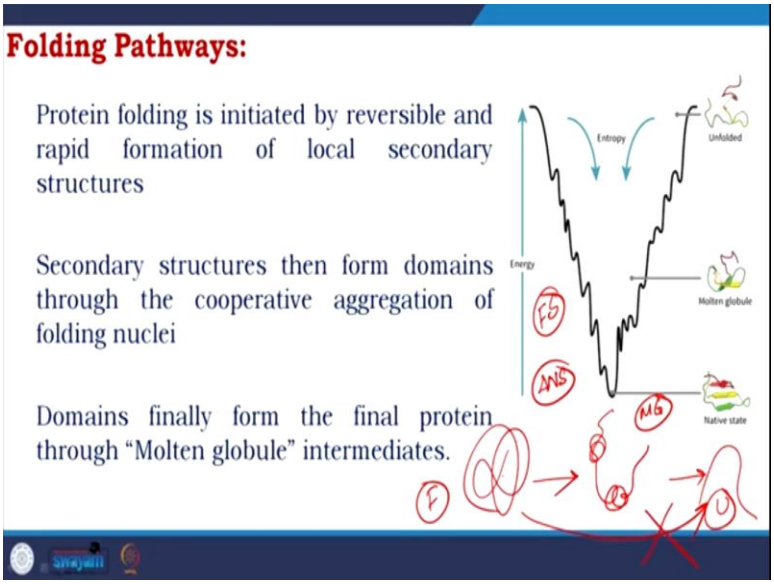

So, for any given protein, there is one conformation with significantly lower free energy than any other state. It would be achieved through the kinetic pathway of unstable intermediates, from unfolded protein to fold in between other states.

**(Refer Slide Time: 34:25)**



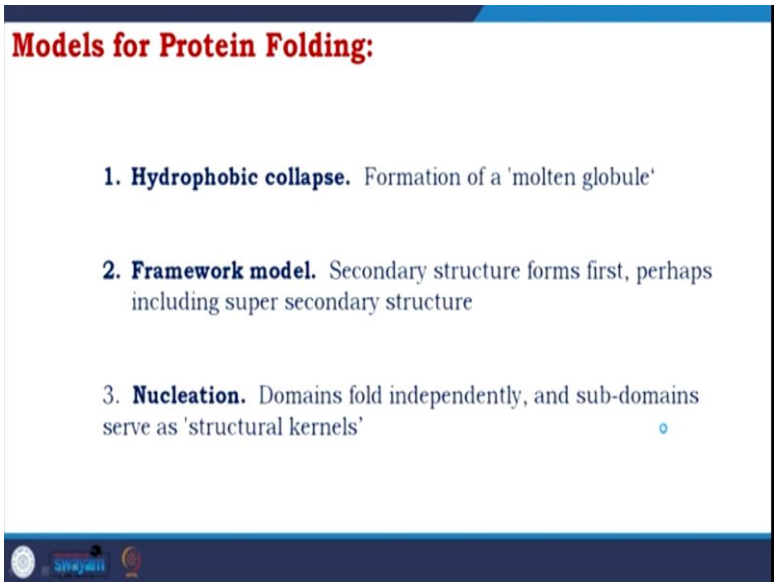

The folding process is through a definite series of steps or a pathway. A protein does not try out all possible rotations of conformational angles but only enough to find the pathway. The final state may not be the most stable conformation possible, but it could be the most stable conformation accessible in a reasonable amount of time.

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



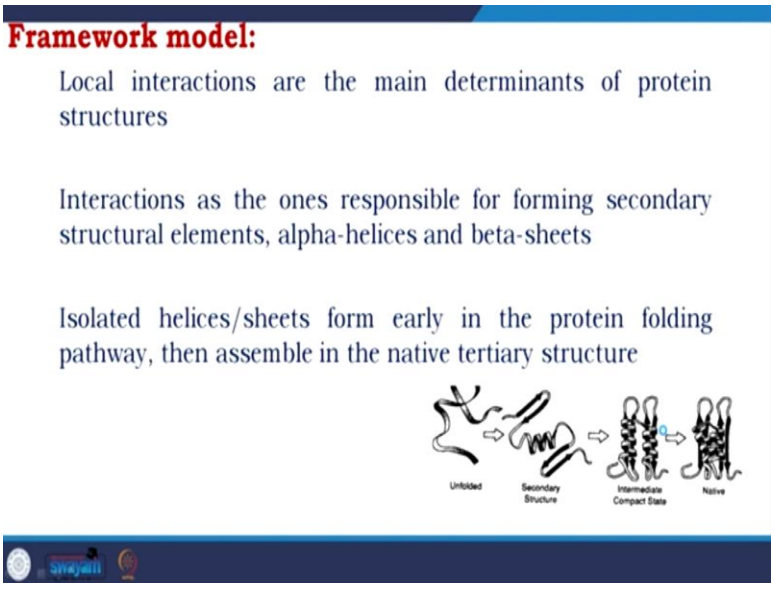

Protein folding is initiated by the reversible and rapid formation of local secondary structures, alpha helix and beta sheets and secondary structure, then from domains through the cooperative aggregation of folding nuclei. And domains finally formed the final protein through the molten globule intermediate molten globules are where the hydrophobic collapse happened.

**(Refer Slide Time: 37:40)**



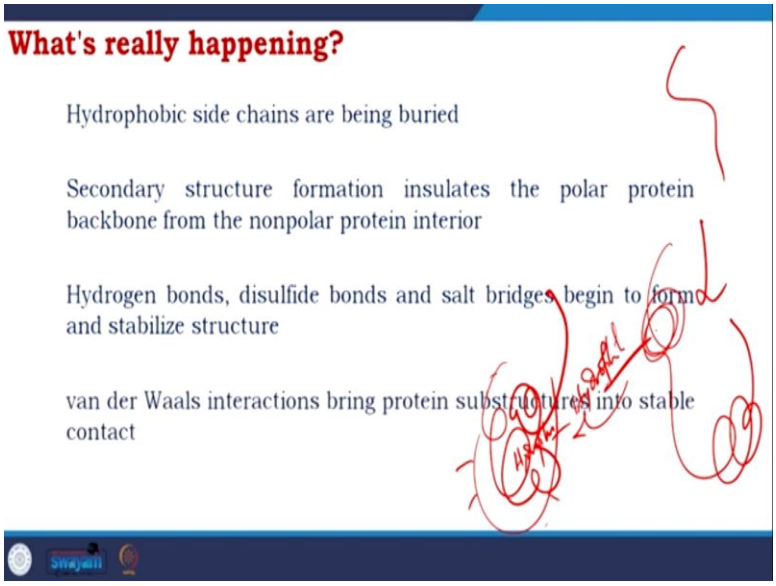

There are different proposed models for protein folding. Hydrophobic collapse formation of molten globule; In Framework models, the secondary structure forms, perhaps including super secondary structures, and then goes to the final structure, or nucleation model in which domains fold independently and sub domains are best structural kernels to help the protein domains fold.

**(Refer Slide Time: 38:05)**



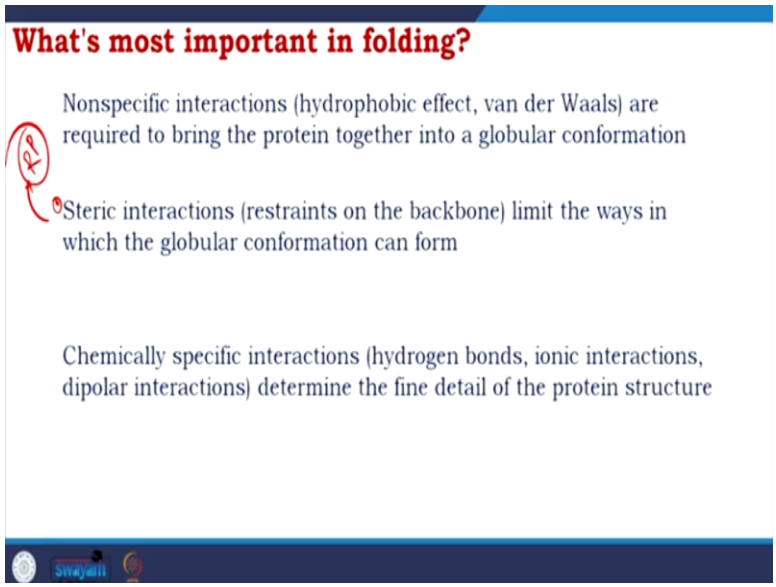

It is the local interaction, and they are the main determinants of protein structure. The interactions are responsible for forming secondary structural elements, isolated helices and sheets from early in the protein folding pathway, then assembled in the native tertiary structures.

**(Refer Slide Time: 38:44)**




The hydrophobic side chains are being buried. Secondary structure formation insulates the polar protein backbone from the nonpolar protein interior. Hydrogen bond disulfide bonds and salt bridges begin to form and stabilize the structure, van der Waals interactions bring proteins sub structures into stable contact.

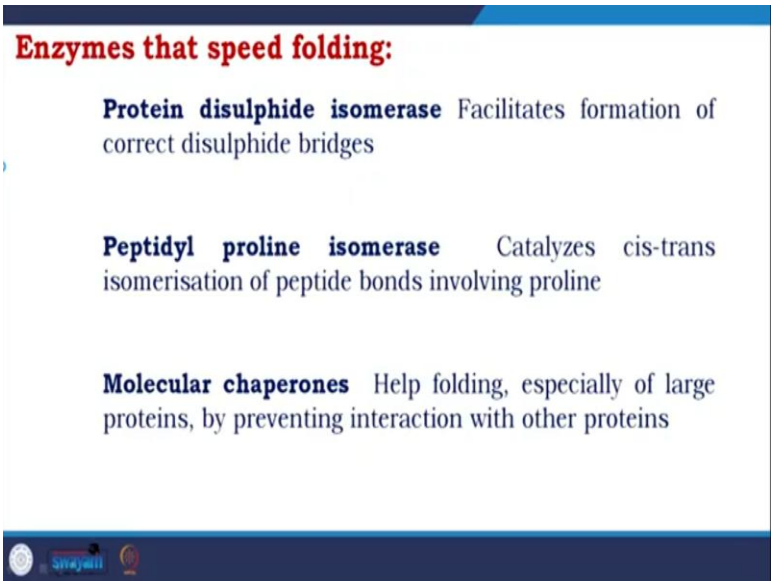
**(Refer Slide Time: 39:36)**




What are the most important in folding? Nonspecific interactions, hydrophobic effect, van der Waals are required to bring the protein together into a globular conformation. Steric interaction restraints on the backbone limit how the globular conformation can form; if you remember, this is where we are getting our Ramachandran plot. Chemically specific interactions like hydrogen

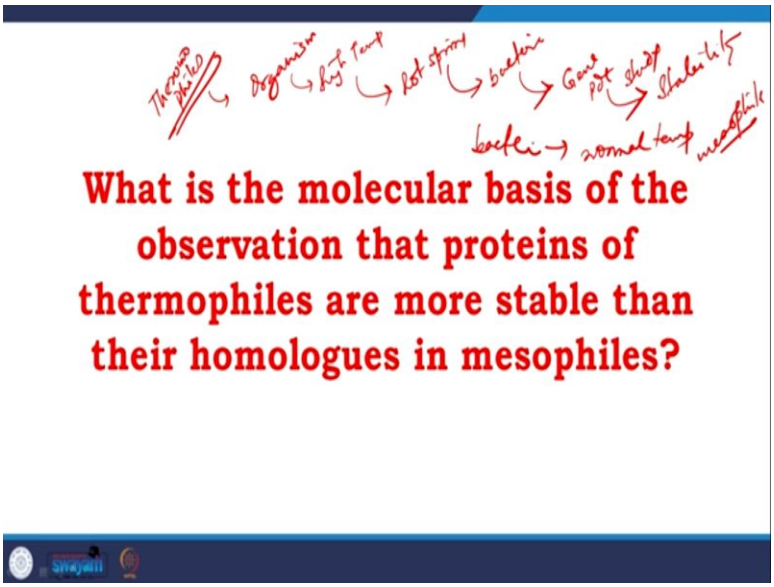
bonds, ionic interactions, solid breeze dipole interaction determined the fine detail of the protein structure.

**(Refer Slide Time: 40:09)**



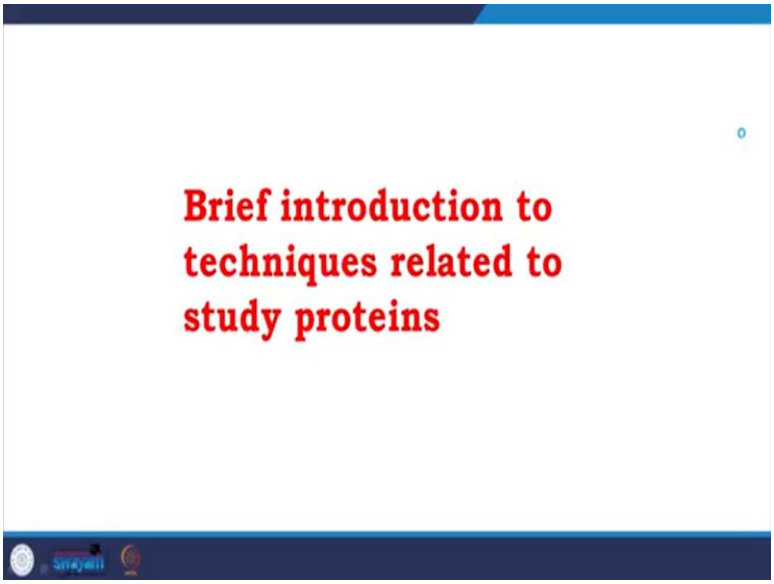

Some enzymes speed up the folding. Protein disulfide isomerase facilitates the formation of correct disulfide bridges. Peptidyl proline isomerase catalyzes cis-trans isomerization of peptide bonds involving proline. Molecular Chaperones helped to fold, especially large proteins, by preventing interaction with other proteins.

**(Refer Slide Time: 41:08)**



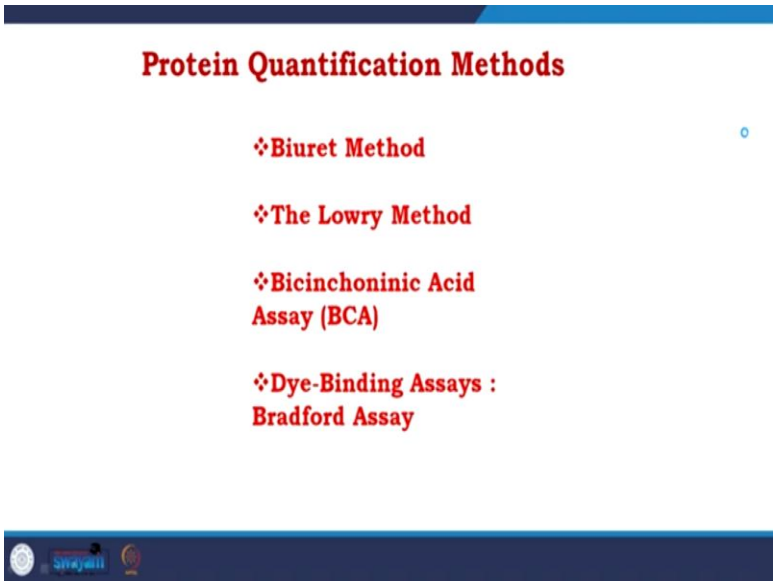
What is the molecular basis of the observation that proteins of thermophiles are more stable than homologs in mesophiles? Thermophiles are organisms that live in high temperatures. For example, go to a hot spring, and you collect bacteria from there. You will find that a protein, a gene product, I would say coming from that bacteria if you study, will show more stability than the same protein in a bacteria that is staying in normal temperature called mesophile.

**(Refer Slide Time: 42:41)**



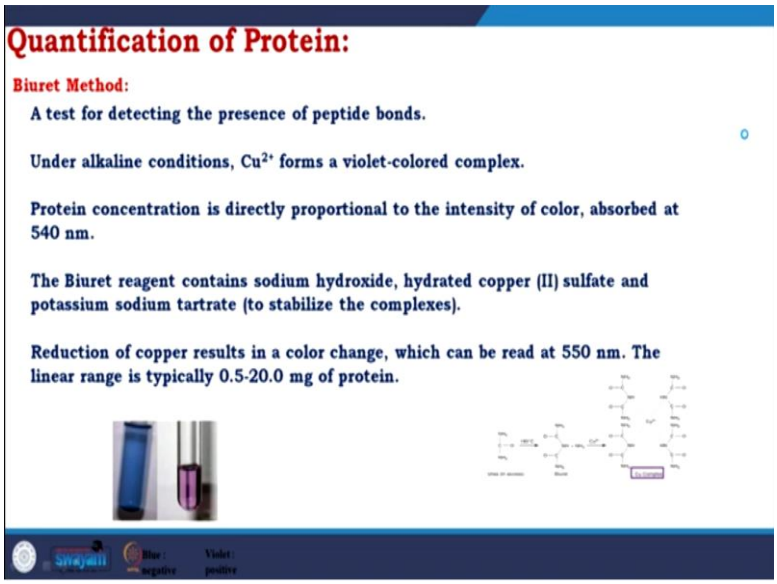

In the second part of today's lecture, I will give you a very brief introduction to the techniques related to studying proteins; I will start with a very simple protein quantification method.

**(Refer Slide Time: 42:54)**



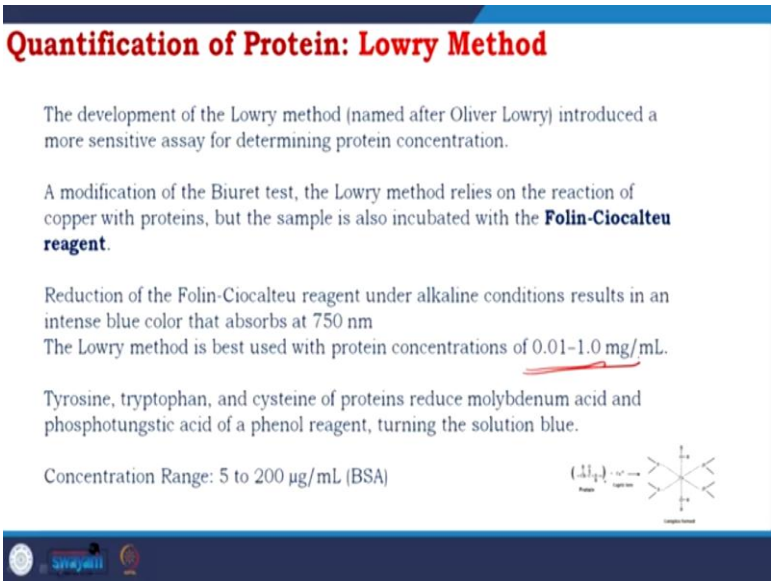
There are different methods, but I will talk about Biuret method, Lowry method, Bicinchoninic acid or BCA Assay, and Dye-Binding Assays. I will take an example of Bradford Assay.

**(Refer Slide Time: 43:06)**



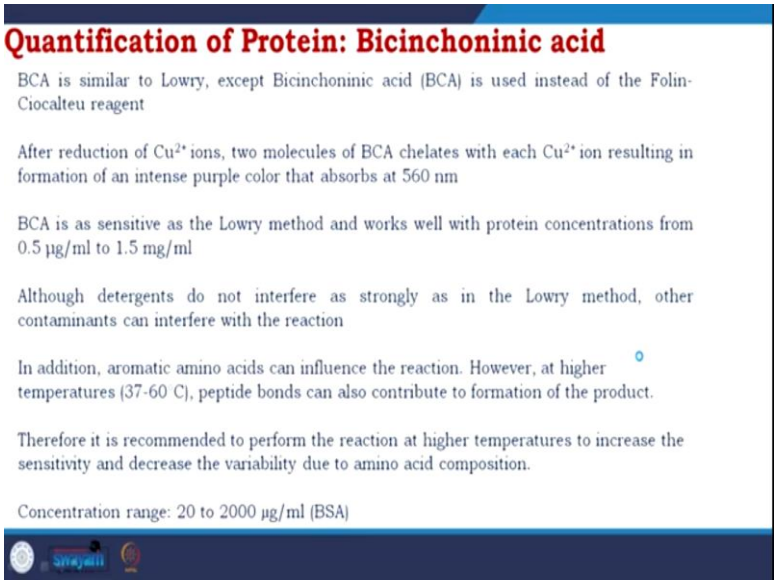

The Biuret method is a test for detecting the presence of peptide bonds. Under alkaline conditions, copper forms a violet-colored complex. Protein Concentration is directly proportional to the intensity of color absorbed at 540 nm. To stabilize the complex, the Biuret reagent contains sodium hydroxide, hydrated copper II sulfate, and potassium sodium tartrate. Reduction of copper results in a color change. The linear range is typically 0.5 to 20 milligrams of protein

**(Refer Slide Time: 44:06)**



It is named after Oliver Lowry, who introduced a more sensitive assay. It is a modification of Biuret tests. The Lowry method relies on the reaction of copper with protein the same, but the sample is also incubated with a reagent called Folin-Ciocalteu reagent. The reduction of the reagent under alkaline conditions results in an intense blue color that absorb at 750 nm. The Lowry method is best used with a Protein Concentration of 0.01 to 1 mg per ml. Tyrosine, Tryptophan, and Cysteine of protein reduce molybdenum acid, and phosphotungstic acid have a phenol reagent, turning the solution blue. The concentration range is 5 to 200 micrograms per ml. So, this is estimated with the protein bovine serum albumin.

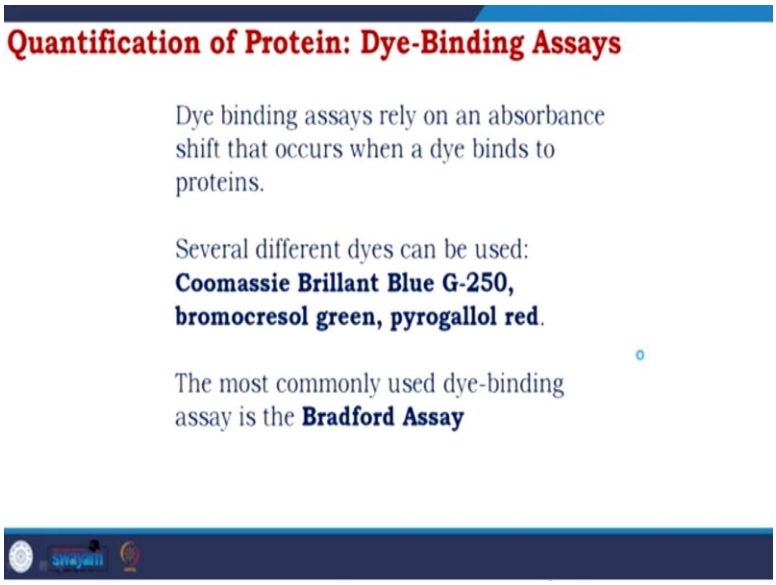
**(Refer Slide Time: 45:27)**




Bicinchoninic acid is again similar to Lowry, except the BCA replaces the Folin-Ciocalteu reagent. After the reduction of copper ions, 2 molecules of BCA chelates with each copper iron, forming an intense purple color that absorbs at 516 nm. This test is as sensitive as the Lowry method and works well with a Protein Concentration from 0.5 micrograms per ml to 1.5 milligrams per ml.

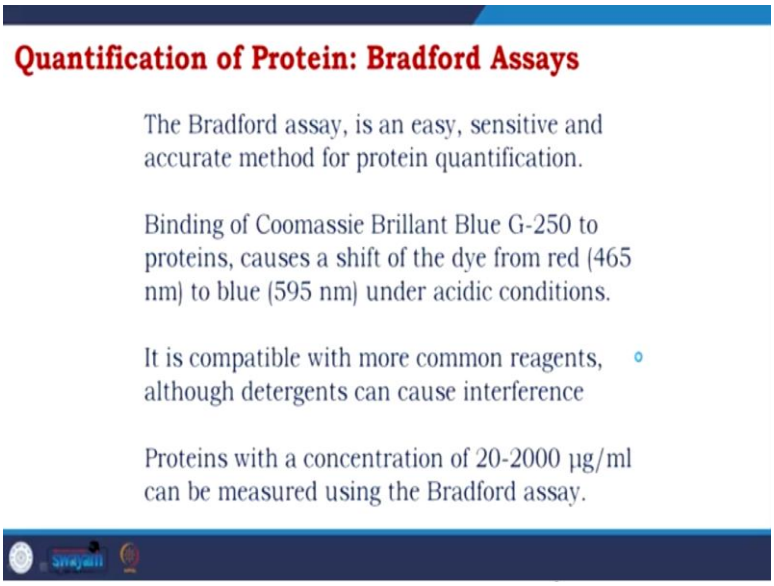
Although detergents do not interfere as strongly as the Lowry method, other contaminants can interfere. In addition, aromatic amino acids can influence the reaction. However, at higher temperatures, peptide bonds can also contribute to the formation of the product. Therefore, for these problems, performing the reaction at higher temperatures is recommended to increase the sensitivity and decrease the variability due to amino acid composition—concentration range from 20 to 2000 micrograms per ml.

**(Refer Slide Time: 46:37)**



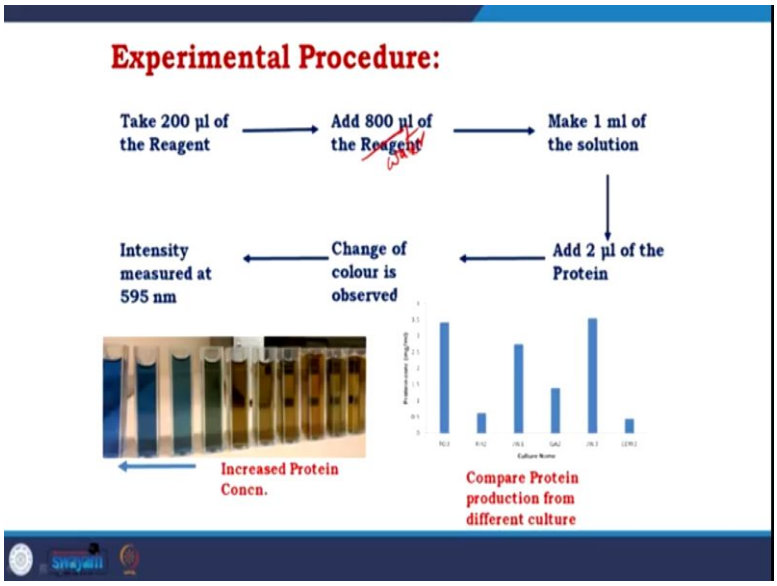

Dye-Binding Assays rely on an absorbance shift that occurs when a dye binds to a protein. Several different dyes could be used Coomassie Brilliant Blue G-250, Bromocresol green, Pyrogallol red, but the most commonly used dye-binding assays are Bradford Assays.

**(Refer Slide Time: 47:07)**



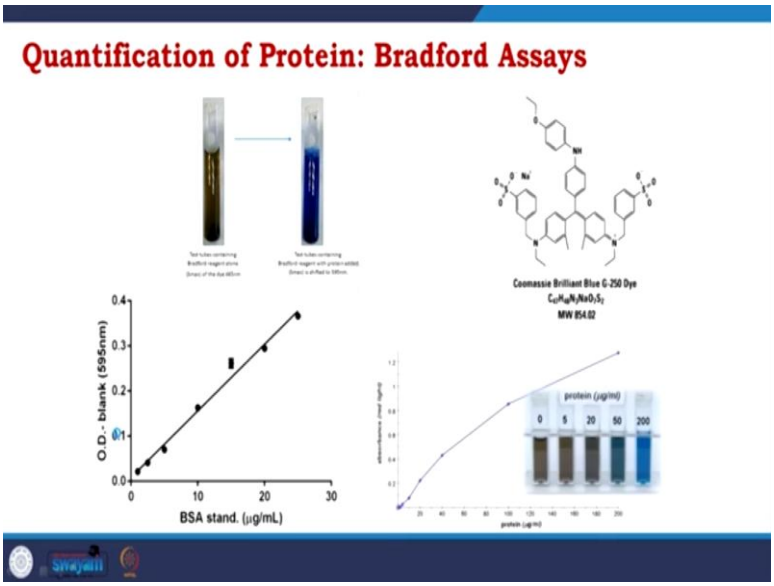

The Bradford Assay is an easy sensitive, and accurate method for protein quantification. The binding of Coomassie brilliant blue G-250 to proteins causes a shift of the dye from Red 465 to blue 595, nanometer under acidic conditions. It is compatible with more common reagents, although detergent can cause interference. Protein with a concentration of 20 to 2000 micrograms per ml can be measured using Bradford Assay.

**(Refer Slide Time: 47:38)**



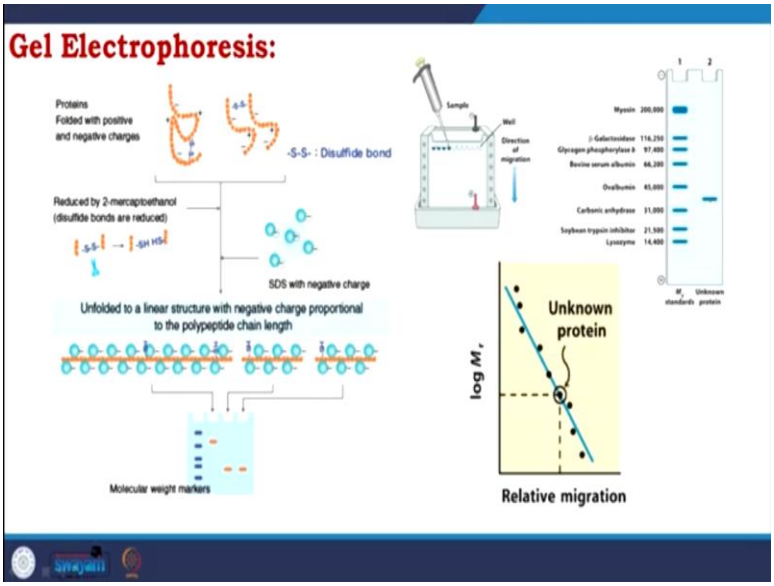

So, how is it used to measure concentration? Take 200 microliters of the reagent, and add 800 microliter of water. So, you have 1 ml of the solution and add 2 microliters of protein, a change of color is observed, and intensity is measured at 595. You see that when you keep changing the higher protein concentration, you gradually start getting a more intense blue color.

**(Refer Slide Time: 48:50)**



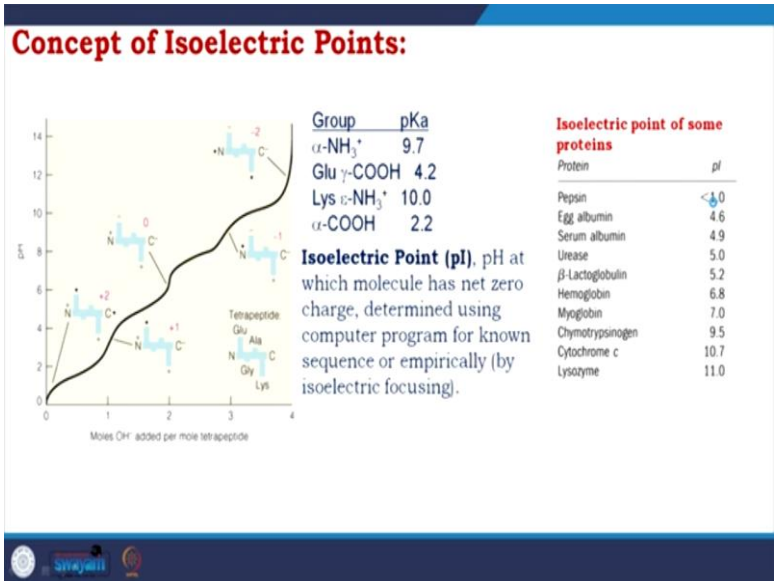

You could also do real quantification. So, for that, you have to use the BSA standard. So, using those reagents, you could quantify protein.

**(Refer Slide Time: 49:38)**

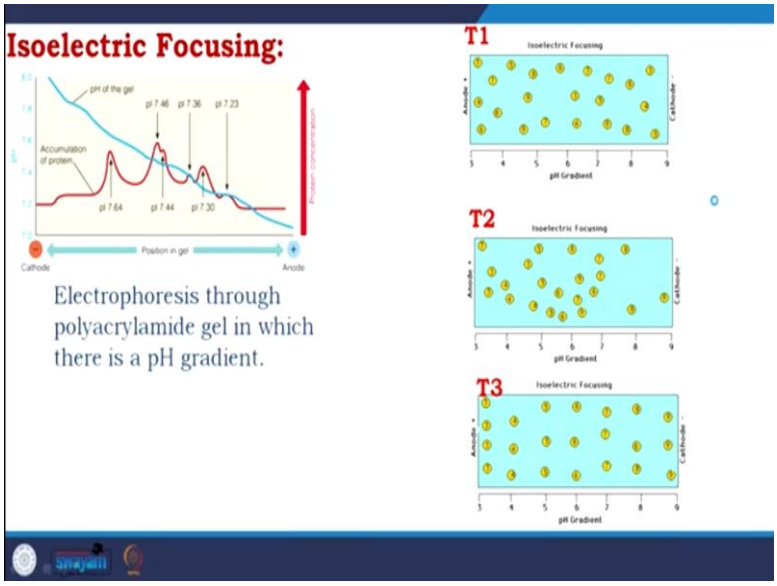



Another very interesting technique is gel electrophoresis. You have the protein. The protein might have a disulfide bond and a positive or negatively charged surface. To reduce the disulfide bond, add 2-mercaptoethanol and boil. Also, you add sodium dodecyl sulfate, which is called SDS gel electrophoresis. More importantly, SDS is now the negatively charged ion, so the entire protein is bound with SDS. So, all of them are negatively charged. So, when you run the gel, you stain it and see the bands, and nowadays, there are advanced markers that will give bands even without staining. Also, from the relative migration, you could get the mass of the unknown protein more accurately.

**(Refer Slide Time: 52:03)**

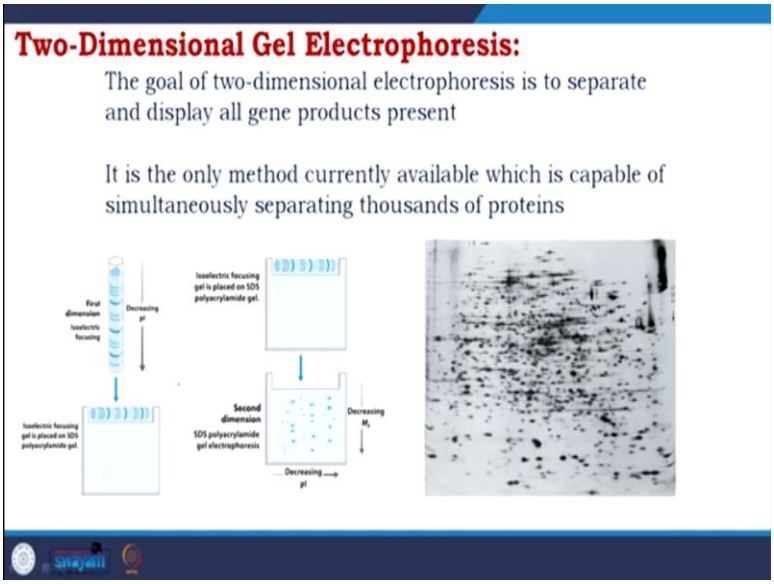



**(Refer Slide Time: 53:27)**



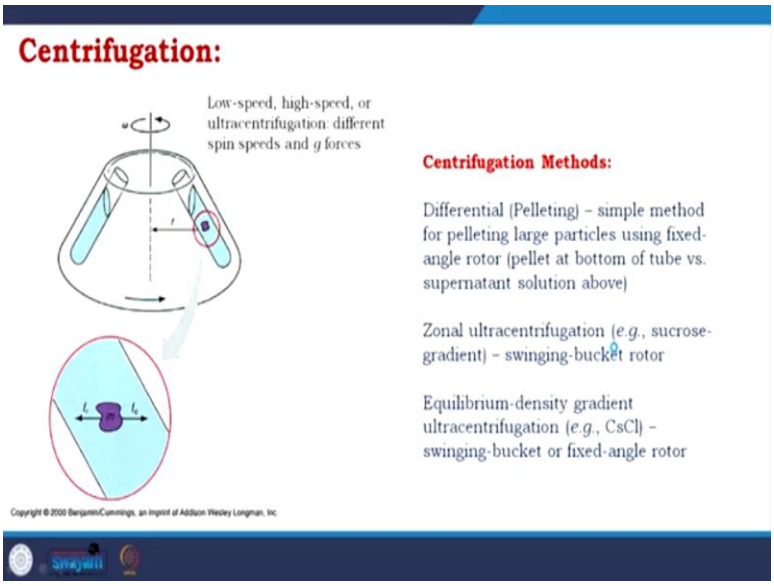

According to that pI, you could perform isoelectric focusing electrophoresis through polyacrylamide gel with a pH gradient. So, you start in time 1; they are random, then gradually separated, and ultimately, they are separated from anode to cathode because of the Isoelectric Focusing the difference in the pH gradient along with the different pIs.

**(Refer Slide Time: 54:02)**



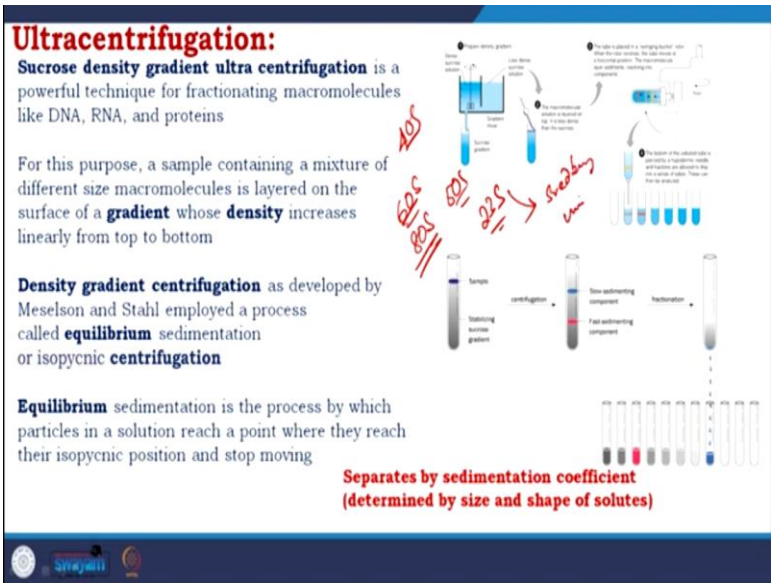

Now, you combine these 2 techniques, Isoelectric focusing and gel electrophoresis, called 2-dimensional gel electrophoresis. 2-dimensional electrophoresis aims to separate and display all gene products present. Let us say in a cell you are experimenting with, it is the only method currently available that can simultaneously separate 1000s of protein. That is the beauty.

**(Refer Slide Time: 55:16)**




Centrifugation is a very important technique. Low, high, or ultra-centrifugation different spin speeds and G forces. So, it needs different work to have a cell grow. You want to get a pallet. So, you have the lowest one if you have protein to isolate, so, your higher speed and then ultracentrifugation, you have to differentiate between cell organelles or something.

**(Refer Slide Time: 56:30)**

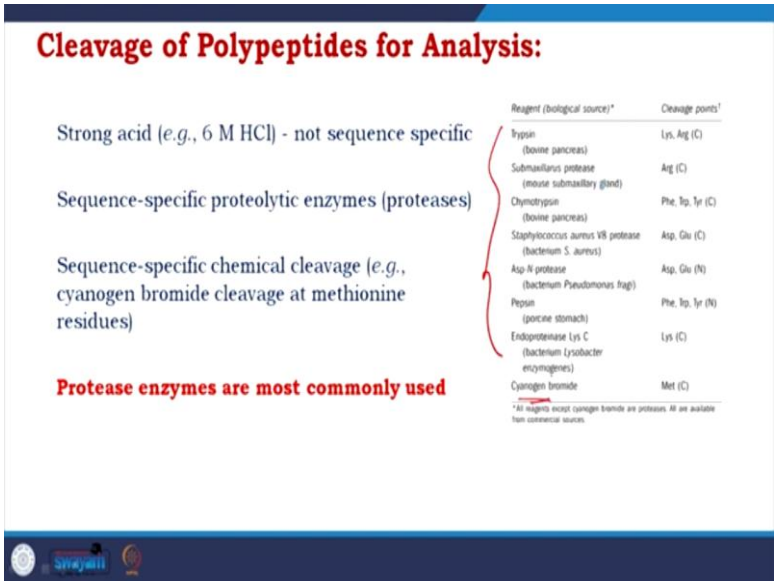



Sucrose density gradient ultracentrifugation is a powerful technique for fractionating macromolecules like DNA, RNA, and protein organelles. For this purpose, a sample containing a

mixture of different-sized macro molecules is layered on the surface of a gradient whose density increases linearly from top to bottom.

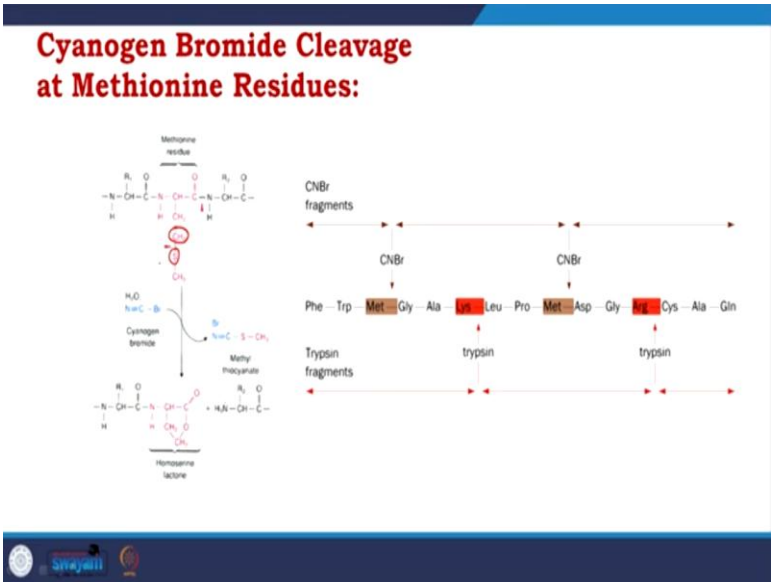
Meselson and Stahl developed this in their famous experiment, which employed an equilibrium sedimentation process, or Isopycnic centrifugation. Equilibrium sedimentation is a process by which particles in a solution reach a point where they reach their Isopycnic position and stop moving.

**(Refer Slide Time: 58:13)**



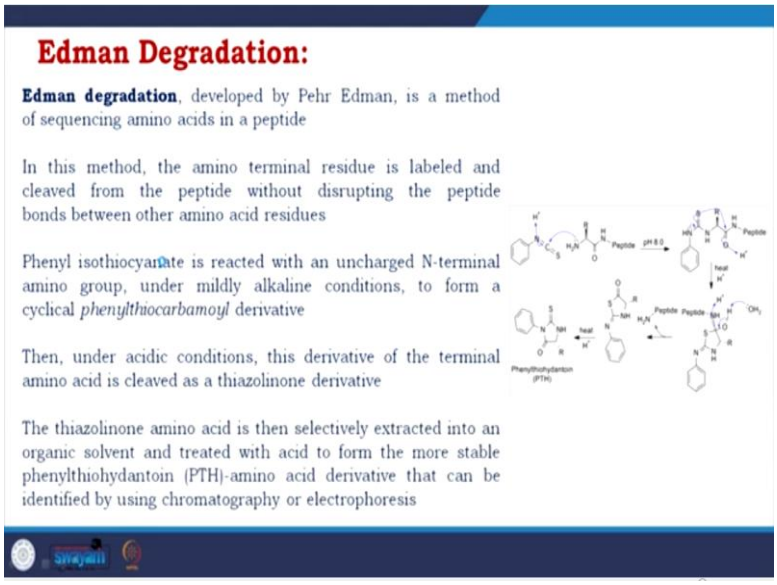

So, how a protein or polypeptide would be cleaved for further analysis? You could use strong acid like 6 molar HCl, but you would not get any specificity there. You use sequence-specific proteolytic enzymes, which are called proteases. You could use sequence-specific chemical cleavage and cyanogen bromide cleavage at methionine residues. Among them, the protease enzymes are most commonly used and are diverse. So if you look, trypsin cleavage points are basic amino acids, lysine, and arginine. Then you get Submaxilarus protease which cuts in arginine specifically, Chymotrypsin, phenylalanine, tryptophan, tyrosine, Staphylococcus aureus V 8 protease aspartate, and glue the acidic ones. Asp-N-protease, aspartate and glutamate, pepsin phenylalanine, tryptophan tyrosine, endoproteinase Lysin.

**(Refer Slide Time: 59:31)**




So, how cyanogen bromide keeps the Methionine residue, you will see that it attacks and keeps the residue between the methyl group and the sulfur, forming a homoserine lactone. It could mainly work on Methionine and be used in combination with trypsin.

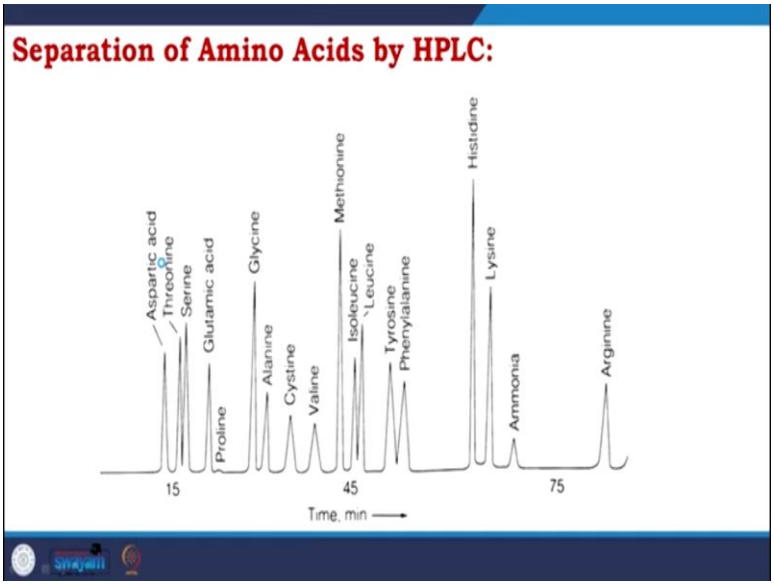
**(Refer Slide Time: 1:00:02)**




Edman degradation is a very famous method. Phil Edman developed it as a peptide sequencing method of amino acids. In this method, the amino-terminal residue is labeled and cleaved from the peptide without disrupting the peptide bonds between other amino acid residues. Phenyl isothiocyanate reacts with an uncharged N-terminal amino group under mildly alkaline condition to form a cyclical phenylthiocarbamoyl derivative and does not affect the other peptides.

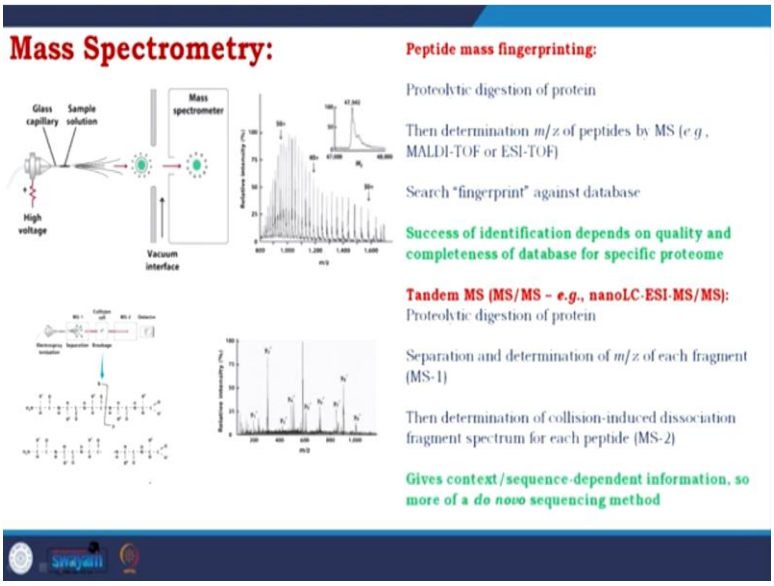
Then under acidic conditions, this derivative of the terminal amino acid is cleaved as a thiazolidine derivative. And ultimately, this isothiazolinone amino acid is selectively extracted into an organic solvent and treated with acid to form a more stable PTH or phenylthiohydantoin. This can be identified by using chromatography or electrophoresis. So, in this way, you could keep identifying the N-terminus amino acid.

**(Refer Slide Time: 1:01:25)**




When you cleave them, you can separate them using high-performance liquid chromatography, a very popular method of identifying small molecules.

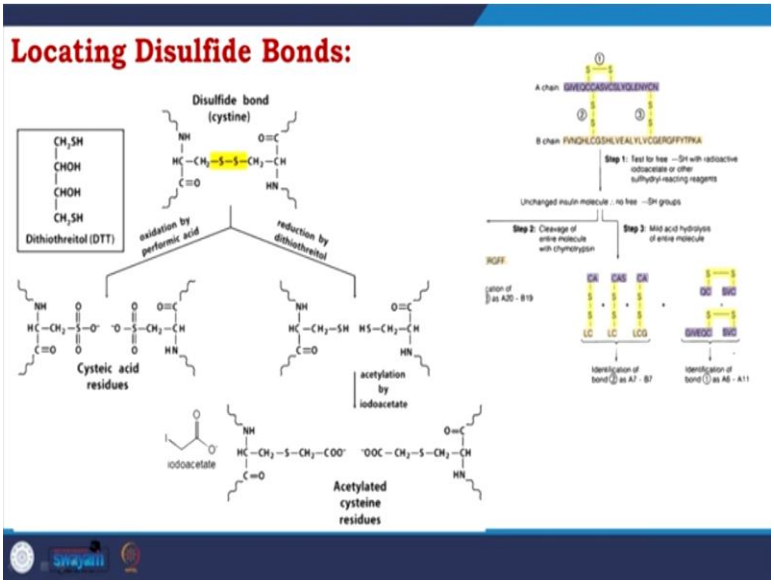
**(Refer Slide Time: 1:01:35)**



Mass Spectrometry is a very popular method. Generally, 2 types of mass spectrometry method are used here, and I am talking about peptide mass fingerprinting. You have to do the proteolytic digestion of protein for any process. You have to do that first, and then you determine the m / z of peptides by mass spectrometry either by MALDI-TOF or ESI-TOF electrospray. Then you search for fingerprints from the database.

And the success of this method depends on the quality and completeness of the database of the specific proteome where you are working. Another method is known as Tandem mass spectrometry. Here the same thing you do is the proteolytic digestion of protein, but then you do a separation, you could say. So, you could do the proteolytic digestion of protein instead of going there to do m / z. You do separation and then a determination of m / z. This is called the first step and then the determination of each peptide's collision-induced dissociation fragment spectrum separately.

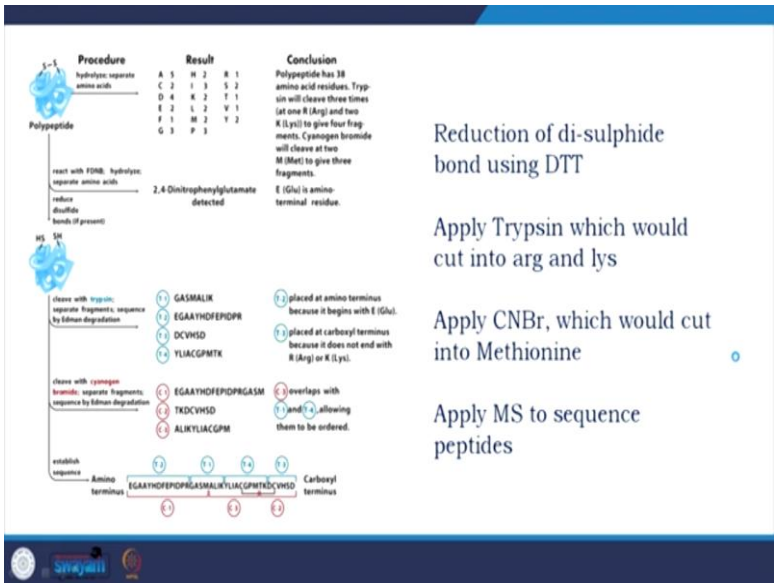
**(Refer Slide Time: 1:03:14)**




How you look at the disulfide bond if you have a disulfide bond present, you either do the reduction by adding dithiothreitol do the oxidation by performic acid. In the case of performic acid, you get cystic acid residues. In the case of dithiothreitol, disulphide bond would be reduced, an individual S-H group would be formed, and then you put iodoacetate, it would do escalation, and now you get asked related cysteine residues.

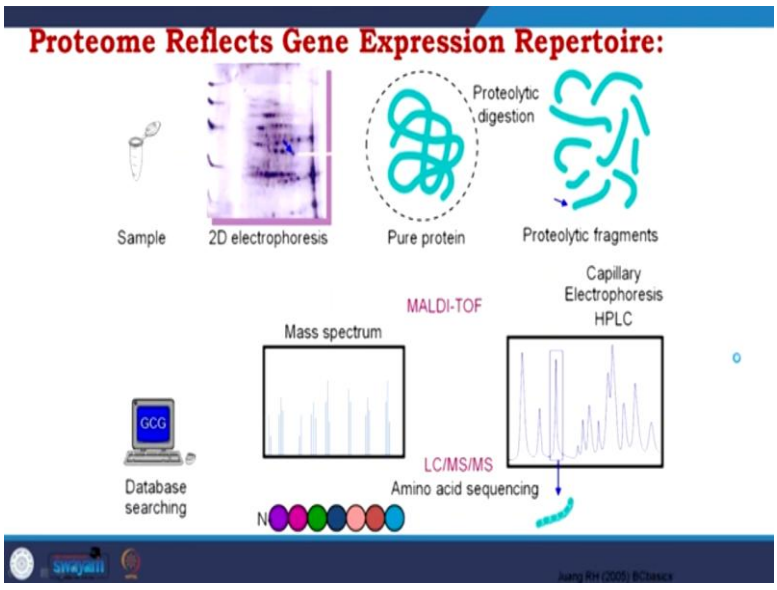
So, in the protein, you identify the escalated cysteine residues and know where the disulfide bonds are.

**(Refer Slide Time: 1:04:26)**




And what you do is the reduction of disulfide bond using dithiothreitol. You apply trypsin, the protease which would cut into arginine and lysine. You apply CNBr, which would cut into Methionine, and then you do Edman degradation, know the N-terminus ones. You apply mass spectrometry to get a fingerprint of the sequence.

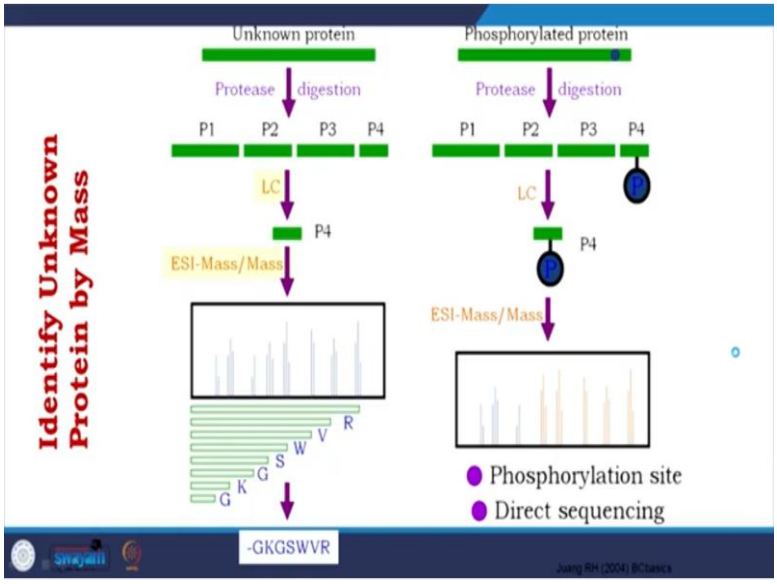
**(Refer Slide Time: 1:04:52)**




You have the sample, and you do the 2d electrophoresis, you get the pure protein, you do the proteolytic digestion and get the proteolytic fragments that proteolytic fragments go through the

MALDI-TOF you get the mass spectrum, or you purify with capillary electrophoresis. Then you go for LC-MS-MS for amino acid sequencing, whereas with the mass spectrum or fragment, you go for database searching.

**(Refer Slide Time: 1:05:28)**




But then you have an unknown protein, you do protease digestion again, you get the fragments p1, p2, p3, p4 with them, you go for MALDI-TOF these fragments are there, you get m / z values for each of them, and you know their mass. Now, you search the database, get the candidate protein, make the digestion simulation, calculate the molecular weight, compare the peptides, and identify the unknown proteins.