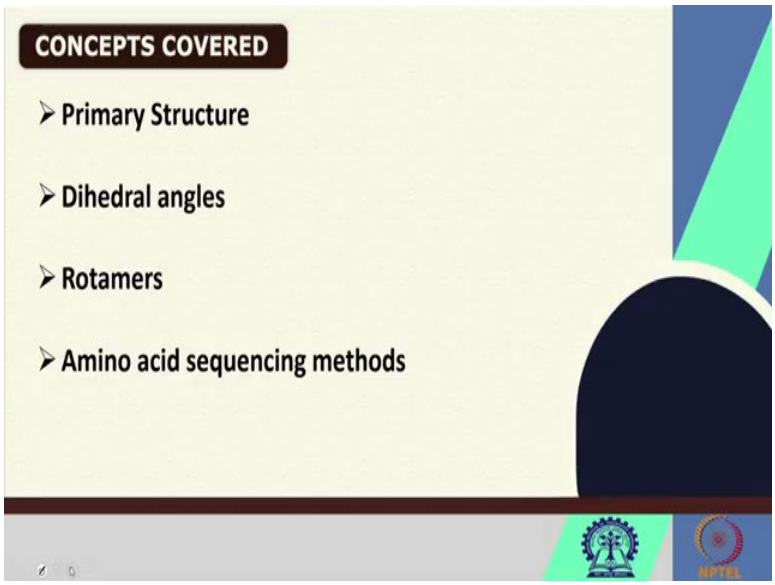
Fundamentals of Protein Chemistry

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

Module - 02
Protein Architecture
Lecture - 06
Primary Structure

After our discussions on the amino acids, their properties and the formation of the peptide bond, we now want to see what this polypeptide chain entails in terms of its structure. So, in the second module, our first description here is going to be on the primary structure of the proteins; that is going to be concerned with the protein architecture.

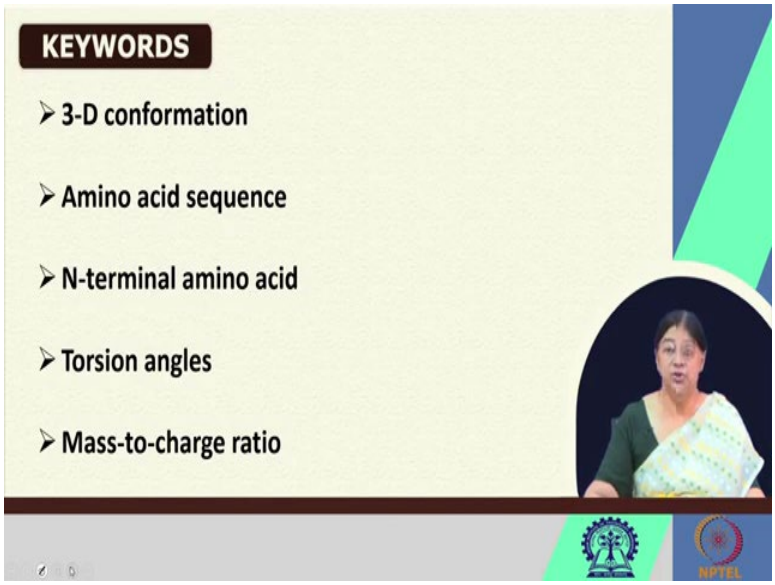
(Refer Slide Time: 00:42)



When we consider the primary structure, before we go into an understanding of what the primary structure is or what the amino acid sequence is, we have to understand a bit about the geometry of the backbone as it is called.

We understood that when we have the peptides linked together, we have the peptide bond itself that has a partial double bond character with restricted rotation, but in addition to the single bonds the other single bonds that are present, there is a possibility of rotation and because of this rotation there are various factors that play an important role in protein architecture.

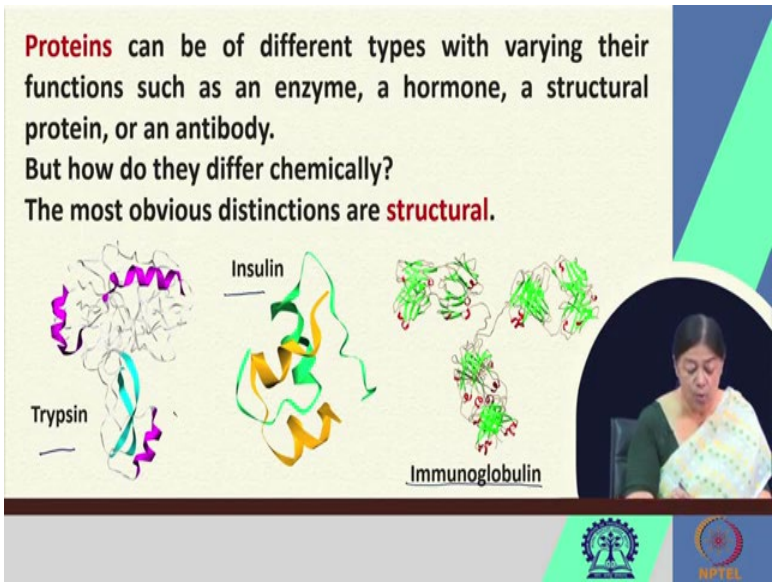
(Refer Slide Time: 01:33)



We will look at the 3-dimensional conformation of proteins, the amino acid sequence, the N terminal, the torsional angles and also the mass-to-charge ratio in terms of the total protein and how we can go for specific sequencing methods. We will also understand the necessity of doing

this or understanding this, because we want to know more about proteins, the protein chemistry, the protein function and so on.

(Refer Slide Time: 02:05)

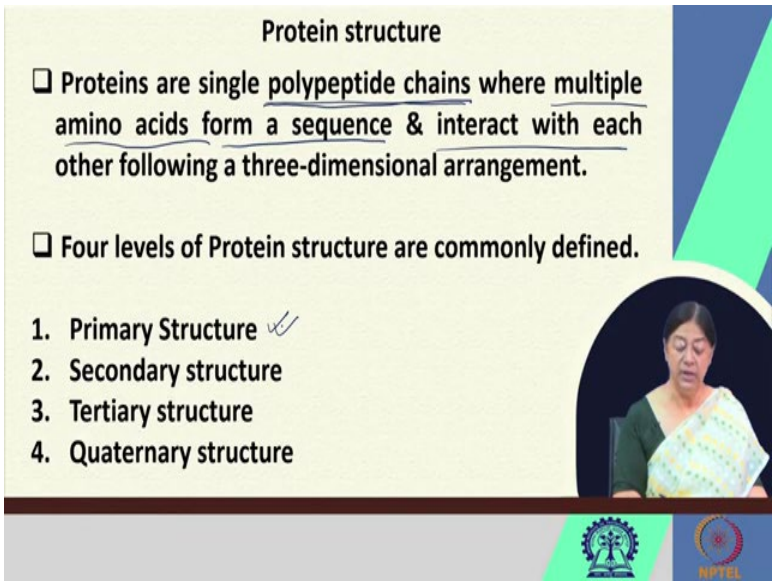


So, what are proteins? Proteins can be of various types, they can be enzymes, they can be hormones, they can have structural proteins or even antibodies. So, if we look at the overall structure of the proteins, they are distinct in nature.

They are composed of the common amino acids; the ones that we studied in the previous module, but they are linked together in varying fashions. They fold in varying fashions and because of that they result in different types of three-dimensional structures that lead to different types of functions and this is what we will gradually understand as we go through the course.

So, to look at proteins such as trypsin, insulin and immunoglobulin-which is an antibody, we will try and realize what these particular representations mean and why do we have these representations. Why is the structure folded as in terms of its interactions. An extremely important factor is in trying to understand what we mean by protein folding.

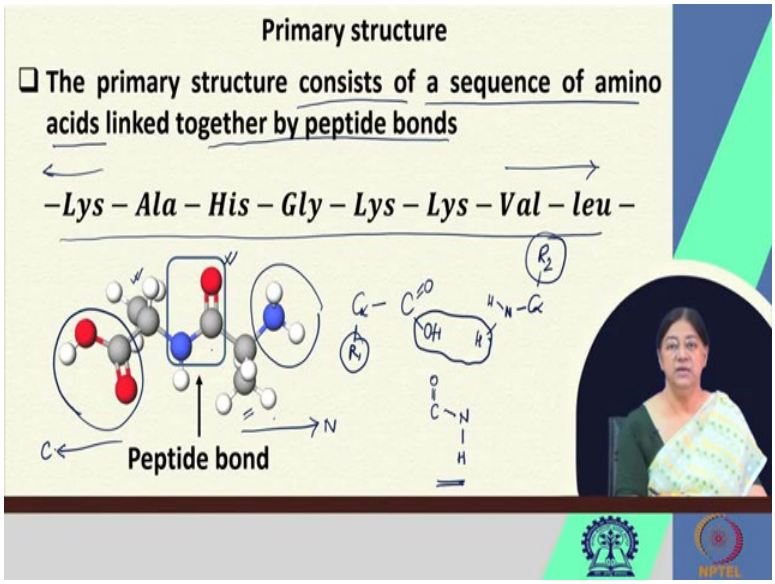
(Refer Slide Time: 03:30)



Now, when we consider the protein structure there are different levels; the primary structure is one such level. What do we have for these proteins? As we saw in the class for the peptide bond, we had the formation of a polypeptide chain. Now, what this polypeptide chain does is, it has a combination of multiple amino acids that form a sequence and the interaction with each other follows or finally, forms a three-dimensional structure.

For now, we know that we have a polypeptide chain linked by peptide bonds, that is what forms the primary structure. After this we have the secondary structure, the tertiary structure and the quaternary structure that is formed by some proteins which we will see as we go along in the lectures.

(Refer Slide Time: 04:33)



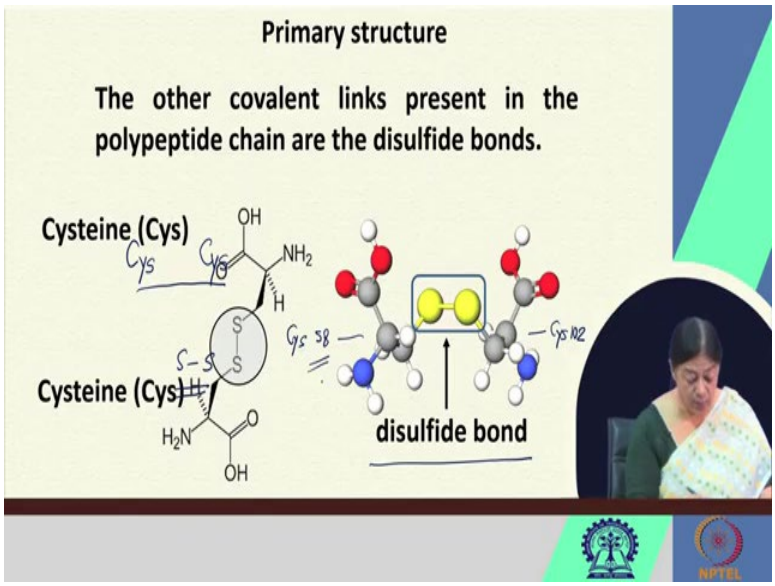
When we look at the primary structure of proteins, we see they consist of a sequence of amino acids, linked together by the peptide bonds. What we know is on the left-hand side we have the N terminus on the right hand side we have the carboxyl terminus, because the peptide bond is formed from a combination of the CO and the NH. Now, if we look at the way the peptide bond has been drawn [refer to slide], we see the N terminus and the C terminus. One R group is amino acid number 1 and we also have amino acid number 2. The amino acid that begins the protein is the N terminus.

We know how the peptide bond is formed. It is formed from the COOH of amino acid number 1 say, that is going to condense with amino acid number 2 say. And what do we have? We have the loss of water, as we saw in the lecture on the peptide bond and we have the C=O-N-H form.

The C=O and the NH is the peptide bond. But we realize that this C=O has to be from the previous amino acid. So, this is the direction of the N terminus and this is the direction of the C terminus. And you have to know how to identify the directionality of the polypeptide chain, because it has a definite direction to it.

So, the primary structure consists of the sequence of peptides, sequence of amino acids linked by peptide bonds.

(Refer Slide Time: 07:05)

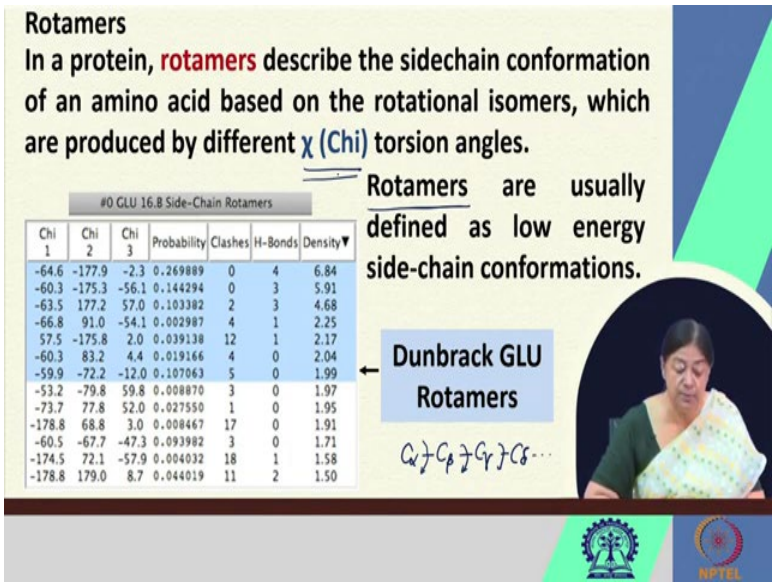


The other covalent links that are present in the polypeptide chain are the disulfide bonds. We saw this when we studied the cystine, where we had the S bond formation from two cysteine amino acids. When two cysteine amino acids link together to form the disulfide linkage in proteins it is called a cystine and this is helpful in bringing a folded structure in the terms of having different parts of the polypeptide chain link together.

This may be cysteine number 58, this may be cysteine number 102 in a large protein [refer to slide]. So, what happens is when they come together, they bring different parts of the protein close to one another in three dimensional space which we will see later on.

So, the covalent bonds apart from post translational modifications as mentioned earlier, these are the other covalent linkages that are present in proteins, the disulfide bonds in addition to the peptide linkages that we studied. So, this is where we have the formation of the disulfide, two cysteines coming together forming a disulfide linkage - a covalent linkage.

(Refer Slide Time: 08:53)

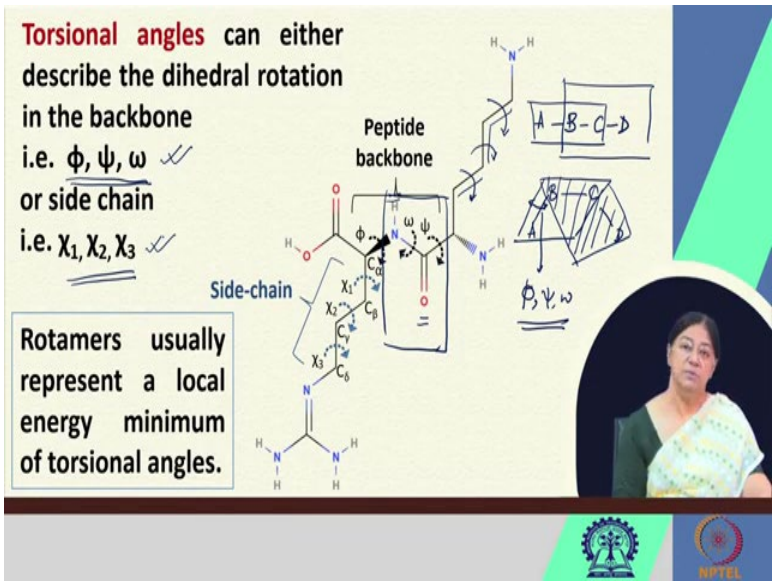


If we just go back a slide and look at the amino acids themselves, there are several single bonds. If we have a longer polypeptide chain, we have several single bonds about which we can have rotations. Suppose we have a Cα followed by a Cβ followed by a Cγ even a Cδ and so on and so on. These are amino acids with longer chains.

Now, all of these are linked by single bonds. So, there is a possibility of rotations about these bonds. This gives what are called rotamers giving us chi angles.

We see different kinds of structural isomers that we can have in terms of a cis trans for the peptide, but when we look at the geometry it is important to understand that the amino acids themselves, because of the presence of the single bonds, can have specific side chains that can rotate, as we have single bonds that link the carbons together.

(Refer Slide Time: 10:25)



So, if we look at the torsional angles that describe the dihedral rotation, there are specific dihedral rotations, which we will study in a minute, very important for the protein backbone geometry and the side chain rotamers that give a library of possible structures because of the rotation along the chi 1, chi 2, chi 3. What do we mean by this? It is an amino acid where we have a peptide bond C=O-N-H.

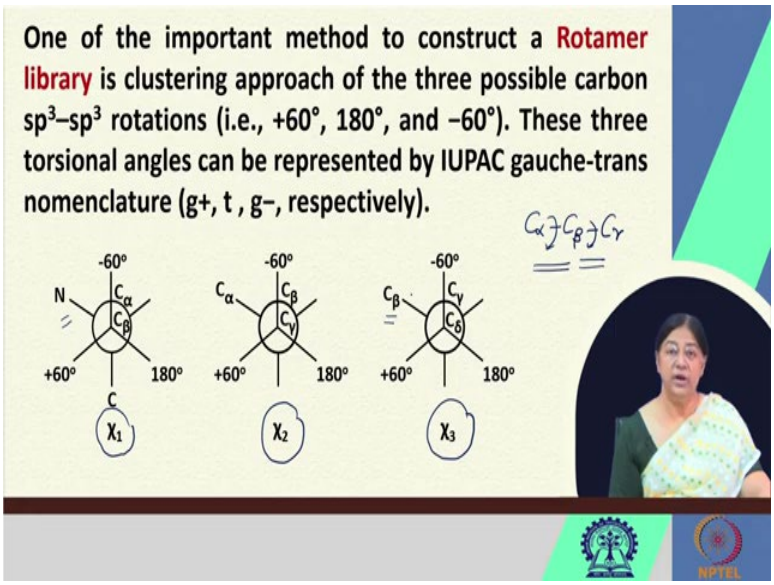
The definition for a dihedral angle is, it is an angle present between two planes. So, for this we have to define say four atoms. In a sense if we have four atoms linked together A, B, C, D, we know that we can define a plane by using just three points or three atoms.

So, if I have three atoms I can define a plane using A B C, I can define another plane using B C D and the angle between these two planes is what we call the dihedral angle.

So, if I look at a plane that comprises A B C, and I look at another plane that comprises B C D - the angle formed between these two planes is our dihedral angle. So, depending on the definition it will be φ, ψ, ω angles that we will see as we go along.

So, the dihedral rotation is in terms of the backbone angles φ, ψ, ω, but when we look at the side chain as we looked at the different single bonds. So, this is the arginine amino acid residue and we see there are several single bonds. Now, there is rotation possible about these single bonds, which is what is termed as the chi 1, chi 2, chi 3 giving us the rotamers that usually represent a local energy, minimum in terms of the torsional angles, giving us the best possible structure or the best possible orientation that is going to give us the minimum energy.

(Refer Slide Time: 13:33)



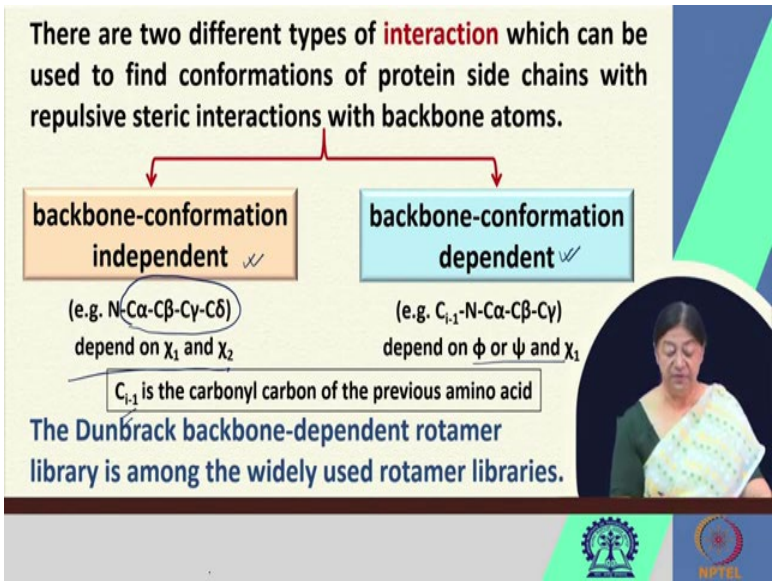
The important method to construct a rotamer library is what is called a cluster approach. Given that we have the C single bond linkages, the carbon atoms are sp^3 hybridized, which means that if we have a connectivity between the Cα and the Cβ followed by the Cγ, we can have rotations about these angles.

Now, these rotations give us angles of chi 1; that is a rotation about the Cα Cβ, a chi 2 that is around the Cβ Cγ and a Cγ Cδ is the chi 3 [refer to slide]. So this is our Cα and then we have the Cβ Cγ. This is where we have the nitrogen that is part of the backbone of the amino acid residue.

Now, when we look at these specific rotations, we just see the Cβ that is in turn connected to the Cα. So, depending upon the geometry there can be specific orientations possible, so that there are no steric clashes of the atoms coming together or the atoms bumping into each other.

So, this is where we have specific geometrical constraints depending upon the structural characteristics of the amino acid side chains, the chi 1, chi 2, chi 3 angles.

(Refer Slide Time: 15:09)

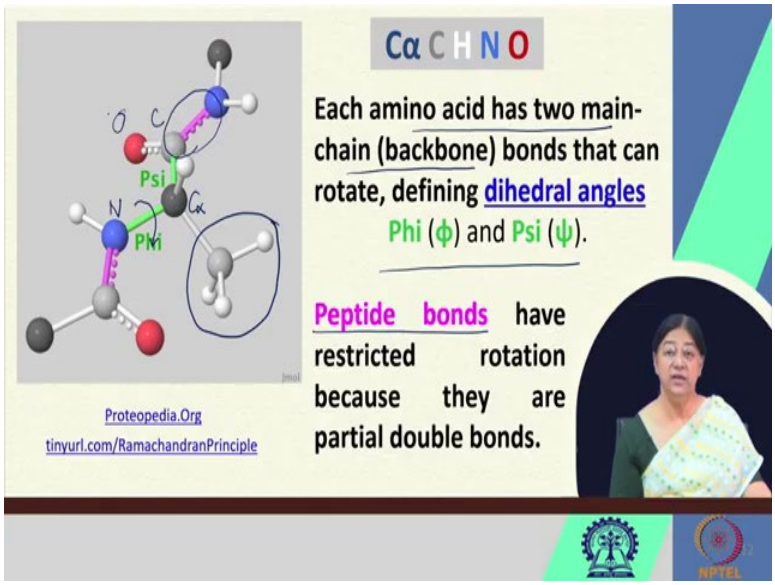


We need to go a step further in understanding the types of interactions to find the conformations of protein side chains with the repulsive steric interactions that are possible. This is what is important.

So, we have some that are backbone conformation dependent. We have these are the chi 1 chi 2, where we have these linkages that we were looking at. Then we have the backbone, these are independent of the backbone conformation, these are dependent on the backbone conformation, because this will comprise the φ, ψ and the chi 1, where C_{i-1} is the carbonyl carbon of the previous amino acid.

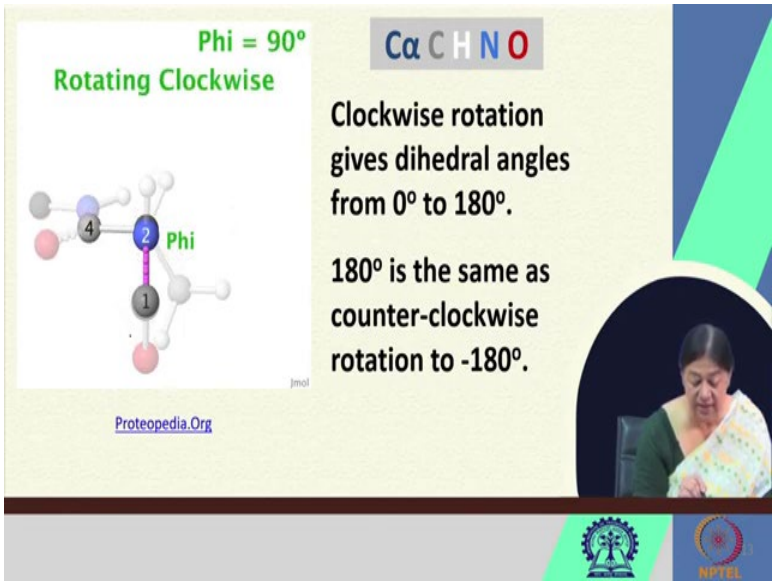
We will go into the geometry a bit more because we are looking at protein architecture. So, we have a specific model that defines this.

(Refer Slide Time: 16:07)

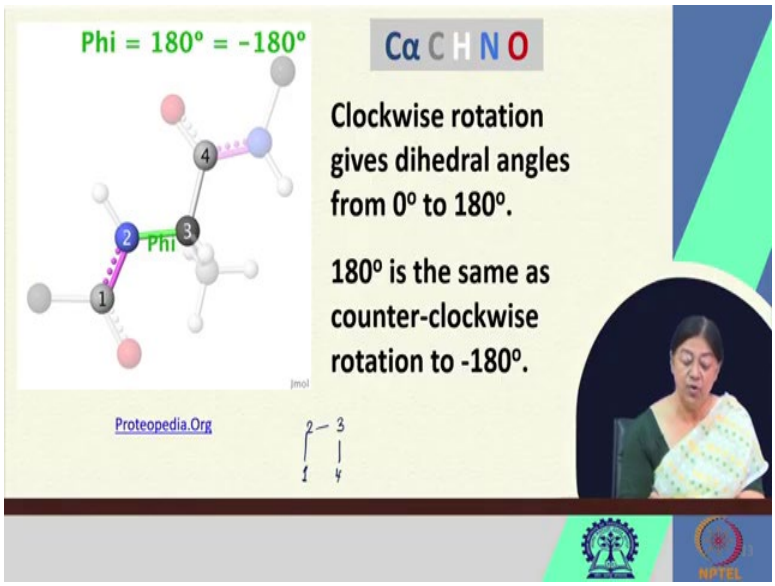


But when we look at the φ, ψ angles, we see that each of the amino acids has two main chain backbones defining dihedral angles φ and ψ and we know already that the peptide bonds have restricted rotation because of the partial double bond character, that is represented by a dotted line [refer to slide], where we have the CO and the NH. So, there is the Cα which is the side chain. Now, what happens is rotation about the bond, giving us the dihedral angles that are extremely important in the protein geometry. What we have is the Cα carbon, we have the C, there is the C = O, then there is the nitrogen and the oxygen. So, this is what is important in understanding the protein geometry.

(Refer Slide Time: 17:19)

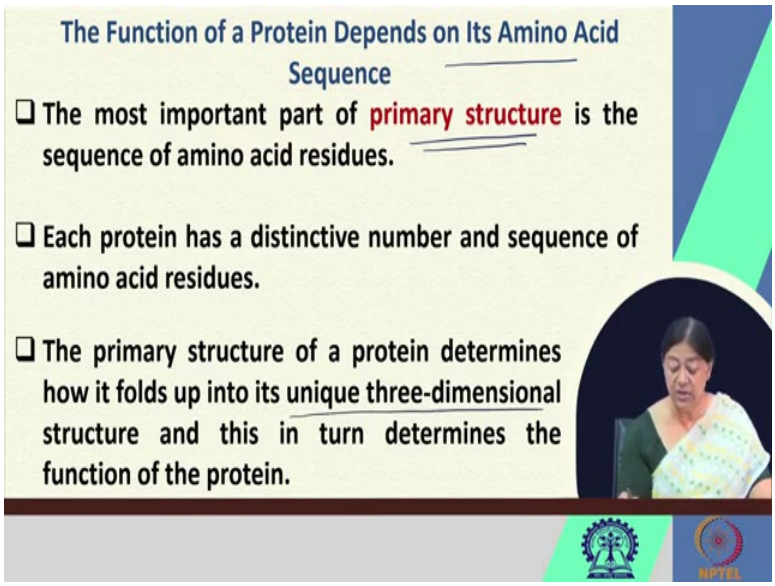


(Refer Slide Time: 17:42)



So, when we look at the φ angle, we see what the dihedral looks like. Now when we rotate this, we can see that we have a 90° φ angle. So, the first thing we see is that there are 4 atoms; 1 2 3 4 and by rotating them 1 and 4 clash. When looking along 2 and 3 we see a clockwise rotation that gives us 0 to 180. We can also have a anticlockwise rotation that will give us a -180. So, this gives us what is the φ angle; that is the angle between the 4 atoms that have to be defined are the Cα the C the N and the O. So, that is what is going to define our geometry for the φ angle.

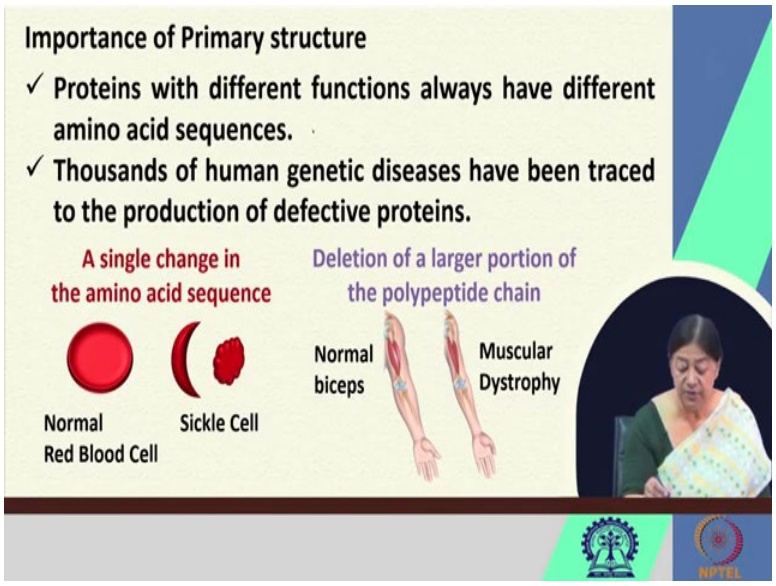
(Refer Slide Time: 18:28)



We similarly see a geometry for the ψ angle that is also going to give us the factor from the specific portions that we looked at in defining the backbone atoms. So, the function of a protein is going to depend upon its amino acid sequence and the most important part here is the primary structure, which is the link of the different amino acids one after the other in forming the polypeptide chain.

Each protein has a distinctive number and sequence of amino acid residues. And given that there is no limit to the number of amino acid residues that you can use in forming this polypeptide chain you realize that there can be many different proteins that can be formed. So, the primary structure of a protein depends on how it actually forms its unique three-dimensional structure that ultimately determines its function.

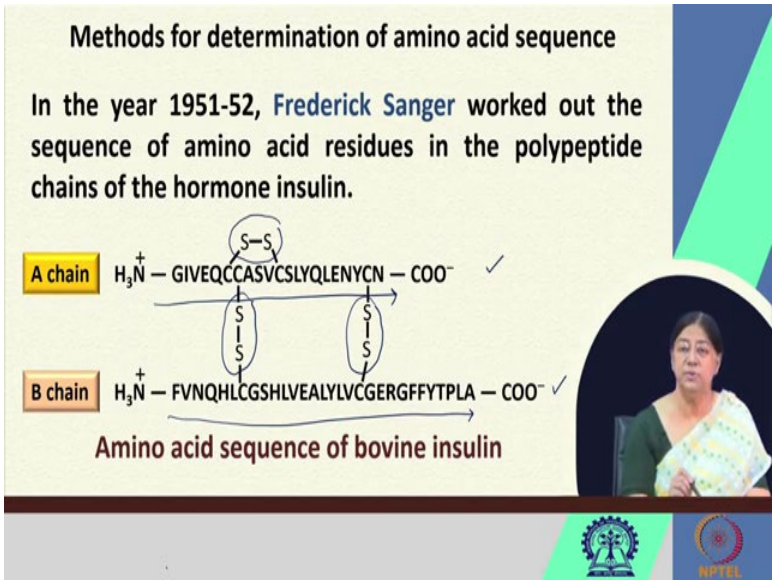
(Refer Slide Time: 19:36)



We know that proteins with different functions always have different amino acid sequences, but the overall structure sometimes is pretty much the same. We looked at how the variation in a single change of amino acid from glutamic acid to valine, could result from a normal red blood cell to a sickle cell depending on the change or the mutation in the type of amino acid that we are talking about, because we are changing the characteristics.

Similarly, when we have muscular dystrophy there is a deletion of a large portion of a specific polypeptide chain, leading to impaired function. So, we realize the importance of the primary structure, the importance of the sequence of amino acid residues in the specific way they form, because that is going to be dependent on how they fold in three-dimensional space.

(Refer Slide Time: 20:42)



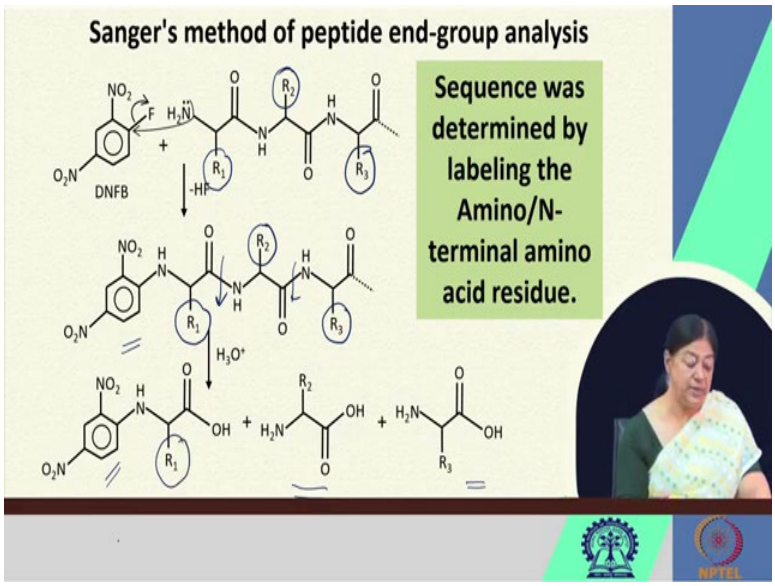
Now, if we want to look at the amino acid sequence, there are methodologies that help us in determining the sequence of the proteins. Now, if we look at insulin, this is the amino acid sequence of bovine insulin, and we find some specific features here. There are two chains in insulin; the A chain and the B chain.

[Refer to slide], we see we have a disulfide linkage in the A chain, that is an intra chain meaning it is in chain A itself. We have in addition two disulfide linkages that link chain A and chain B.

Now, if we want to understand or we want to know what the sequence of chain A is, what the sequence of chain B is, then there are specific methodologies that have to be followed in order for us to identify the specific amino acids that are linked in the specific order that forms what the A chain is or what the B chain is or for that matter whatever protein it is.

So, an understanding of that sequence is extremely important, because that is what defines its function.

(Refer Slide Time: 22:10)



So, just for a preliminary idea to understand Sanger's method for peptide end-group analysis, it is just a labelling. So, if we have a methodology by which we can label the N terminus of the polypeptide chain, and we can cleave it, means we can break it off from the rest of the polypeptide chain and have a methodology to analyze which amino acid has been broken off, then the N terminus can be found out.

So, this is the methodology that is commonly used. And we will see how we can actually use these for understanding the sequence. There are very interesting methods more like puzzles and riddles that actually tell us or give us an interesting way to find out what the sequence is all about. What do we want to know? We want to know what is R_1 what is R_2 and what is R_3.

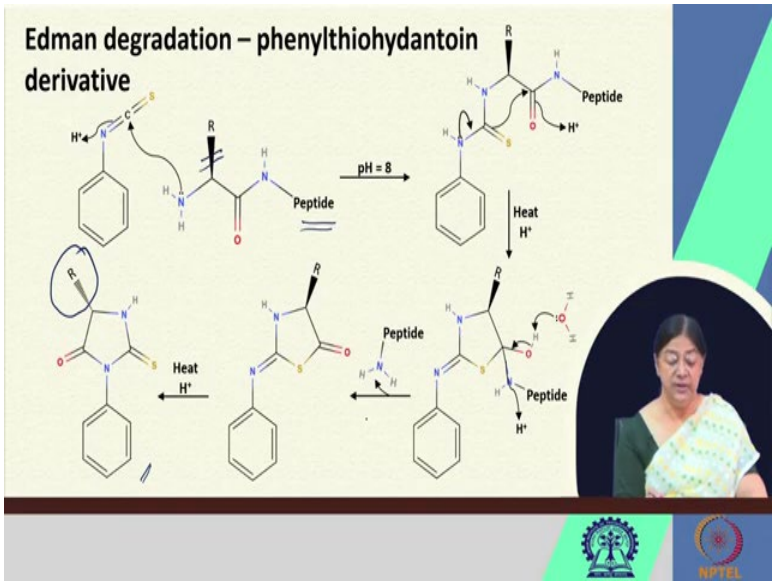
So, the method in this case is using a reagent dinitroflurobenzene and there is a specific reaction that occurs. After this reaction occurs, we have the formation of this adduct. Now, this means

that there is the R_1, R_2, R_3 in the polypeptide chain and we now have this product formed from the reaction with D N F B.

Now, when we have hydrolysis what happens is, the hydrolysis results in the breakage of this bond. Like we looked at last time and we have condensation, we have the formation of the peptide bond, when we have hydrolysis, we can have the disruption of the peptide bond. Now, when we have the disruption of the peptide bond what we are going to see is, we are going to see a series of amino acids.

But the information that we get is we have to look for which amino acid is staged along with the D N F B? And this would be amino acid number 1. So, this is one method in which we can determine this.

(Refer Slide Time: 24:25)

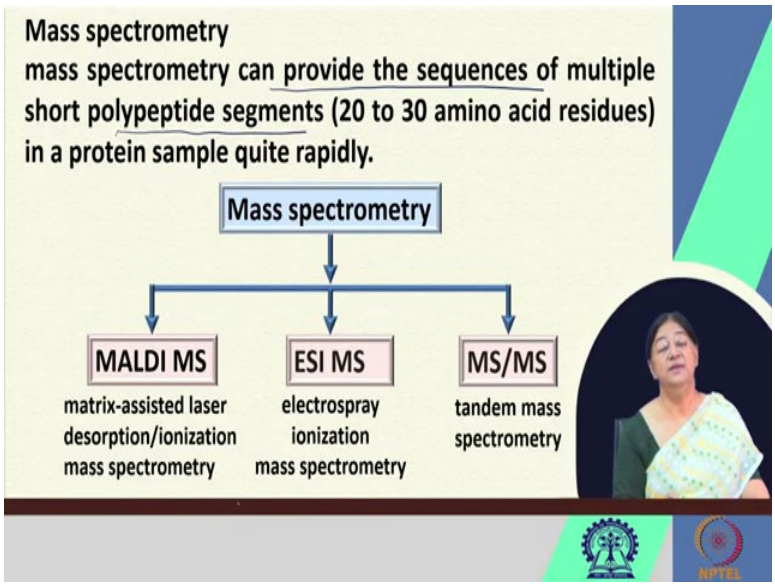


Another method is known as Edman degradation which forms what is called the phenylthiohydantoin derivative. So, here we have the peptide, here we have a specific reagent that comes into the picture where we have a reaction. This reaction occurs at a specific pH, giving us again a derivative that is known as a phenylthiohydantoin derivative.

And this reaction then gives us the peptide along with the first amino acid of the sequence, and this is important in an understanding of what the identifier is for the first amino acid, because this is the rest of the peptide. Now, the interesting part here is, that we can now subject the rest of the peptide to the same reaction, as now we have a new N terminus. So, we can find out what that N terminus is, based on the specific sequence of the reaction.

So, in this degradation we can find out the N and know what amino acid is there, what R group is present there. So, this is how the specific reactions occur in the Sanger's method and the Edman degradation. We will look at specific methods to find out the sequence of polypeptides in a further lecture later on.

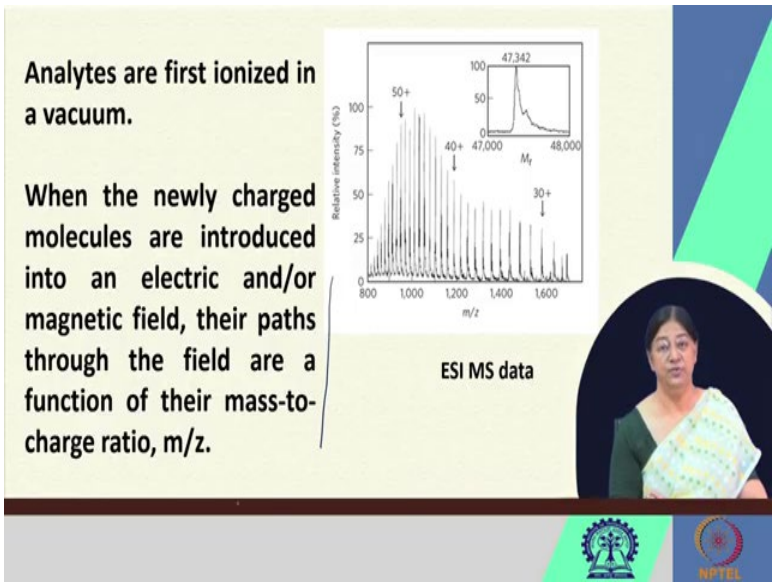
(Refer Slide Time: 26:04)



Another way in which we can find out this, is using mass spectrometry, where we can provide the sequences of multiple short polypeptide segments in a protein sample quite rapidly, where we have actually a fragmentation.

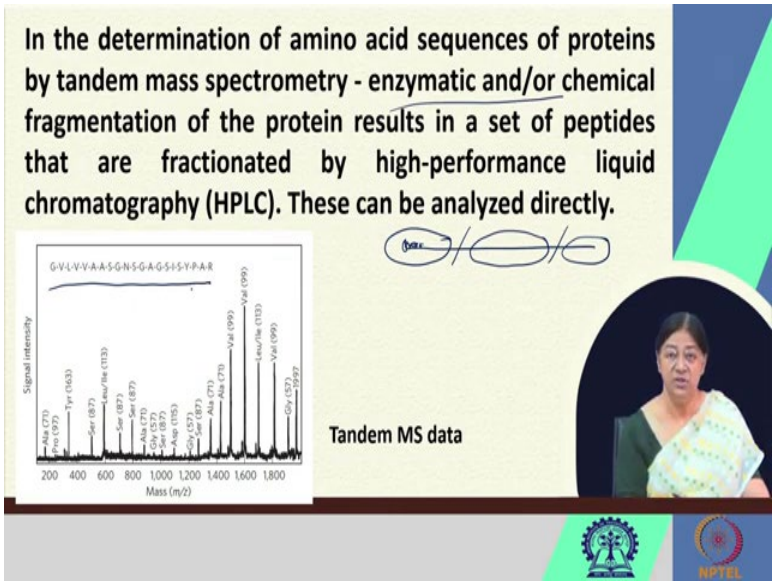
What happens in this case is in mass spectrometry, we have different methods, we have MALDI MS that is matrix assistant laser disruption, we have ESI electro spray and we have also MS/MS. Now, these are different mass spectrometry methods that tell us or give us an idea as to how we can determine the sequence.

(Refer Slide Time: 26:50)



So, in this case the analytes are ionized in vacuum. So, we have different fragments that are formed, and these fragments are now charged. Now, because of the newly charged molecules they are put into an electric field and they migrate along this electric and magnetic field because of their mass to charge ratio. So, this is what we see in an electro spray ionization MS data; a fragmentation of the polypeptide chain.

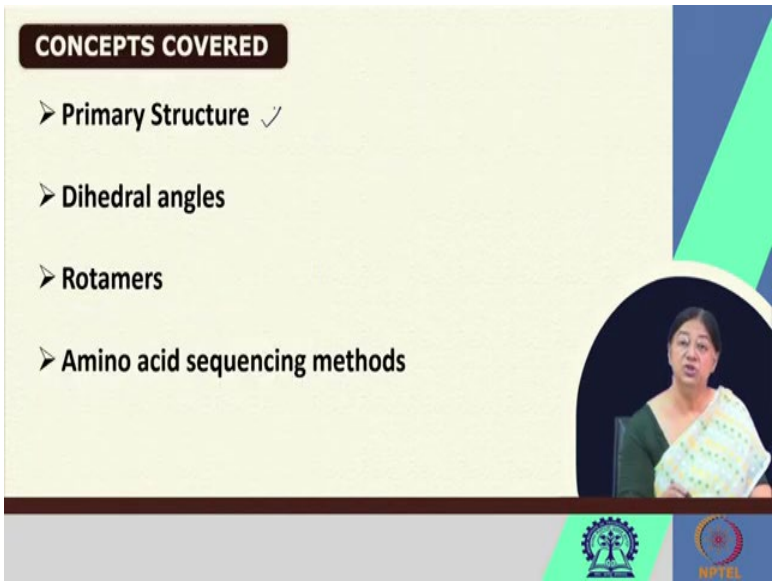
(Refer Slide Time: 27:25)



Similarly, if we look at a tandem mass spectroscopy we have enzyme and chemical fragmentation, which means that if we have a sequence of amino acids, we can break it off enzymatically and then we analyze the different fragments that we have.

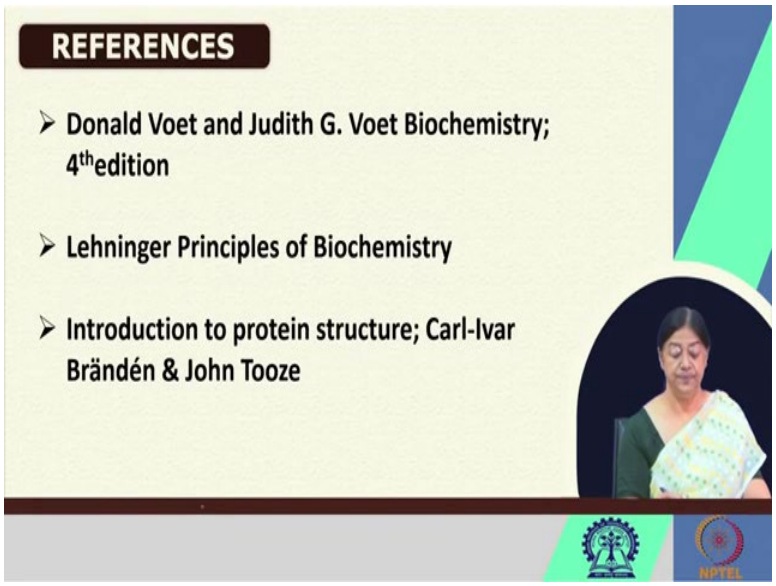
And this analysis can be done through tandem mass data or tandem mass spectrometry, where we can look at a specific sequence of the polypeptide chain and that is extremely important in an understanding of the primary structure of the protein.

(Refer Slide Time: 28:05)



So, we looked at what the primary structure of the protein is, the sequence of amino acids, the dihedral angles and the rotamers. Dihedral angles the backbone angles of amino acids, which we will revisit when we look at the secondary structure, the rotamers for the amino acids themselves and specific amino acids sequencing methods, giving us an idea of what the primary structure of the protein is.

(Refer Slide Time: 28:31)



Thank you.