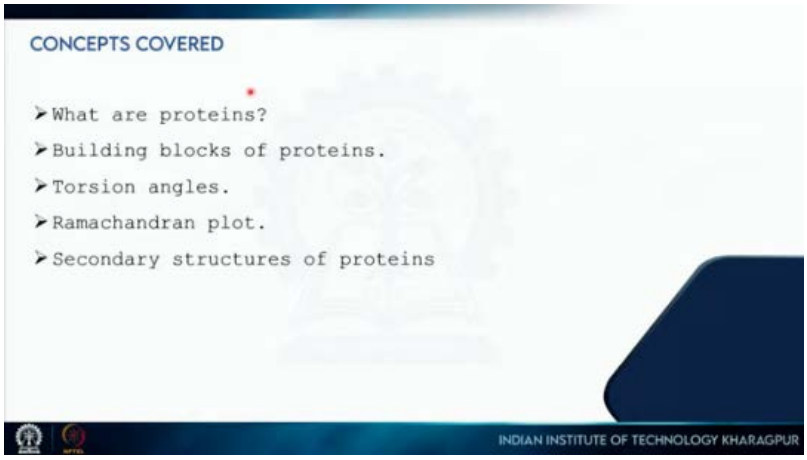
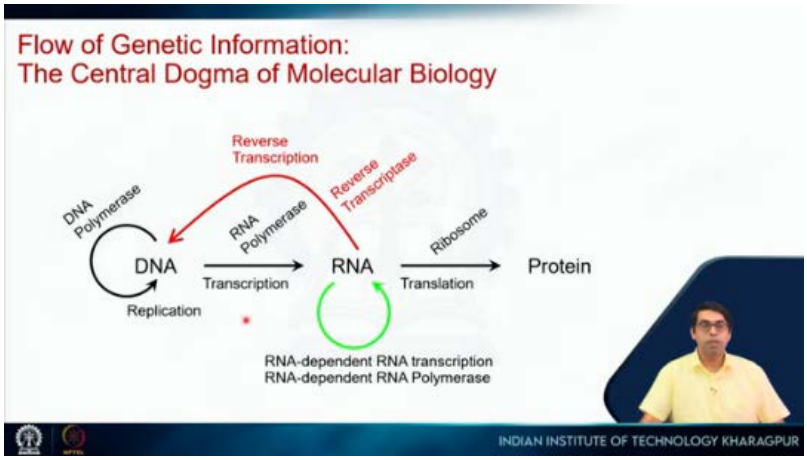
Introduction to Complex Biological Systems
Professor Dibyendu Samanta and Professor Soumya De
Department of Bioscience and Biotechnology
Indian Institute of Technology, Kharagpur
Lecture 11
Amino acids, hierarchy of protein structure

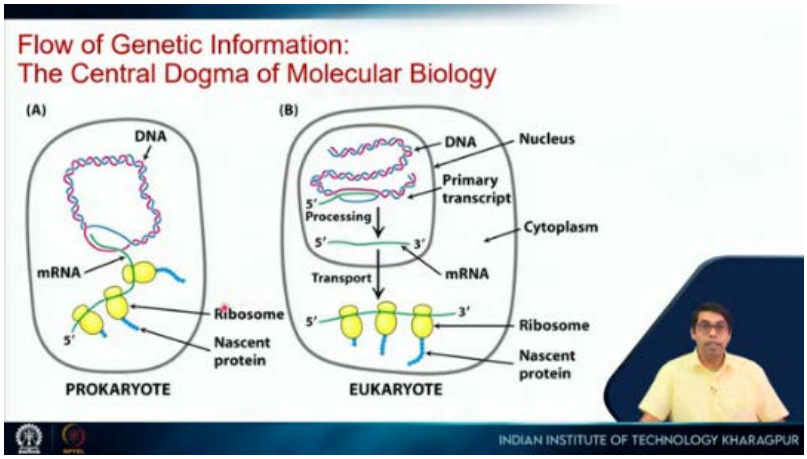
Welcome, students. So, I am Professor Soumya De from IIT Kharagpur. Today, I am going to talk about proteins, which are the nanomachines of any living system. So, this is Module 3, and I am going to give you the basic idea about what proteins are, what the building blocks of proteins are because proteins are, after all, polymers.



Then, I will go into the concepts of torsion angles, the Ramachandran plot, and the secondary structures of proteins. So, you have already seen this. This is the central dogma of molecular biology. This states that DNA is the molecule that stores all the genetic information. And from that genetic information, part of that information is copied into messenger RNA, which is used to ultimately make proteins.



So, DNA is the molecule that stores information, and proteins are the molecules that carry out all the functions that happen inside a living organism. And then, of course, there are all these events that also happen; you can copy RNA to RNA or also copy DNA from RNA. So, today we are going to focus mostly on proteins. Again, this is a brief recapitulation of what you have already seen that protein synthesis can happen inside the cell in one place in prokaryotes or bacteria, or it can be compartmentalized in eukaryotes. So, in bacteria, the synthesis of messenger RNA and the synthesis of protein occur at the same time, whereas in eukaryotes, it happens in two different places.



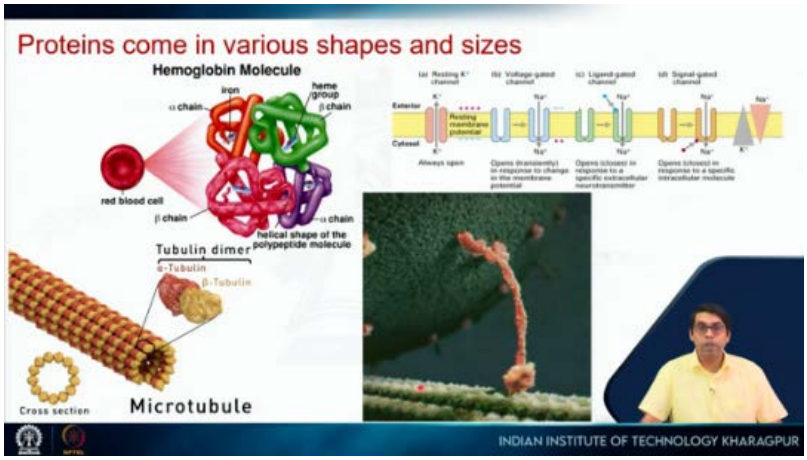
The mature mRNA goes out of the nucleus, and then the ribosomes bind, and proteins are synthesized. So, it turns out that proteins come in various shapes and sizes. So, any function that you can think of is carried out by proteins. For example, if we think about respiration, we are breathing; we are breathing in air.

And what happens is that in our lungs, the oxygen is captured and then it is transported to different parts of our body. And that is done by a special protein called hemoglobin. We will see hemoglobin in more detail in one of the lectures this week, most probably in lecture 4. So hemoglobin is a carrier protein. It binds to oxygen.

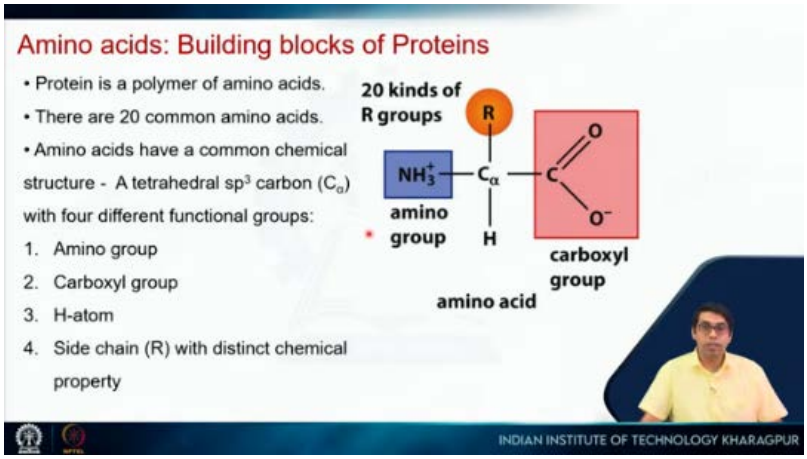
It carries it to different cells inside our body. Similarly, there are all these different proteins which are called ion channels. So, ion channels are very good and very important for our sensation. There you can see several ion channels which are shown here, and these proteins are embedded on the cell membranes. So, we will talk about cells in more detail later, but briefly, all cells have a covering which is called a cell membrane.

Ions cannot pass through this cell membrane, so they need a special type of transporter called ion channels. So you can see there is an ion channel for potassium, there is an ion channel that lets sodium go through, and then there are also ion channels for calcium and other ions, like chloride ions, etc. And these ion channels are very specific. For example, the ion channel which allows sodium to pass through will not allow potassium to pass through it. And we will see that these are all because of proteins, and they can have very specific shapes and very specific functions. Then, there are proteins like this.

So, this is called a microtubule, and it is formed of proteins which can form a polymer. So, protein itself is a polymer, but then these proteins themselves come together and form these huge structures, and microtubules are responsible for various things that happen inside the cell. They are responsible for the cell shape, and they are also responsible for the transport of different molecules from different parts of the cell. So, you can think of what you see here below as a microtubule. On that, what you see here, which looks like something is walking on it, is another protein called dynein. This protein walks on the microtubule and carries this huge cargo called the vesicle. There can be different things inside that vesicle, and it carries them from one part of the cell to another part of the cell. So, this is very similar to our railroad system.

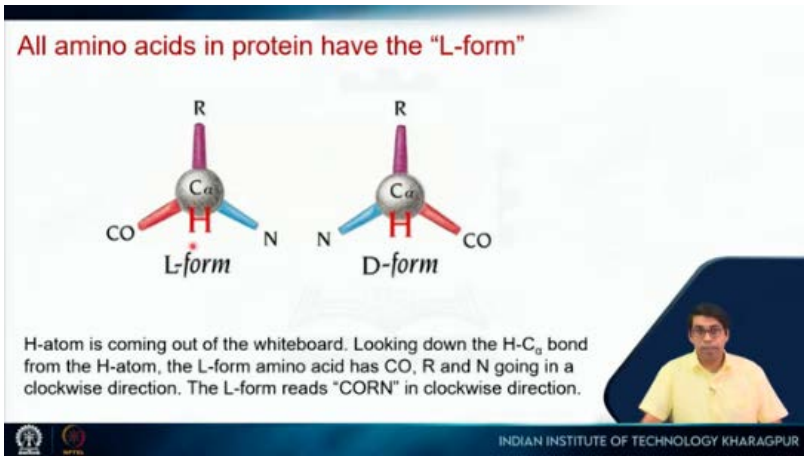


So you can see that proteins can act as so many different things. And then, of course, in the next week, I will talk about a specific type of protein, which are called enzymes, which carry out the catalysis of all the chemical reactions that happen inside our body. So, proteins come in various shapes and sizes. But if we go and look deeper, we will see that all these proteins are nothing but polymers, and these polymers are made up of monomers which are called amino acids. So, this is the basic structure of the building blocks of proteins, which are called amino acids.



It is called an amino acid because it has an amino group and an acidic group. So, there is this alpha carbon, and we know that when carbon is sp3 hybridized, it has four hands; it can bind to four things. So, one is the amino group, the other one is the carboxyl group, the third one is a hydrogen, and the fourth one is referred to as an R group, which can be different. It turns out that for common proteins, there are 20 different types of amino acids, and they have these different R groups. These amino acids also have a particular

stereochemistry since there are four different groups connected to this carbon. This will have a particular stereochemistry, and we can have two different forms, one is referred to as the L form, and the other one is referred to as the D form. So, this L form and D form are actually mirror images of each other. It turns out that all natural proteins are made up of amino acids which are composed of the L -form.



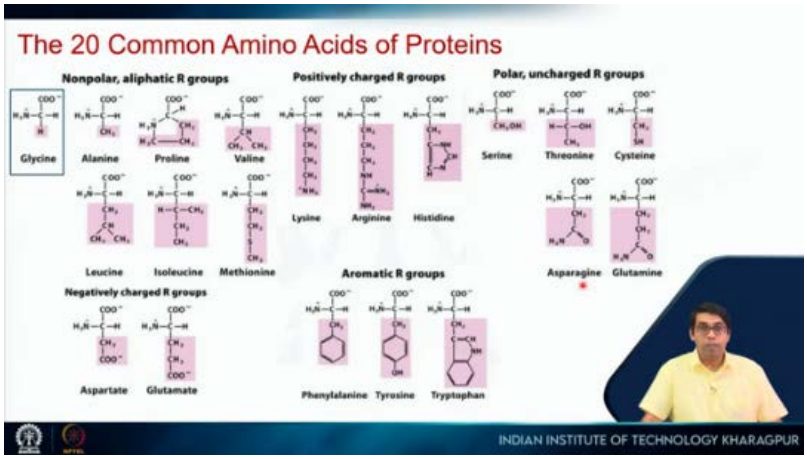
So, all natural proteins are made up of the L-form of amino acids. There are, of course, exceptions. We will see that as we progress through this course that in certain cases, D-form amino acids are used, but they have very specific functions. So, how do you know which one of these two is the L-form? If we write the structure of the amino acid in this particular manner, where H is coming out of the plane of the screen, then if we go clockwise, it will spell C O R N or CORN. So, C O stands for the carboxyl group, R is for the side chain, and N stands for the amino group.

So, as I said, there are 20 common amino acids of proteins, and they can be grouped into different groups depending on the nature of the side chain. The simplest form of protein or amino acid is glycine. So, in glycine, the R group is hydrogen. So, all the R side chains are highlighted here in pink or purple. So, this hydrogen, since the R is hydrogen, glycine has two hydrogens.

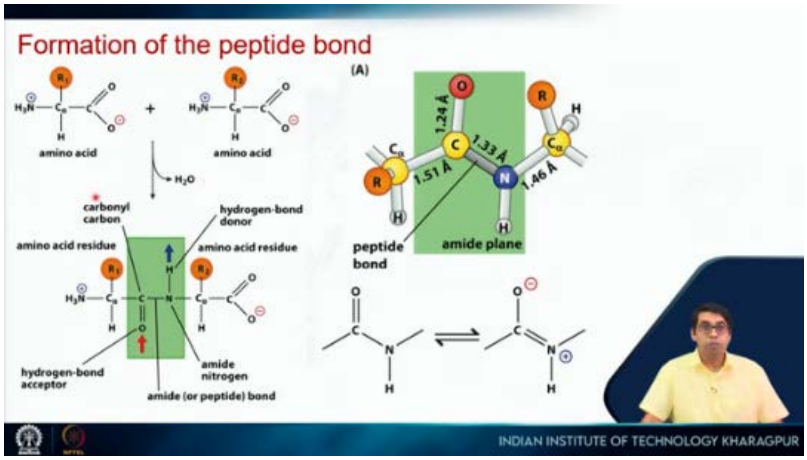
So, it is the only amino acid which is not chiral. And then there are all these different amino acids. You will see that they are not charged, and these are mostly aliphatic. So, these are nonpolar aliphatic amino acids. Then there are amino acids which have a positive charge, and these are also basic amino acids. Examples are lysine, arginine, and histidine. Then

there are amino acids which have a negative charge, so these are acidic amino acids. There is aspartic acid and glutamic acid. We have polar amino acids, so these are not charged, but they have polar groups like this OH group, SH group, or NH_2 CO NH_2 groups.

And finally, we have aromatic amino acids where the side chain has these aromatic groups. So, phenylalanine, tyrosine, and tryptophan are examples of aromatic amino acids. So, these different groups will become useful when I start talking about protein folding, and we will see that these aromatic and aliphatic groups are the ones which like to or tend to interact with each other, and these charged amino acids which they tend to interact with water and also with each other. Also, the polar groups tend to interact with water, and all of these things become very important when I talk about protein folding.



So, we have the monomeric units which are amino acids. These two amino acids come together to form a polypeptide. So, if I take two amino acids, then the carboxylic group and the amino group will react. So, this is a condensation reaction. A water molecule goes out, and this bond is formed, which is called the amide bond, but in this case, we will call it a peptide bond. You have already seen this during protein synthesis. This is the steady capitulation of that.

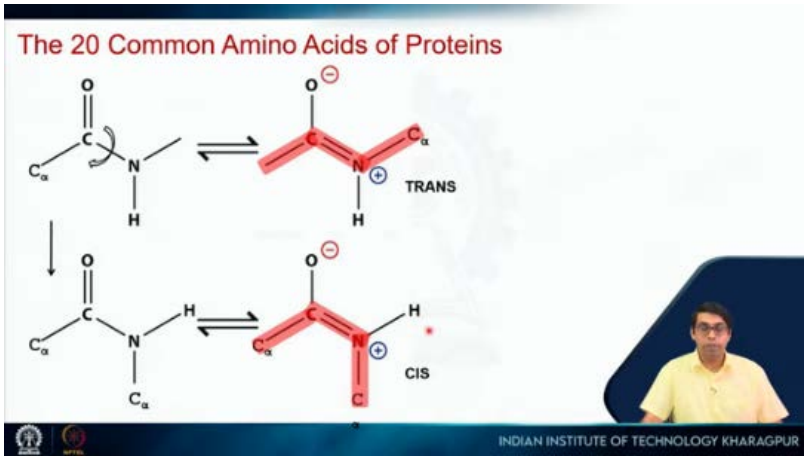


Now, if we focus on this peptide bond, it turns out that it can have this partial double bond character because the lone pair of electrons which are on this oxygen can—sorry, the lone pair of electrons which are on this nitrogen can go here, and this can open up, resulting in a partial positive charge here and a negative charge here and a double bond here. So, this resonating structure results in a partial double bond character of this peptide bond. So, what are the consequences of that? This actually has a very important consequence in terms of protein structure and folding. So, let us look at that again. This is the partial double bond character. So, if this is a partial double bond, it means that this carbon is sp2 hybridized.

So, all these four atoms are in the same plane. This nitrogen is also sp2 hybridized, which means that this carbon, this C alpha (α), and this hydrogen are also on the same plane. So, it turns out that all these six atoms are on the same plane. So, this is a planar structure. But when it is in this form, the nitrogen is sp3 hybridized.

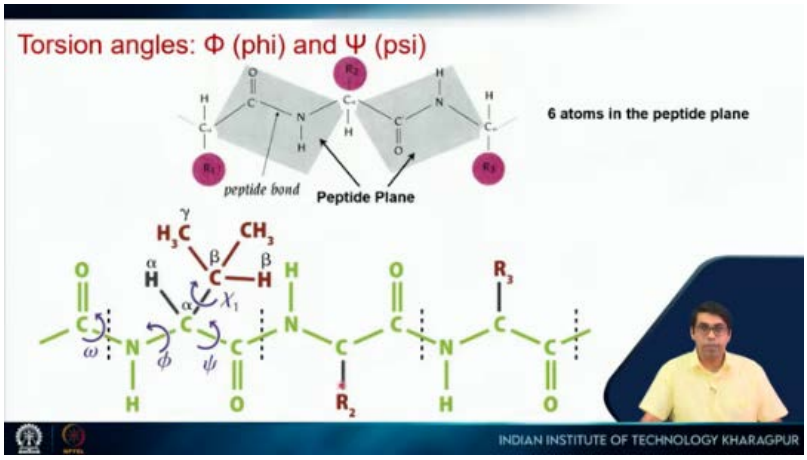
So, this is a single bond. And we can have rotation about this single bond. So, let us say it rotates like this. So, what I have done is I have flipped these two groups. So, the C alpha(α) has come here, and the H has gone there.

So, it looks like this. And again, it can form this resonance structure. We can see that there are these two different structures which are possible, where the two C alpha(α) groups are on the opposite side of this peptide bond, and the two C alpha(α) groups here are on the same side of the peptide bond. These two are referred to as the trans form and the cis form. So, the peptide bond can exist either as a trans peptide bond or a cis peptide bond.



So again, if we look at a polypeptide chain. So, this is one amino acid; it is connected to one amino acid here and another amino acid here. So, this peptide bond will form one peptide plane, and this peptide bond will form another peptide plane. Just a reminder that there are six atoms in the peptide plane. So, these six atoms are the C alpha(α), carbonyl carbon, carbonyl oxygen, nitrogen, another C alpha(α) of the next amino acid and the hydrogen. So, together they form this peptide plane. Now, if we look at a long chain of amino acids which are connected to each other, they will look like this. So, you see this is one amino acid. It is connected to the next one.

This is the second amino acid. It is connected to the third one. So, this is another amino acid. And this dotted line shows the peptide bonds. So, we have already seen that we cannot have rotation about this; we can have this in two forms, either the cis or the trans form, but what about the other bonds which show up here?

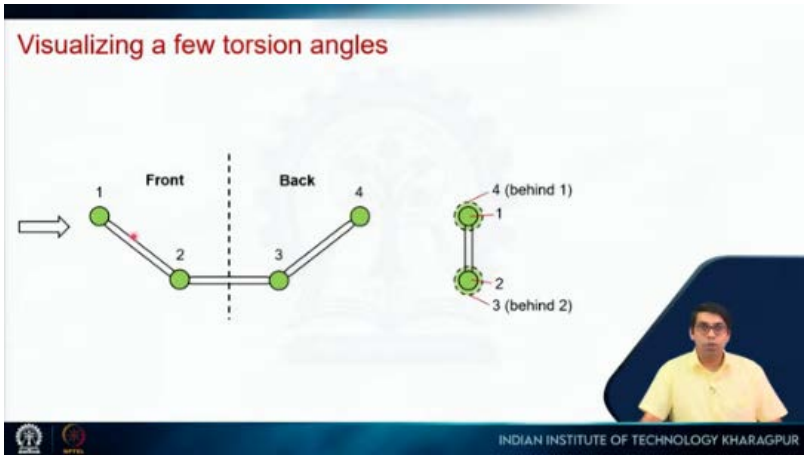


So, if you see, there are three types of bonds which show up in regularity. This peptide bond, then this nitrogen to C alpha(α), C alpha(α) to carbonyl. Again, the peptide bond, nitrogen to C alpha(α), C alpha(α) to carbonyl, right? So, these three bonds keep repeating, and we can have rotations about these two.

These are not partial double bonds. So, we can have free rotation about these two. The rotation about this bond, this nitrogen C alpha(α), defines the phi (Φ) torsion angle, and the C alpha(α) to carbonyl carbon defines the psi (Ψ) torsion angle. So, what are torsion angles? Let us look at this.

Normally, we define an angle in plane geometry between three points, but this is not in plane geometry. This will be a solid geometry definition. So let us think of four atoms, 1, 2, 3, and 4, which are connected by these three bonds. So if all these four atoms are on the plane of your screen, it will look like this. Let us say we are looking from this direction.

So if we look from this direction, then they will look something like this. So you will see these two atoms, 1 and 2, in the front. And these two atoms, 3 and 4, in the back. So, I have drawn these 3 and 4 in dotted lines because you will not see them; they are of the exact same size, as I have drawn here. So, they are all on the same plane.

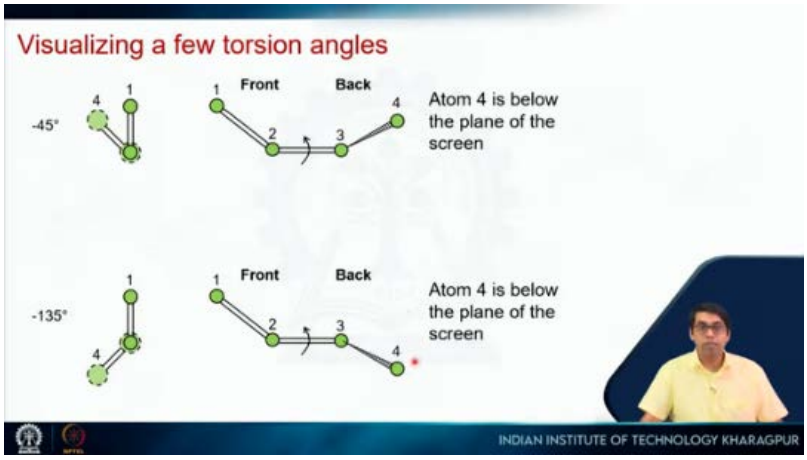


So, I have drawn it again, and this particular configuration will be defined as 0 degrees, where 1 and 4 are exactly behind each other. Now, if I rotate about this bond, which is between atoms 2 and 3, so if I hold this and pull it out towards you, or in other words, if I hold this and rotate it towards the right or clockwise, then rotating it in this direction defines

a positive angle. So it goes from 0 to 90 to 45 degrees. So if I am seeing in this view, it will look like this; this is 45 degrees. And if I see in this view, this 4, the fourth atom, will come out of the screen, or the plane of the screen, towards you, and that is shown as this solid wedge. Similarly, if I rotate in the opposite direction, then this defines negative angles. So, this is 0, this will be minus 90, and this will be minus 180, right?

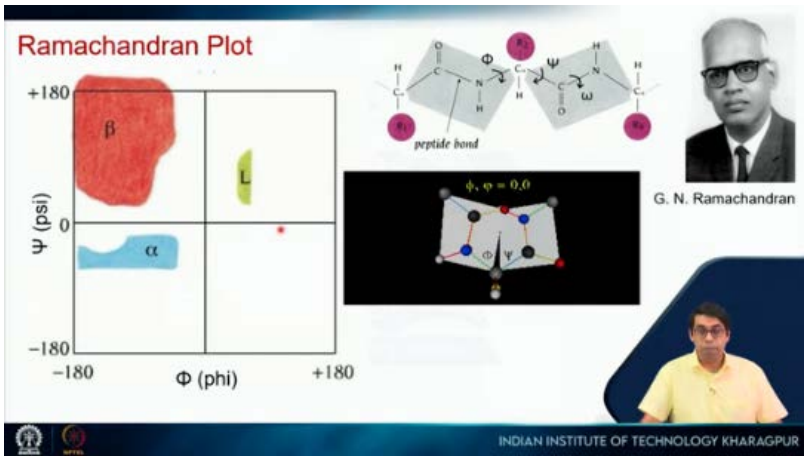
So, if I rotate it like this, this is minus 45, and in this case, it will go towards the back side of the plane of the screen. So now think about this. So you can pause the video here and think about how this minus 135 degrees will look like. Will it be on the left-hand side or the right-hand side?

Will it be above or below? And here, will it be above or below, or will it be in front of the screen or behind the screen? So if you have seen, if you think about something like this, then you are right. So a minus 135-degree rotation will be towards the left-hand side. So this is 90, so it goes below, and in this case, it will be behind the plane of the screen. Now it turns out that Ramachandran So, he made actual models of dipeptides, and he thought that, if I rotate about these phi (Φ) and psi (Ψ) torsion angles, are all torsion angles possible?

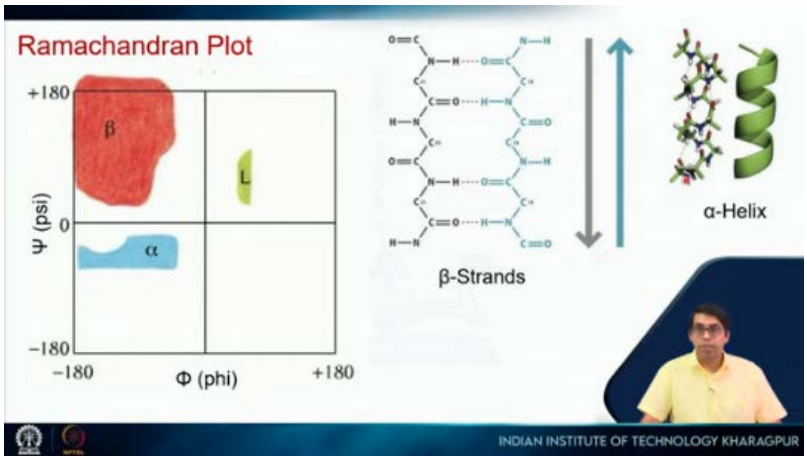


And it turns out that not all torsion angles are possible because there will be steric clash. So, you can see this GIF image, and you can see that at some point, the carbonyl This oxygen will clash with this hydrogen, right? So, it means that this 0° angle will not be possible. So, to represent this, what Ramachandran did was he plotted the phi(Φ) and psi (Ψ) torsion angles in a plot, which is now famously called the Ramachandran plot.

And when he plotted this, he found that there are only certain regions which are allowed in this. So, it means that in his model, there were all these empty spaces. So, these particular regions were not even sampled because of steric clash. And now it turns out, because we have solved a lot of protein structures, we see that this is exactly true. Indeed, proteins occupy only these regions, and it also turns out that when a polypeptide chain, a series of amino acids, has this particular torsion angle, it has a particular shape.

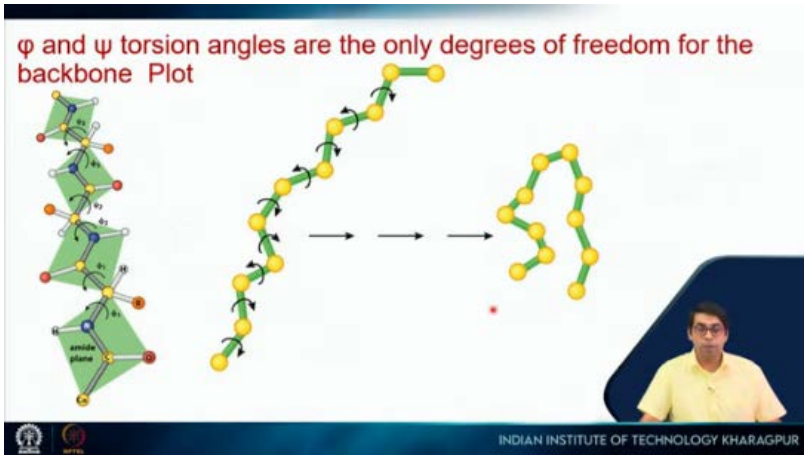


When they have these torsion angles, they have a different type of shape, and those shapes are called alpha(α) helix and beta strands. So, when consecutive amino acids have torsion angles in this region, they form this spiral-like structure, which is called an alpha(α) helix. And when consecutive amino acids have torsion angles in this region, they will have an extended structure, which forms beta strands. And then, of course, there can be loops, which can show up in other places. So, these alpha(α) helices and beta strands turn out to be something called the secondary structures of proteins. So, how do proteins fold?

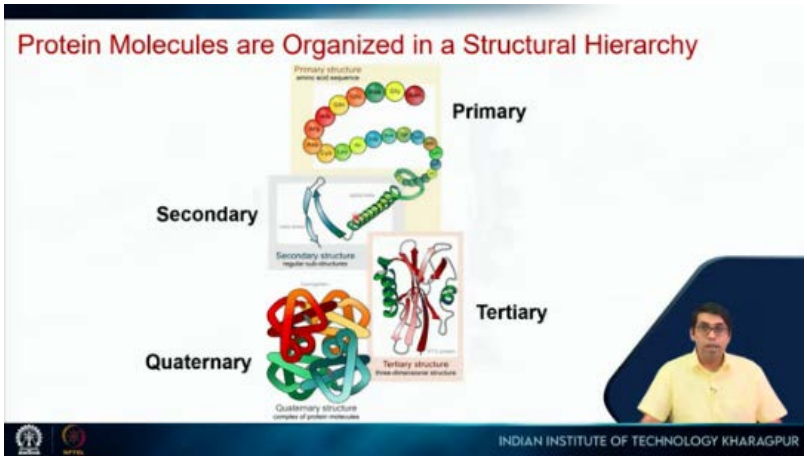


So if I just draw the extended structure of a protein, it will look something like this. Now you can see that again there is this peptide bond. So where is the peptide bond? So there is the peptide bond. And then we have the alpha(α), the phi (Φ), and psi (Ψ) torsion angles.

So the peptide bond is here. Phi(Φ) and psi(Ψ) torsion angles. And we have already seen that you cannot have rotation about the peptide bond, but we can have rotation about the phi (Φ) and psi (Ψ) torsion angles. So if we keep on rotating about this, this extended chain can collapse into something like this, which will be our folded protein structure. So I will talk about protein folding in more detail in the next lecture.

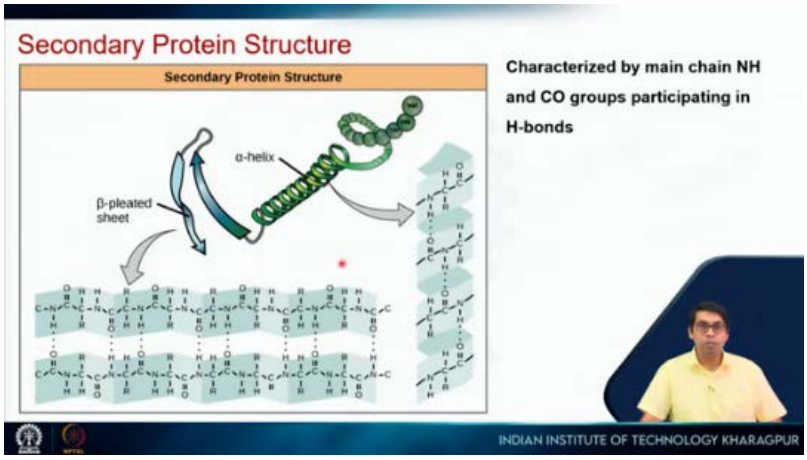


So, coming back to these protein structures or the hierarchy of protein structure, it turns out that protein structure can be divided into four levels. The sequence in which amino acids are connected to each other is called the primary structure or the primary sequence of the protein. These sequences can fold into helices or strands, which are called the secondary structure of the protein. And these helices and strands can fold and form tertiary contacts. So this is called the tertiary structure of a protein.



So there are three-dimensional contacts between them, and it forms a tertiary structure. So most proteins will have primary, secondary, and tertiary structures. Some proteins do not stop here. They can have another higher-order organization where multiple polypeptide chains can come together and form a quaternary structure. So the structure that you see here is the structure of hemoglobin, where four polypeptide chains come together to form a quaternary structure.

So let's look at the secondary structures in more detail. So I told you that there is one structure called the alpha(α) helix. So when consecutive amino acids have a particular torsion angle, they form an alpha(α) helix, and the other one is a beta strand. So these secondary structures are characterized by the torsion angles as well as the hydrogen bonding pattern between the main chain amide NH and the carbonyl group. So, there is a particular pattern of this hydrogen bond which stabilizes these structures.



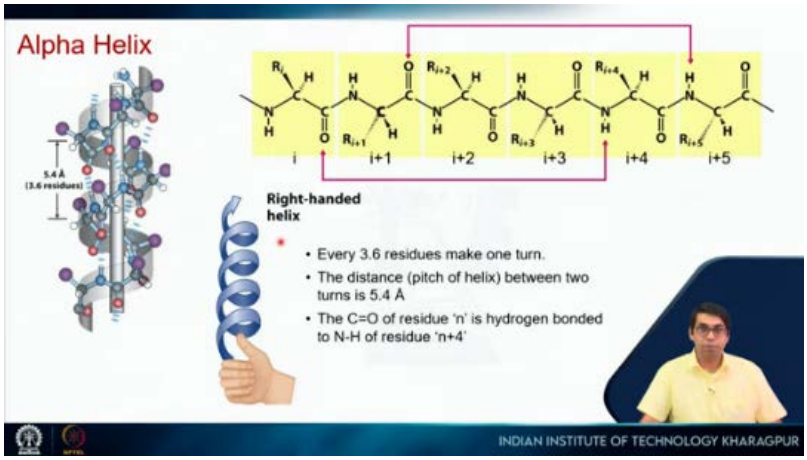
So, let us look at the alpha(α) helix first. So, if I draw this peptide chain where I have amino acids i, i+1, i+2, i+3, i+4, and i+5, this folds into a helical structure like this and it forms a particular pattern of hydrogen bonds. The pattern is such that the carbonyl oxygen of the ith residue forms a hydrogen bond with the amide hydrogen of the i+4 residue. So, you see this NH will form a hydrogen bond with this CO. Similarly, the next amino acid NH will form a hydrogen bond with the next amino acid CO.

So, i with i+4. i+1 with i+5, i+2 with i+6, and so on. So, there is always a difference of 4 amino acids, and you can see that. So, there is this amide NH. So, the blue is the nitrogen, and this small red sphere, white sphere, is the hydrogen.

Similarly, these carbons are shown here. So, this is the carbonyl oxygen, and this is the carbonyl carbon. So, C double bond O is here, NH is here, and they form a hydrogen bond here. Now, it turns out that all these hydrogen bonds are oriented in the same direction. So, you can see that all the NH groups are pointing up and all the CO groups are pointing down, and this has a consequence which we will see in the next slide.

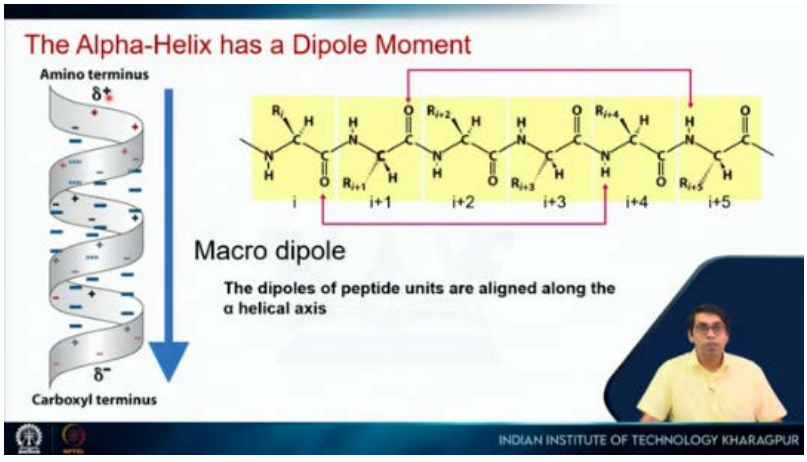
So, here are some of the characteristics of a hydrogen bond. So, I told you that the pattern is i+4 or n+4. If we go like this, then 3.6 residues make 1 turn. So, in 1 turn, there is not an integral number of amino acids. So, it is not 3, it is not 4, but it is 3.6.

And the distance or the pitch of the helix between 2 turns is 5.4 angstroms. So, if I measure it from here to here, it will be 5.4 angstroms. And these are all right-handed helices. So, most alpha(α) helices are right-handed helices. Some can also be left-handed helices, and they occupy a different region in the Ramachandran plot.



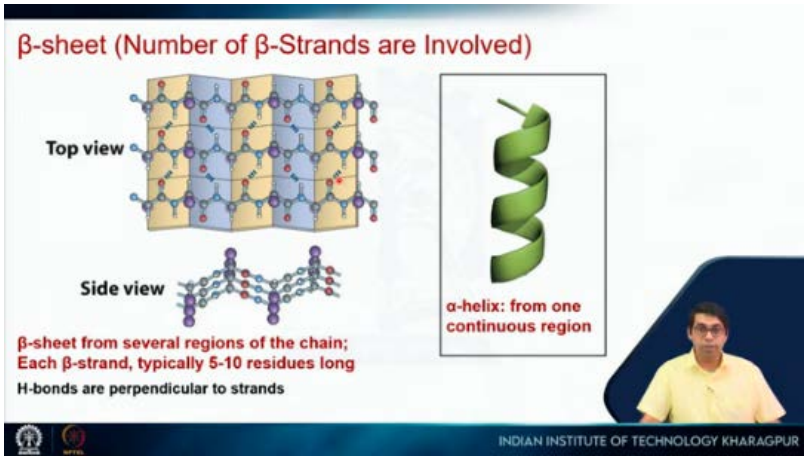
So, what is the consequence of this orientation of the amide and the CO groups? So, all the amide groups are pointed up, and all the CO groups are pointed down, and we can think of this amide group as a small dipole. So, nitrogen will have a partial negative charge, and hydrogen will have a partial positive charge. So, negative positive pointing in this direction. Similarly, oxygen will have a partial negative charge, and carbon will have a partial positive charge.

So, it is also oriented in the same direction. So, within the helix, all these partial negative and positive charges can cancel out each other, but towards the end of the helices, towards the amino terminus, there will be three carbonyl groups. So, there will be these three positive charges; these are not satisfied by hydrogen bonds; these are partial positive charges, and towards the carboxyl terminus, there will be these partial negative charges. Coming from the oxygen, and these partial positive charges are coming from this hydrogen. So, there is a net positive charge here, and there will be a net negative charge here.



But again, it is not a formal positive or negative charge. These values can be roughly around 0.5 of a formal charge of a proton and 0.5 of an electron. However, these charges are separated by a long distance, which is the length of the helix. So, this results in something called a macro dipole. So, the helices will have this macro dipole, and these become important not only in protein folding but also in the interaction of these proteins with other molecules.

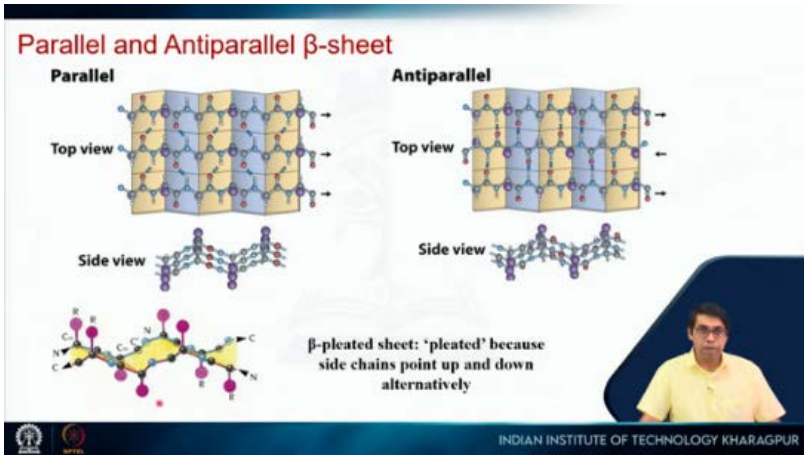
And we will see examples of this in the coming lectures. So, most helices are right-handed, but they can also be left-handed, and these two occupy two different regions in the Ramachandran plot. The other one is the beta strand. In the beta strand, two strands come together to form hydrogen bonds. In this case, hydrogen bonds are not within the same strand but are inter-strand hydrogen bonds. In a helix, all the hydrogen bonds are within the same helix, but in beta strands, the hydrogen bonds are within two beta strands. And these two beta strands can have these two orientations. So, they can be parallel, where all these strands have the same orientation.



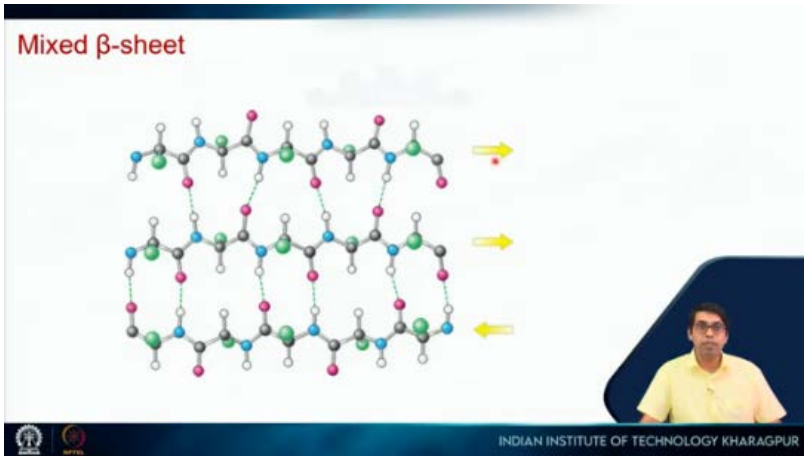
So, N-terminus on the left-hand side, C-terminus on the right-hand side, N-terminus on the left-hand side, C-terminus on the right-hand side, and they can be antiparallel. So, N-terminus here, C-terminus here, now N-terminus here, C-terminus here. So, they all form hydrogen bonds, but the pattern of the hydrogen bonds is slightly different. So, in this case, you can see that all four atoms in the hydrogen bond are collinear. So, they form a straight line, but in this case, they are not collinear.

So, you see this and this are not in the same straight line. So, if we count the number of hydrogen bonds, they will be exactly the same per amino acid. The dipole-dipole interaction here is collinear. These hydrogen bonds are slightly more stable energetically than these hydrogen bonds. But we see both parallel and antiparallel beta strands in proteins.

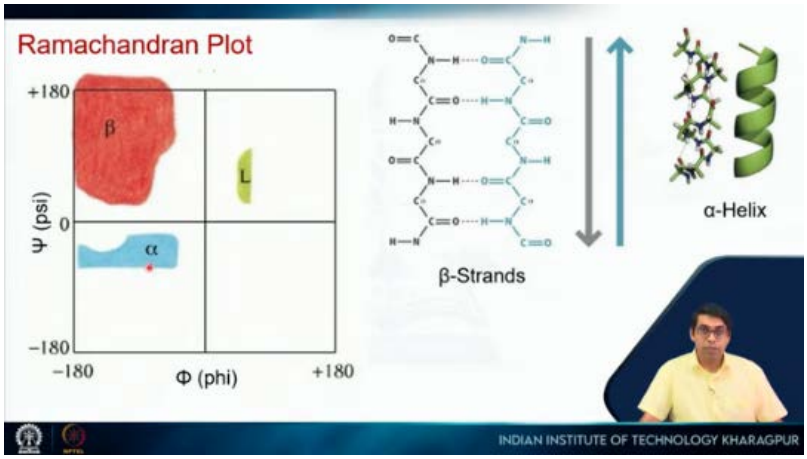
And if we see from the side view, we will see that the side chains alternate as going up and down. And this results in this pleated structure, which is why these are also referred to as beta-pleated sheets (β-sheet).



So again, we can have mixed beta strands: one beta strand goes in this direction, another goes in this direction, so these two are parallel, and the third one goes in the opposite direction, so these two are antiparallel. So, we can have parallel, we can have antiparallel, and we can have mixed orientation.



So, beta strands occupy a particular region in the Ramachandran plot, alpha(α) helix another region in the Ramachandran plot, and the left-handed helix a different region in the Ramachandran plot.



And all these structures together, so once these beta strands and alpha(α) helices are formed, they come together to form the tertiary structure of the protein. So, whenever we look at protein structures, we will often see that alpha(α) helices are represented like this helix structure, this spiral structure, and beta strands are represented by these arrows. So, this is not how it looks in the actual protein. These are called the ribbon structures, and these are shown so that we can easily identify the secondary structure.

So, in this case, you can see there are alpha(α) helices, beta strands, and then these gray regions which do not form any secondary structure. So, these are called loops. So, this is a protein which has almost all beta strands, a small turn here; this is also almost all beta strands, a small turn here; and this one has two alpha(α) helices and several beta strands. So, protein structures are typically solved by three different techniques now. One is X-ray crystallography, the other one is NMR spectroscopy, and the third one is cryo-electron microscopy.

And whenever these protein structures are solved by different labs throughout the world, they are all deposited into a data bank or a repository which is called the Protein Data Bank. Here is the link for the Protein Data Bank; you can go there and check out these protein structures, and every protein structure is assigned a unique code which is called the PDB code. PDB stands for Protein Data Bank, so every protein will have its unique

alphanumerical code. And that defines that protein structure, and if you go to the data bank, you can get a lot of information, not only about the structure but also the function of the protein. So, these are some of the books that I have referred to; you can go through one or more of these books. Some books are on general biochemistry, and some other books are specifically for proteins. So, that is all. Thank you for now.