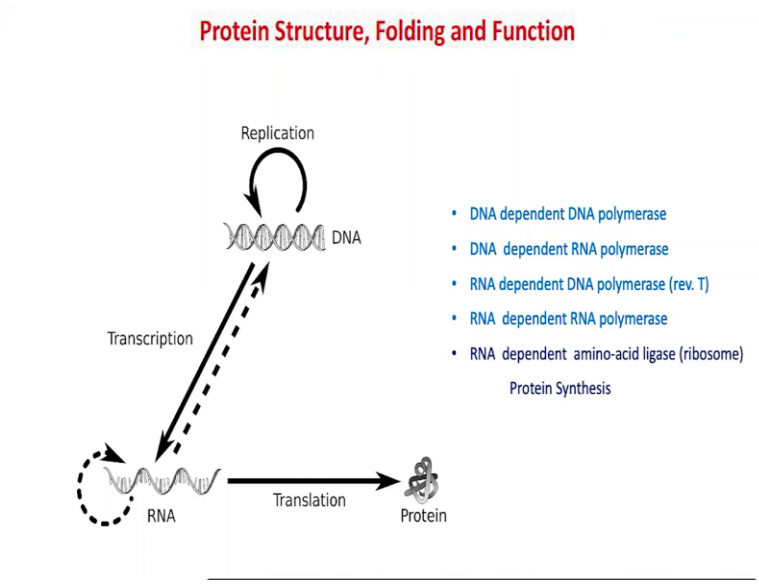
Introduction to Cell Biology
Professor Girish Ratnaparkhi and Nagaraj Balasubramanian
Department of Biology
Indian Institute of Science Education and Research Pune
Protein Structure, Folding and Function - Part 1

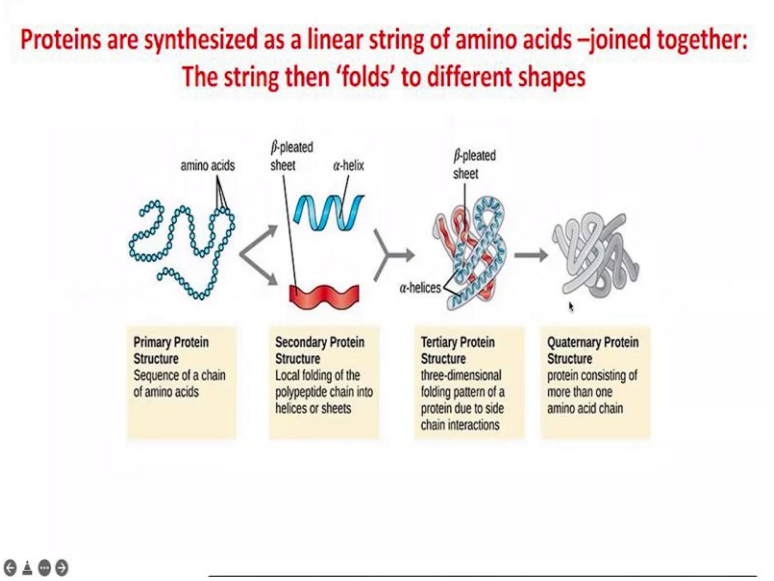
(Refer Slide Time: 00:15)



Good morning, everybody. Please confirm that you can hear me. So, today more or less marks the end of one major section of what I am going to teach you. We have gone through a huge chunk of the central dogma. And hopefully, you have had a very reasonable overview of the different processes involved. You have seen a lot of pictures of molecules. I hope you have got a idea of size scale and also a bit of an idea of reality through many of these animations, which have been shown to you.

My goal today is to wrap up this entire section by talking to you about proteins, which are the end product of the information which starts, which is stored in the DNA, and finally transferred to these executive molecules called proteins, which pretty much do a lot of work in the cell and very obviously by extrapolation, a lot of work in our body.

(Refer Slide Time: 01:19)



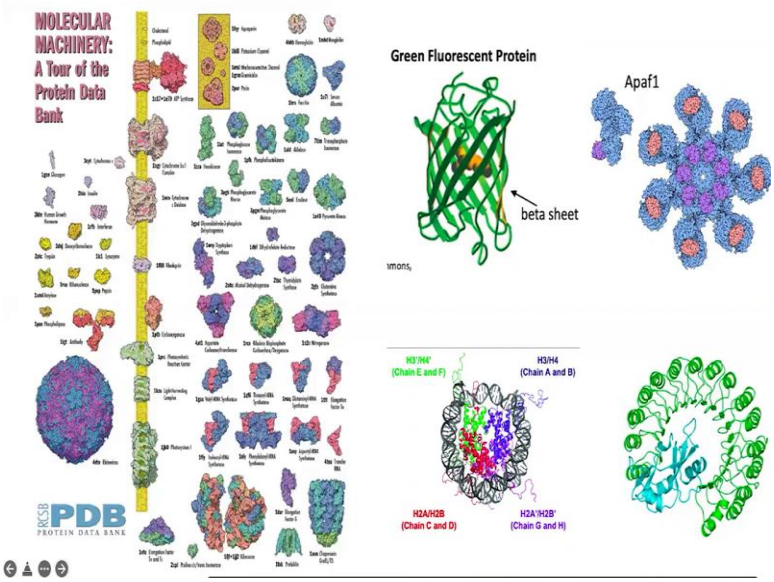
So, you saw proteins as the end product of the translation process. And proteins are nothing but beads on a string or a series of amino acids attached by peptide bonds. Now, the linear sequence of amino acids is something which we define and this is nomenclature as the primary structure of the protein. It is just a sequence from the N terminal to the C terminal of amino acids. Since there are 20 amino acids, one can have a large combination of these 20 amino acids to make a much more complex sequence than, for example, which you saw for DNA and RNA, where there were only four nucleotides.

Now, as part of the nomenclature, this linear sequence of amino acid does not really stay like a string in solution. It has this tendency to form defined structures. And the two major secondary structures are called alpha helices and beta sheets. And these are nothing but let us call this partial folding of the structure along the primary sequence of the amino acid. And I will define this a little bit more clearly in future slides.

These secondary structures then sort of fold onto each other to form what is called as the tertiary structure, so which usually has both beta sheets and alpha helices and many independent proteins with tertiary structure can bind to each other to form what we call as a quaternary protein structure. Now, this quaternary protein structure can have proteins, the same protein basically, two of the same binding to each other or two different proteins binding to each other, it does not matter, it is still defined as a quaternary protein structure.

Now, the sequence of this amino acid the length is usually variable. It can be anything from two amino acids, which is a dipeptide to 2,000 amino acids, which is basically extremely long protein structure. And remember, if you have a large single protein which is coded from a single gene of let us say 2,000 residues, you can make a massive quaternary protein structure by taking multiple 2,000 amino acid protein structures which are folded and bind them to each other and such as how large molecular machines like ribosomes are generated within the cell.

(Refer Slide Time: 03:45)



Now, there is a resource database called as the protein databank and this has been in place for about 30, 40 years. And whenever research group solves the structure of their protein or DNA or RNA in its folded state to atomic resolution or to near atomic resolution, researchers deposit the coordinates of the structure to the protein databank and the protein databank is a repository where all structures solved to date have been placed.

And there is a large variety of structures in the protein databank which represents the hundreds and thousands of proteins which are available in life forms. These are the bacterial proteins, the proteins from cockroach, there are proteins from plants, humans, mice, archaea, so on and so forth. And we also see sequence conservation across evolution. We also see conservation of folds. So all these amino acid sequences fold into different shapes, which you can see on the left hand side, and these shapes define the function of the protein.

As a few examples, remember, we have 10,000, 20,000 examples. I am just randomly picking a few. This is a very famous protein called as the green fluorescent protein. Under UV light, it fluoresces green. It is found in a jellyfish called Aequorea. And these jellyfish very obviously in the right wavelength are fluorescing in the depths of the ocean. And as you can notice, this is made of these strands, which we call beta sheets, and there not too many alpha helices over here.

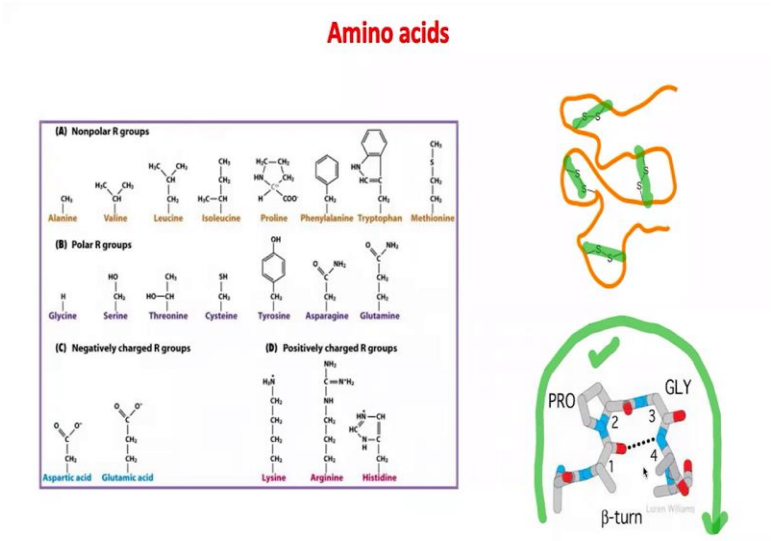
This is a multimeric quaternary structure of a protein called Apaf1. And you can see it is arranged in a very strange symmetric order. This is a quaternary structure. And this will remind you of one of these toys which many of you have played with and they have been very popular in the recent past. Shown over here is the nucleosome, you can see the DNA coiling around the nucleosome, and the histones, which are inside the nucleosome.

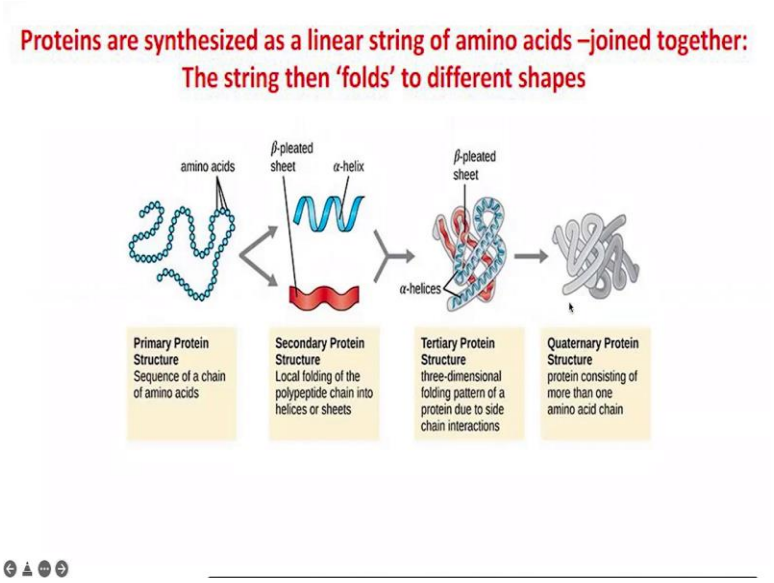
Over here is a protein which cuts RNA in light blue in cyan. This is I think angiotensin. It is a ribonuclease. But very interestingly the second protein in green over here is a ribonuclease inhibitor. And you can see it is horseshoe shaped and it goes and clamps onto a ribonuclease stopping it from cleaving RNA. So, these ribonuclease inhibitors are also very important proteins.

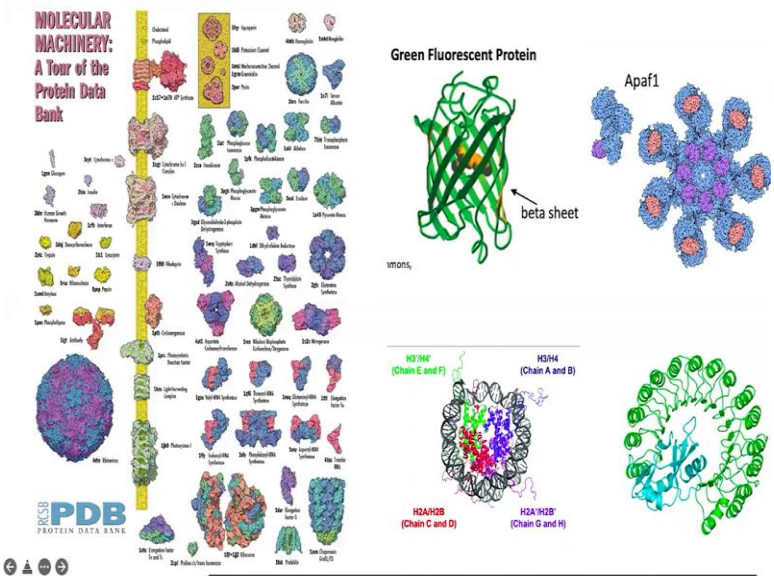
And these are just a few of many, many examples which are shown over here. I will give you a link to the protein databank page. When you go over there you will find if you do a search for DNA polymerase, it will show you structures of every DNA polymerase, which has been solved till date from bacteria to archaea to plants to humans. If you give a search for ribosome it will show you structures in various forms of all the ribosomes which have been solved today.

So, it is a very nice resource. And you can see three dimensional structures, you can rotate proteins, you can zoom into them, you can go down to atomic resolution, and really look at proteins in a very deep manner by going to this website.

(Refer Slide Time: 07:09)







Now, all of you are aware that there are 20 amino acids, and we broadly define them as non-polar hydrophobic, polar negatively charged as well as positively charged. Now, each amino acid has a certain property, a certain shape, a certain size. For example, glycine is the simplest amino acid. It has as a side chain, the hydrogen, only a hydrogen, whereas there is a methyl group for alanine. Tyrosine, phenylalanine, tryptophan have aromatic side chains. And this gives them special spectroscopic properties which we will really not talk about in today's class.

There are negatively charged side chains and positively charged side chains and histidine, for example, as many times a very special role in enzyme catalysis. Not that the others do not, but histidine more or most often than not along with serine is a very important catalytic residue. Now, all these amino acids in their different combinations basically can line up as a linear sequence of amino acids, which folds to all these beautiful shapes, which you see in front of you.

Now, by definition or the way it is, is that polypeptides proteins are not branched. They are basically a made as a linear sequence and they fold into a compact state. The exception is residue called cysteine. Two cysteine residues can basically bond to each other across the polypeptide chain making what are called disulfide bonds. And shown over here is a polypeptide chain a protein with 1, 2, 3, 4 disulfide bonds. So, if you have 8 cysteines, they can come together and make a, make 4 disulfide bonds.

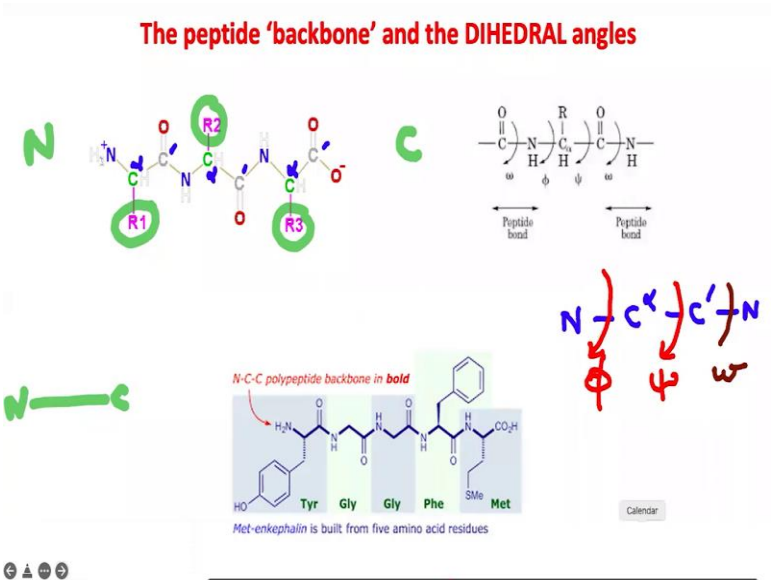
Now, there are proteins which have many, many disulfide bonds, there are proteins with no disulfide bonds and there are proteins which have cysteines, but the cysteines do not form

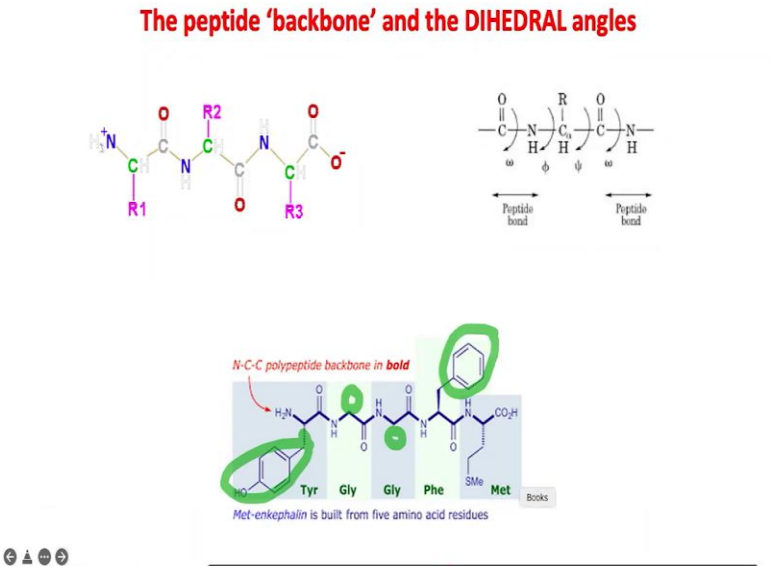
disulfide bonds. So there are all kinds of variants out there. Another interesting feature which is also a part of secondary structure is the fact that one can make turns in the folded state of a protein.

Now, here is an example of a turn. The polypeptide chain will basically take us out of a 180 degree turn over here. And this is not a beta sheet. This is not an alpha helix. It is actually a turn. And you will notice that proline and glycine are favorite residues in turns, because the proline because of its special structure, which I am not going into today, allows polypeptide chain to turn and make the turn a little rigid so that it gives it a little bit of support.

The dotted lines you see over here are hydrogen bonds. And this is one of the features of amino acids. Hydrogen bonds just like in the case of DNA are very, very important in maintaining the stability of the folded state of the protein.

(Refer Slide Time: 10:16)





Now, let me introduce you the concept of what is a protein backbone and this is something you have seen in orange over here, this is a protein backbone over here, and to the concept of what is a dihedral angle. Now, some of you may have studied this. For those of you who have not, you have to pay attention because this is a important concept and not something very easy to get the hang of in the first point.

Now, this is the, these are the atoms which are there in a polypeptide chain. This is a nitrogen, a carbon another carbon, nitrogen, carbon and this is the way the central main chain of the polypeptide repeats itself. The polypeptide has R groups, which is the side chains. This is side chain 1, side chain 2, side chain 3 and this can go on for example for very large protein for 2,000

residues. But this is the N terminal and this is the C terminal. And every protein has a N terminus and a C terminus.

For example, a protein with 100 amino acids will have a N terminus and a C terminus which is a NH3 plus over here and COO minus or COOH. But you will notice that the repeating atoms are always the same. N, C and this C we specifically give a name, we call it a C alpha. There is another C we give it a name, we call it a C prime. There is N again, there is another C alpha, there is a C prime, there is N again, there is another C alpha, there is a C prime. So, this is the repeat structure for the backbone of the polypeptide.

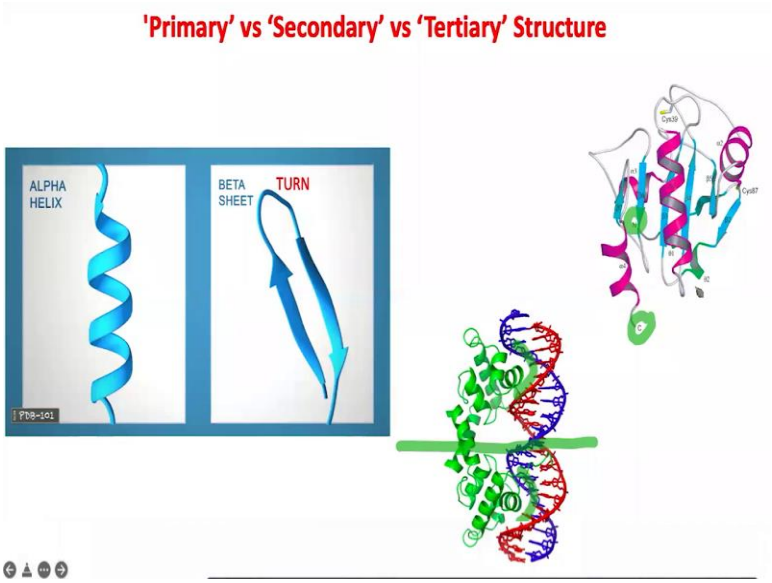
Now, we have defined certain angles which are called dihedral angles, they also called torsion angles, along the main chain of the polypeptide chain. The bond around, so I will start writing this, N, C alpha, C prime N. So, this particular bond over here is called a phi bond, this particular bond over here, the rotation along this bond is called as psi, and the rotation around this final bond is called as omega.

Now, bonds or these angles which are called dihedral angles are not defined by two atoms. So, when I say phi is between N and C one does not formally define phi just with two atoms, phi is defined with four atoms and that is something you have to remember that a torsion angle or a dihedral angle is defined by four atoms and the four atoms which define this are shown in the, in one of the slides which are up ahead. I will come to that slide.

Just remember that there are three torsion angles along the polypeptide main chain phi, psi and omega. And I have defined them partially and I will define them a little more clearly later on. Also, clearly shown over here are the R groups for this particular polypeptide. Again N, C alpha, C prime, N, C alpha, C prime, N, C alpha, C prime, N, C alpha, C prime, N, C alpha. Each time there is a C alpha, you can have an R group, tyrosine over here, glycine over here. I will keep on marking them.

So, this is one R group, N, C alpha, C prime. N, C alpha, this is another R group which is glycine. N, C alpha another R group which is glycine. N, C alpha, this is another R group which is phenylalanine and so on and so forth.

(Refer Slide Time: 14:14)



Now, let us go ahead and look at some of the secondary structure elements. So, the three secondary structure elements are alpha helix, beta sheet and a turn. Now, an alpha helix is formed from a local stretch of amino acids. Let us say there are 10 amino acids or 12 amino acids one after the other. And with the right kind of sequence of amino acids, there will be a fold called as alpha helix which is formed.

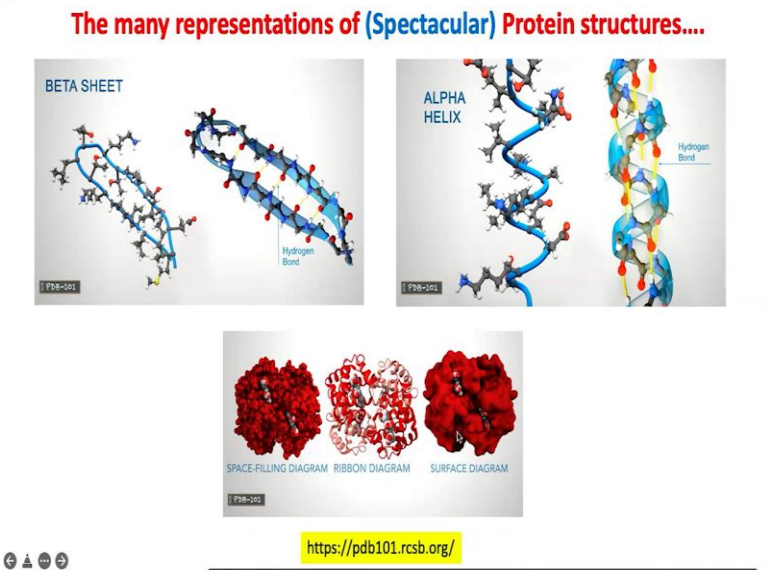
A beta sheet is a more distant structure. You can have sheets coming through which are interacting with each other as shown over here, which are really far away in the protein structure. And most many times or not between multiple alpha helices or between different sheets of a beta sheet, there are turns in the protein structure.

Now, shown on the right over here are just some examples. Here is a protein which is a transcription factor, which is made of nothing but helices and turns and it is called a helix loop, helix protein. This is a broad definition of a protein. And you can see that it is bound to DNA. And you can see it is a symmetrical structure with a symmetry basically somewhere over here, and it is bound to the major groove here and it is bound to the major groove over here.

This is more standard round protein. A globular protein is the term we use because it is rounded, approximately around in shape. It starts from the N terminus over here, goes into beta sheet, beta sheet finally makes a helix. Then this goes again into a beta sheet, again into a helix, couple of beta sheets, and then finally ends on the C terminus with a final alpha helix. And you can notice

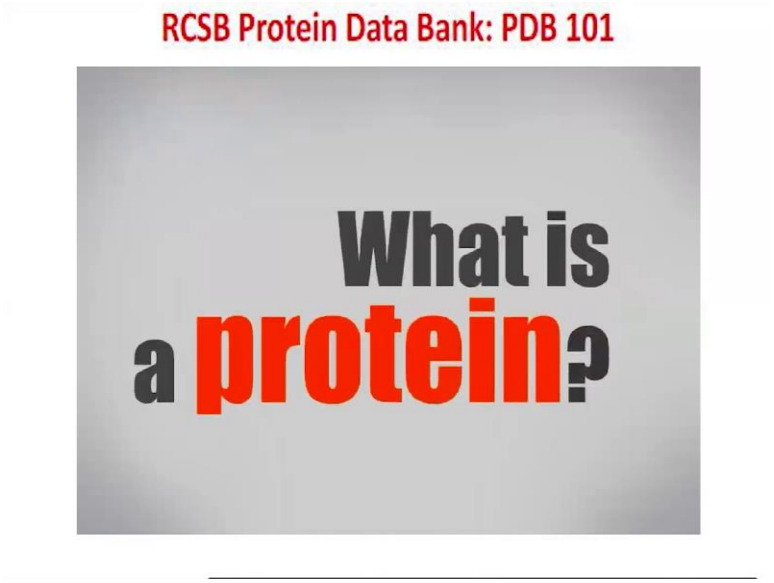
that each of these secondary elements, so this is a two dimensional picture, are all packed against each other to make a final folded state of a protein.

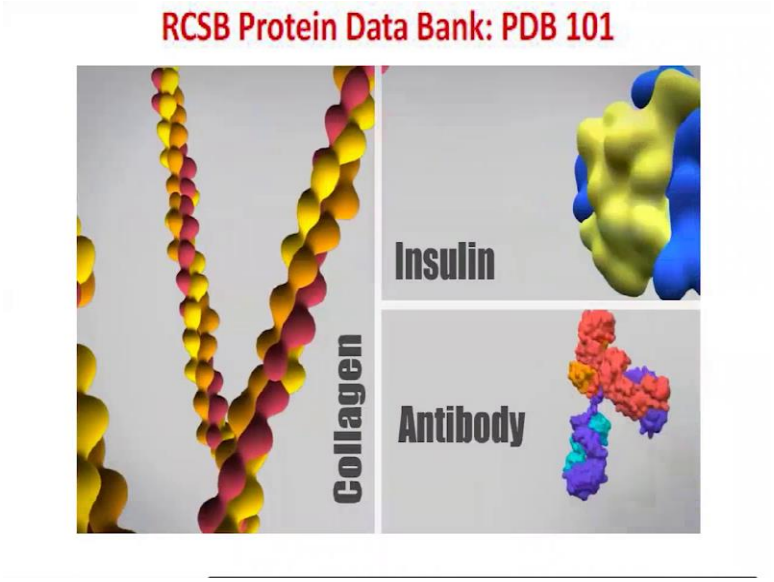
(Refer Slide Time: 16:07)



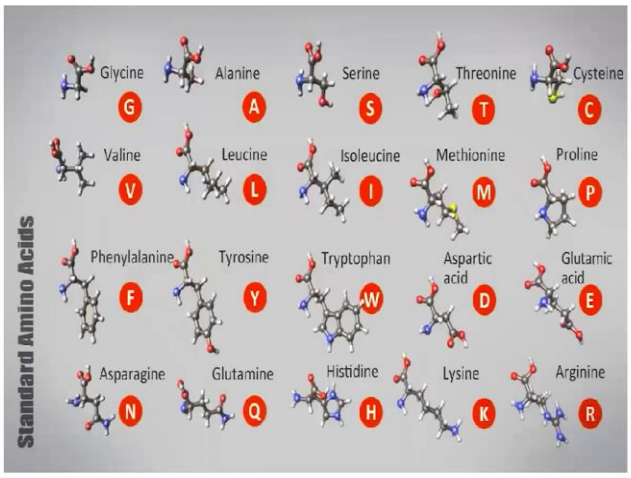
Just more representations of beta sheets, alpha helices, showing you the actual atoms along the length of the beta helix and alpha helix and the beta sheet, also showing you the hydrogen bonds in yellow, which stabilize the alpha helix and also using the different kinds of diagrams which are used to represent proteins, a space-filling diagram, a ribbon diagram, and sort of a surface view of the protein. And in this particular case, in all these three cases, this is a large globular protein of about, let us say, 100 to 200 amino acids in length. And you can look at many of these structures by going through the protein databank.

(Refer Slide Time: 16:46)

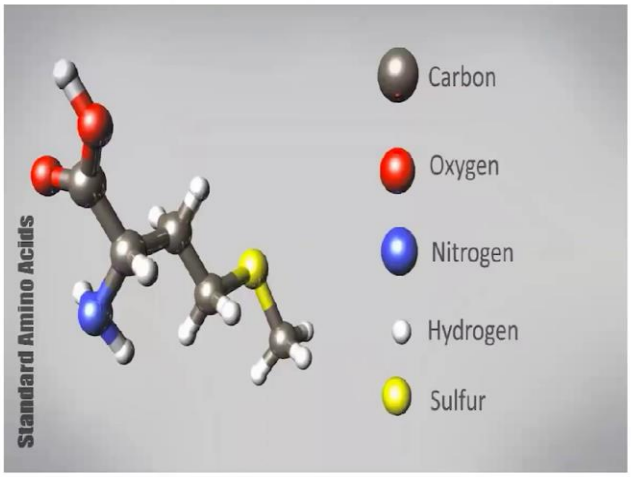




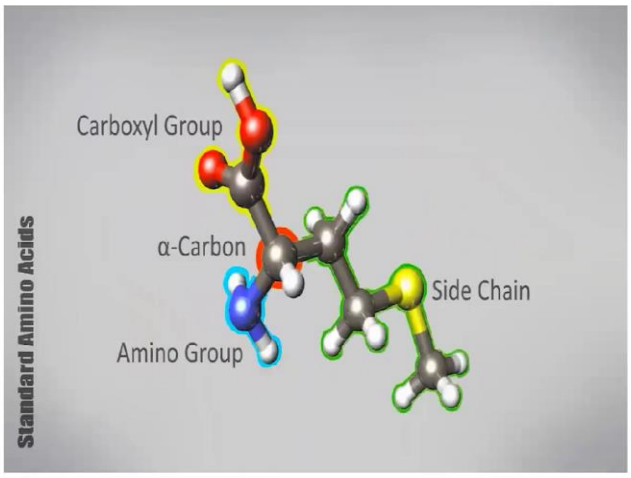
RCSB Protein Data Bank: PDB 101



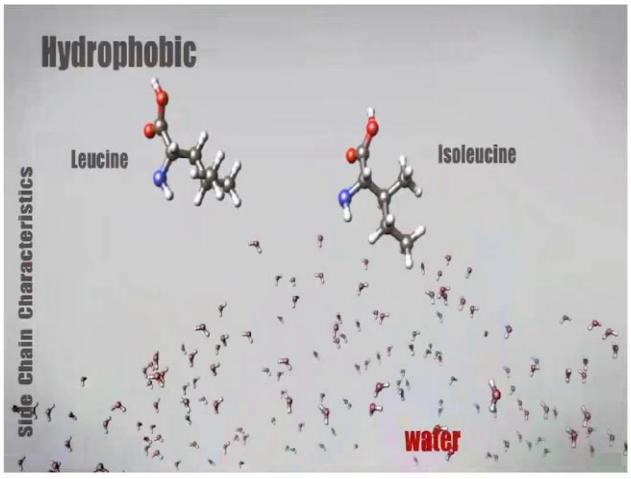
RCSB Protein Data Bank: PDB 101



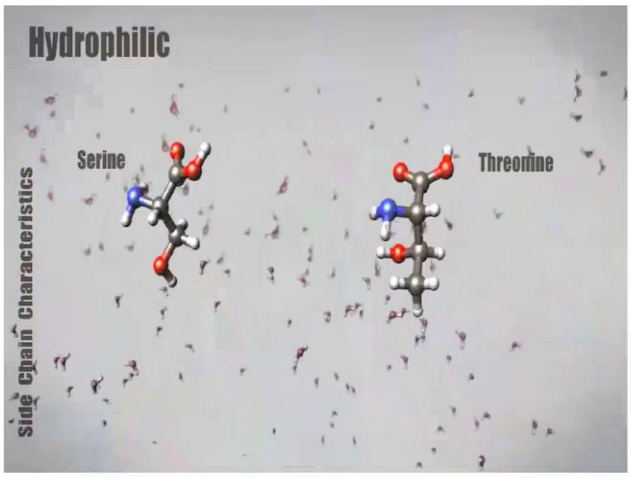
RCSB Protein Data Bank: PDB 101



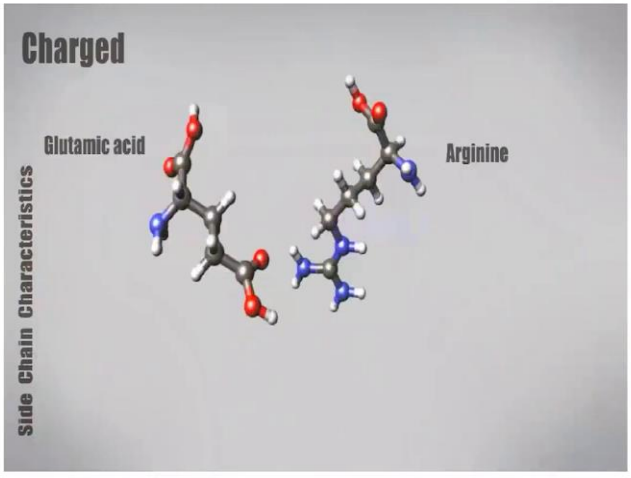
RCSB Protein Data Bank: PDB 101



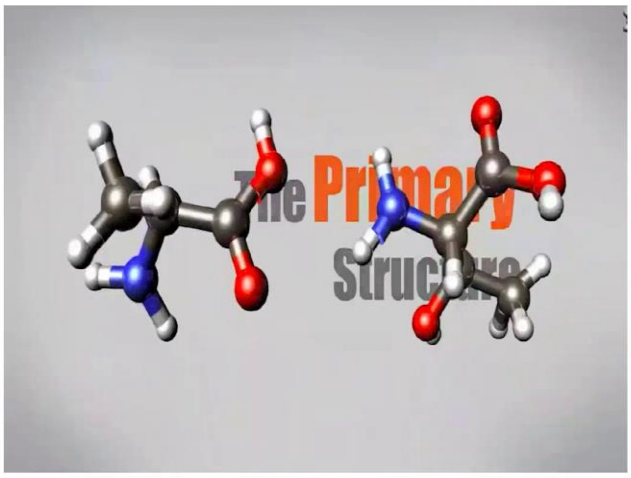
RCSB Protein Data Bank: PDB 101



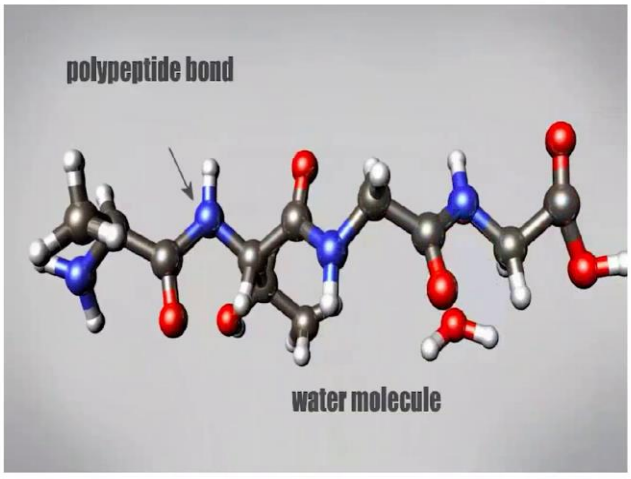
RCSB Protein Data Bank: PDB 101



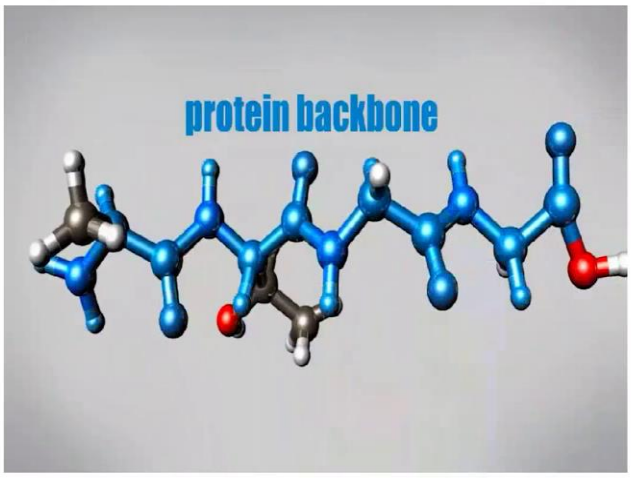
RCSB Protein Data Bank: PDB 101



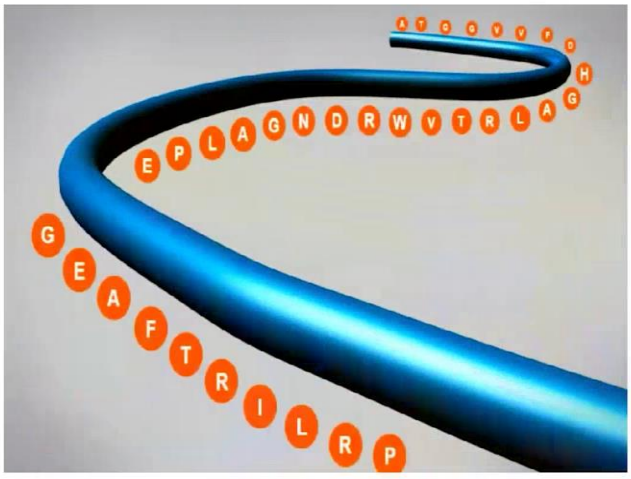
RCSB Protein Data Bank: PDB 101



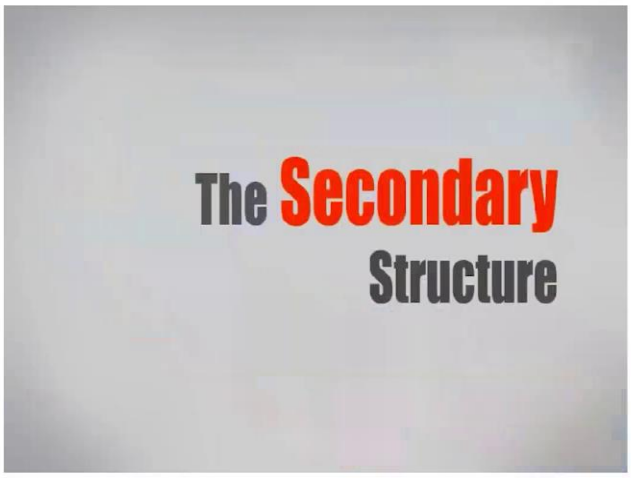
RCSB Protein Data Bank: PDB 101



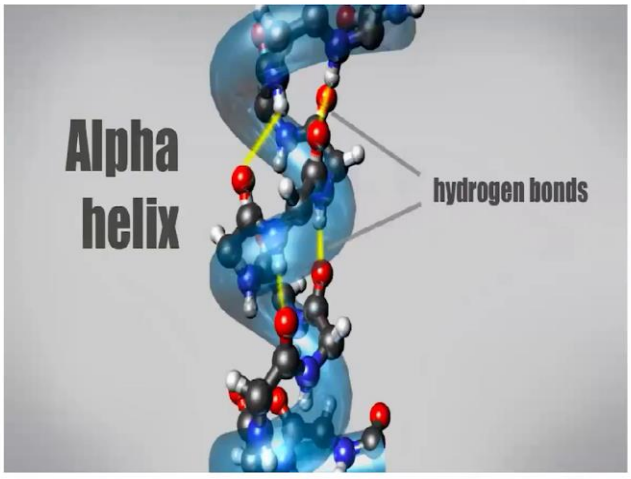
RCSB Protein Data Bank: PDB 101



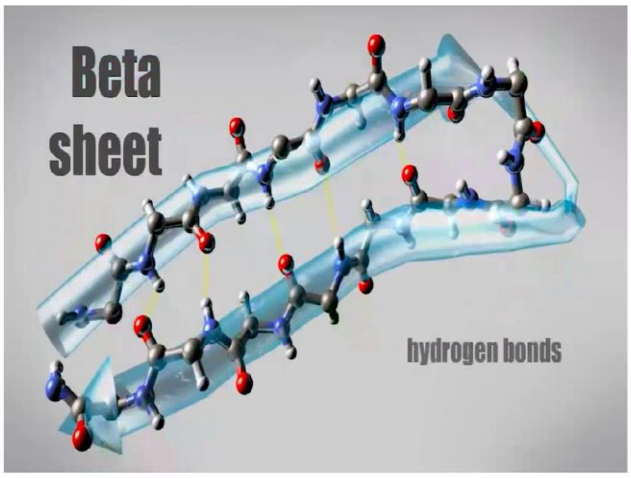
RCSB Protein Data Bank: PDB 101



RCSB Protein Data Bank: PDB 101

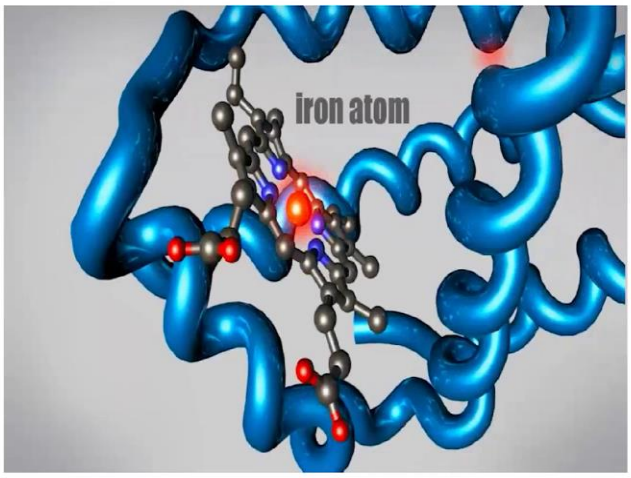


RCSB Protein Data Bank: PDB 101

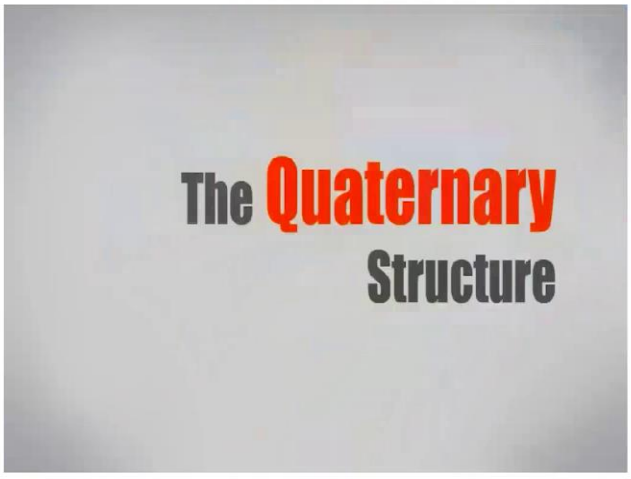


RCSB Protein Data Bank: PDB 101

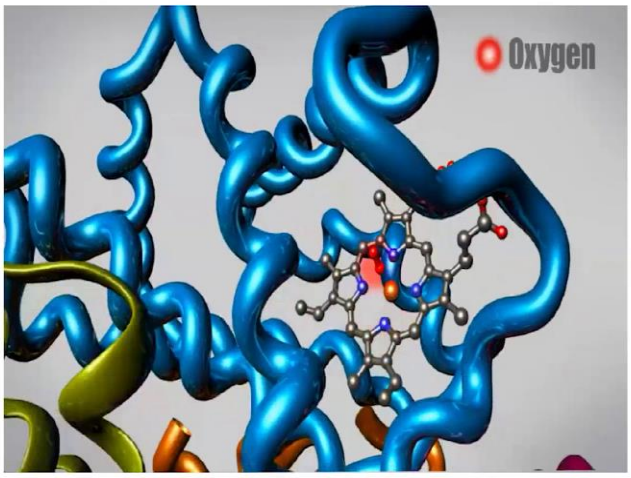
RCSB Protein Data Bank: PDB 101



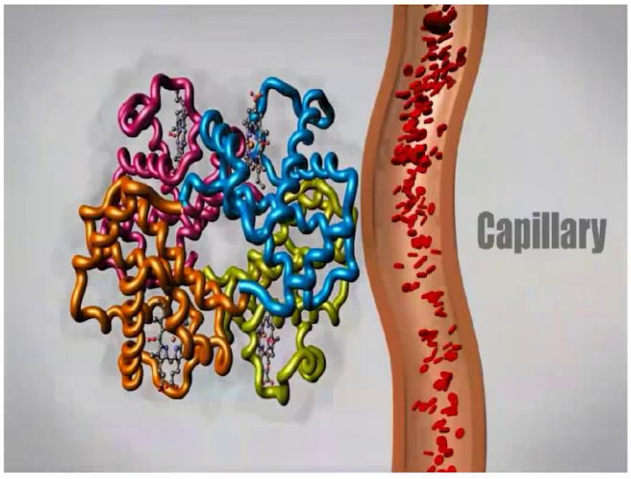
RCSB Protein Data Bank: PDB 101



RCSB Protein Data Bank: PDB 101

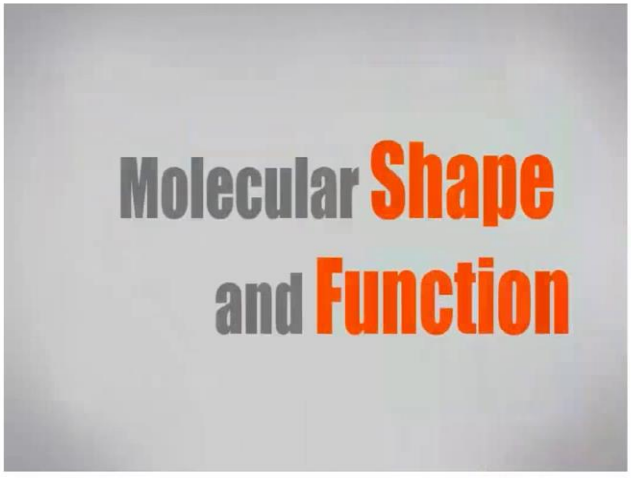


RCSB Protein Data Bank: PDB 101

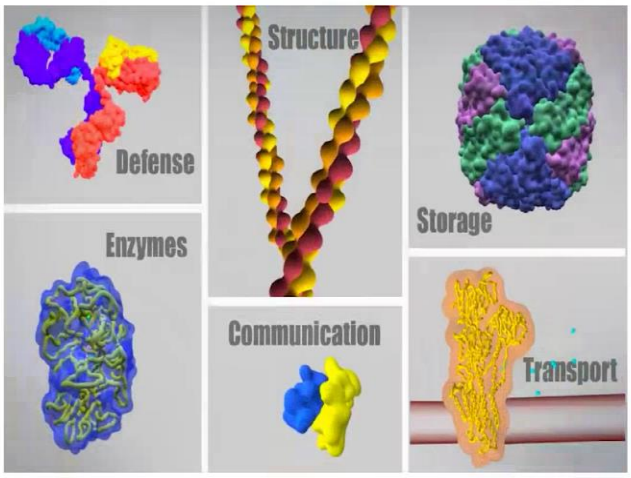


RCSB Protein Data Bank: PDB 101

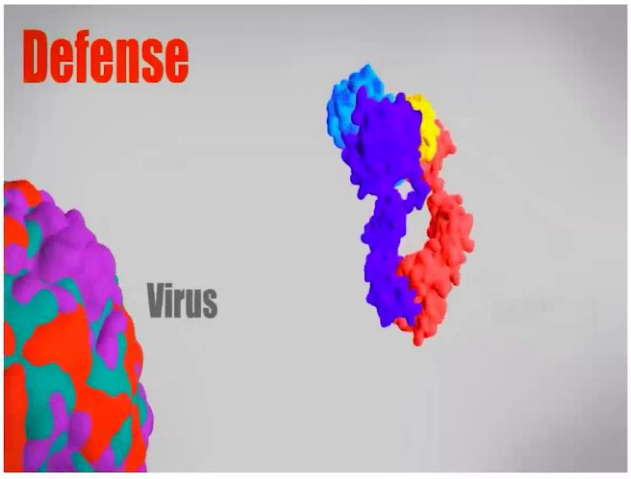
RCSB Protein Data Bank: PDB 101



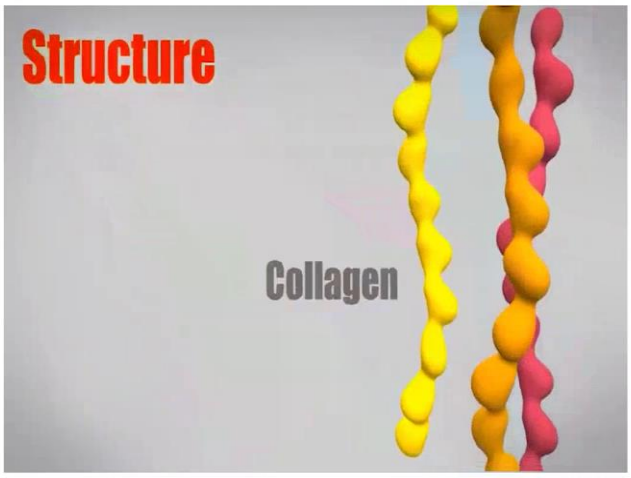
RCSB Protein Data Bank: PDB 101



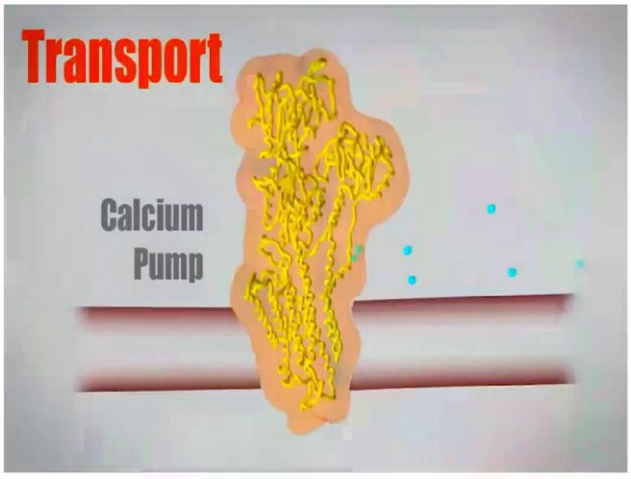
RCSB Protein Data Bank: PDB 101



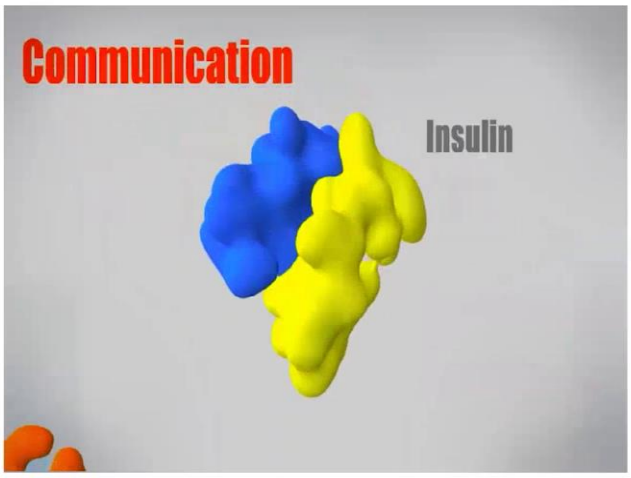
RCSB Protein Data Bank: PDB 101



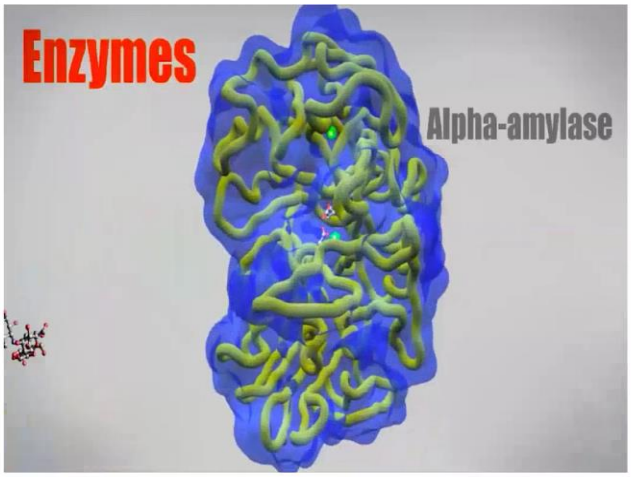
RCSB Protein Data Bank: PDB 101



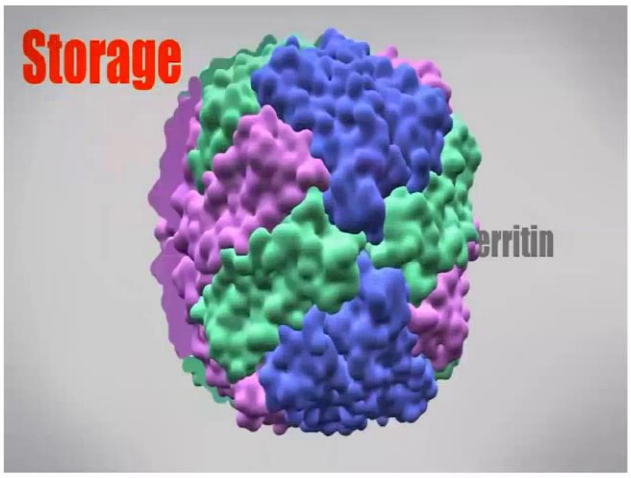
RCSB Protein Data Bank: PDB 101



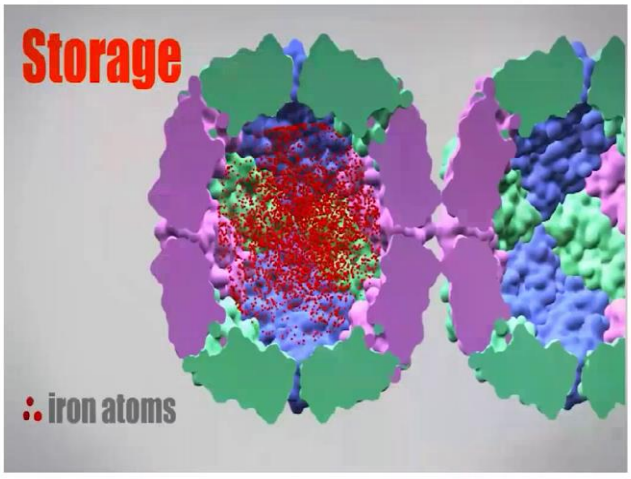
RCSB Protein Data Bank: PDB 101

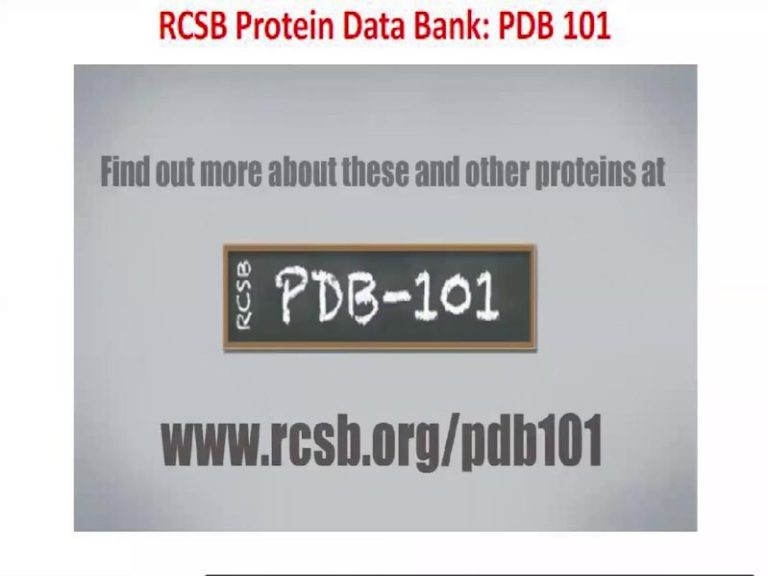


RCSB Protein Data Bank: PDB 101



RCSB Protein Data Bank: PDB 101





Now, let us see a quick movie and after that I will take questions. Proteins play countless roles throughout the biological world and catalyzing chemical reactions to building structures. Despite this wide range of functions, all proteins are made out of the same tiny building blocks called amino acids. The way these 20 amino acids are arranged dictates the folding of the protein into its unique final shape and its function.

Amino acids are made of carbon, oxygen, nitrogen, hydrogen, and sulfur atoms. These atoms form an amino group, a carboxyl group, and a side chain attached to a central carbon atom. The side chain is the only part that varies from amino acid to amino acid and determines its property. Hydrophobic amino acids such as leucine and isoleucine have carbon rich side chains which do not interact.

Now, you can see all these water molecules, which are also been shown in this schematic. Remember, whenever we look at any structure, it is always surrounded by water molecules, a topic I will come to next week.

Hydrophilic amino acids such as serine, threonine interact with water. Charged amino acids like glutamic acid or arginine interact with oppositely charged amino acids or with water. The primary structure of the protein is the linear sequence of amino acids is encoded by DNA. The amino acids are joined by peptide bonds, which link an amino group and hydroxyl group. The water molecule is released each time a bond is formed.

Specific amino acid sequences give proteins the distinct shapes and chemical characteristics. These protein chains often fall into two types of secondary structures stabilized by hydrogen bond. A protein chain can fold into a rigid alpha helix forming regular patterns of hydrogen bonds between the backbone atoms of nearby amino acids.

Backbone atoms of the chain can interact side by side to form beta sheets. Many proteins fold into a compact globular shape with hydrophobic side chains sheltered inside away from the surrounding water. The functions of many proteins rely on this global structure. For instance, hemoglobin forms of pocket protein, a small molecule with an iron atom in the center that binds oxygen. Two or more polypeptide chains can come together to form one functional molecule with several subunits. The four subunits of hemoglobin cooperate so that the contents can pick up more oxygen in the lungs and release it in the body.

Many proteins rely on the ability to recognize the shape of specific molecules in order to function correctly. The flexible arms of antibodies protect the body from disease by recognizing and binding to foreign molecules and thus preventing the viral RNA or DNA to enter the cell. Collagen forms the strong triple helix that is used throughout the body for structural support.

The calcium pump moves ions across cell membranes allowing the synchronized contraction of muscle cells. The hormone insulin is a small staple protein that can easily maintain its shape while traveling through the blood to regulate blood sugar level.

Alpha-amylase is an enzyme with a catalytic site that begins the breakdown of carbohydrates in our saliva. Ferritin forms the hollow shell that stores iron from our food. Learn more about the functions and 3D structures of nucleic acids, proteins and molecular machines at the RCSB protein data bank. So, this was a broad introduction to proteins.