Course Name: I think Biology

Professor Name: Kaustubh Rau

Department Name: Biology

Institute Name: Azim Premji University

Week:3

Lecture:12

W3L12_Biomolecules - Part I

Hello and welcome to the I think biology NPTEL course. This is week three and we are looking at the building blocks of biology. This is the introduction to biomolecules. This lecture is in two parts. In the first part I will be talking about nucleic acids and proteins and in the second part I will be discussing polysaccharides and lipids. So let's start with the periodic table.

You will see that elements are shown in two colors, purple and green in this periodic table. So take a minute and think about what these two different categories might mean for a course in biology. So let me show you what the colors point to. So shown in purple are elements which are essential for humans and shown in green are elements which are suggested to be essential.

And then if you look more closely at particular elements starting on the extreme left with hydrogen, the five elements which are important, I mean which are from which form the major component of our body are hydrogen, carbon, nitrogen, oxygen and phosphorus. Then apart from that on the left hand side you will see sodium, magnesium, potassium and calcium and these are also extremely important as ions, inorganic ions within the body. The other important ion is the halogen which is chlorine which is on the extreme right and then in the middle you will see in the fourth period many other metals which are manganese, manganese, iron, cobalt, copper, zinc, selenium. So why are these metals important for us? And that's something for us to think about. Of course we know the case of iron and it forms a part of hemoglobin and as an oxygen carrier.

But similar to iron, other metals also form the metal center of many important enzymes which are used in redox reactions within the body. And that's the reason why they are important. The other thing is that these metals also point to some of the early origins of cells when these metals in the form of minerals were available to the early biomolecules such as early forms of RNA and they may have been used to catalyze certain reactions and they have since been incorporated within other kinds of biomolecules within our cells. Looking at two entities within our body more closely, one is a molecule which is water and the other one is

the element carbon. So water is a molecule which has polarity because oxygen is more electronegative than hydrogen.

The shared electron pair is pulled slightly towards oxygen, conferring a slight negative charge to that end of the bond and a slight positive charge to the hydrogen end of the bond with the effect that the water molecule has some polarity, which means that it can form hydrogen bonds with other water molecules. And this becomes important within our body when hydrogen bonding plays a major role in deciding the stability of our molecules. Then looking at carbon, the most important form of carbon we need to examine is hydrocarbons. And hydrocarbons are molecules made up of hydrogen and carbon and they can exist as linear chains or as carbon rings or combinations of both. And carbon-carbon bonds can be single, double or triple.

And these have consequences because depending on the kind of bond you have, it affects the molecule's geometry, which means it affects the molecule's function. And that's something we will emphasize throughout the lecture is the relationship between structure and function. So now if you look at ethane and ethene, ethane has a single single carbon-carbon bond, whereas ethene, which is C2H4, has a double bond. So the double bond is a planar bond, which means it constrains the movement or the rotation of one carbon around the second carbon, which means the ethene molecule has certain restrictions in the kinds of rotations or movement it can do. So this can have an implication for biomolecules within our body. For instance, if you have fatty acids and you can have saturated and unsaturated fatty acids. And then they can in turn take certain conformations or 3D orientations, which then have an impact on their properties. So what are the kinds of reactions we need to be aware of? The two main ones are shown here. One is a dehydration reaction, which is used in the synthesis or the making of molecules. So typically two molecules come together and one water molecule is removed.

So that's why it's a dehydration. And these two molecules are then joined together. Shown here is the reaction of a monosaccharide. Two glucose molecules are shown and they come together to form a disaccharide with the exclusion of one water molecule. On the other hand, the other major reaction hydrolysis is the opposite of a dehydration reaction where a water molecule is inserted and used to break up the bond between two molecules.

So again shown is the reverse of the reaction, which I talked about earlier, where a disaccharide disaccharide is broken up to form two monosaccharides due to the insertion of one water molecule. So these two reactions are important to think about when we talk about biomolecules. Finally we also need to consider how the body builds up structures within a cell. So you start with certain building blocks and the building blocks for biomolecules are in four major categories. They're sugars, fatty acids, amino acids and nucleotides.

These then link up through the reactions which I showed you earlier to form larger units

or polymers. So sugars will link up to form polysaccharides, fatty acids will link up to form lipids, amino acids will link up to make proteins and nucleotides will link up to form nucleic acids like DNA and RNA. And then there is a further step that we need to think about in the building up of structure. So you start with these building blocks, they go on to make these polymers or macromolecules or large molecules and then these macromolecules come together to make even larger assemblies. So shown here is the example of a ribosome.

So a ribosome is composed of proteins and RNA and they come together in a particular confirmation to form a ribosome which has a particular structure which is important for its function. And you can see that the ribosome is a fairly large assembly. It can be up to 30 nanometers in size which is quite large on the scale of a molecule. So this is one theme that we need to be aware of when we look at biomolecules and cells is the building up of structure starting from small to large. Then let's think about the composition of a cell.

So shown on this map is the composition of an E. coli cell and the different colors indicate the different classes of molecules. So proteins are shown here in light brown and they form the major component of the cell. Nucleic acids are shown in blue. Again DNA is shown as dark blue and RNA is in light blue.

And you can see that RNA forms a much larger component of the cell as compared to DNA which might be a surprise to many of us. Then you have lipids which are again a major component of the cell. Then you can have storage molecules like glycogen which is a polysaccharide. And then you can have certain composite molecules which are made up of two different kinds of molecules coming together. For instance you have a peptidoglycan,
So this means this has a polypeptide part and then it has a polysaccharide part which is the glycan. And they come together to form this composite molecule or you have a lipopolysaccharide. So you have a lipid and a polysaccharide coming together to form this molecule. So you can see that we don't just have to stick ourselves, I mean stick to these four main categories of nucleic acids, proteins, lipids and polysaccharides.

But these can combine with each other to give these composite molecules with very specific properties which are made a use of. Then you can also have certain other kinds of molecules which could be found in specific cells, for instance These polyamines which are found in bacterial cells or you can have smaller molecules like metabolites which are used for various functions like signaling or buffering and things like that. Here is the same data shown in the form of a table.

Now you're comparing an E. coli or a bacterium with a mammalian cell. And then if you look just look at the non-water components of each cell you will see that proteins make up a large proportion of that component, Over 50% for both cells. Then you have the nucleic acids which are the other major component.

And again here the percentage of RNA is greater than that of DNA. And so that's something

for us to think about. Then you have the phospholipids and you have polysaccharides. The inorganic ions also form a major component of the cell. And we need to think about them when we talk about cells because of the major role they play in the functioning of the cell, So sodium ions, potassium ions, chloride ions, calcium are all very important to think about when we talk about these biomolecules. So now let's look at nucleic acids in a bit of detail. So shown here is a nucleic acid chain and it has three main components. You have nitrogenous base which is shown in green. You have a sugar which is shown in blue.

And you have a phosphate which is shown in yellow. And they are joined together in a particular way so that you get a nucleic acid. First if you look at the sugar, in nucleic acids you have a five-membered sugar, you have pentoses. So all the carbon atoms in that sugar have been labeled from one to five. And if you are looking at DNA, then at the second carbon, DNA is made up of deoxyribose, So at the second carbon instead of having an OH group, you have an H group. So it's a deoxyribose. And the sugar links up with phosphates to form the phosphodiester bond. And there are two main carbons which will participate in this reaction, the carbon at the third position and the carbon at the fifth position. So so since both of these can link up with phosphates, you can get a nucleic acid chain with the formation of these phosphodiester bond.

And then at the first carbon position, it is joined to these nitrogenous bases. And in DNA, they can be a four types, guanine, adenine, thymine, and cytosine. So this makes up the basic structure of a nucleic acid. And if you look at the individual structure, individual units of a nucleic acid, they are made up of nucleotides and nucleotides have two functions.

One is that they form DNA and RNA. So you can say that they are information storage molecules since the sequence of the nucleotides is used to make up our genes. And then from there, we build up proteins. But the other main function of nucleotides is that of energy carriers shown here is ATP and cleaving the phosphate bonds leads to a release of energy, which is used to catalyze reactions within the body. So looking at the structure of DNA in more detail, you have two chains that come together to form the famous double helix. And the important thing to note is that each chain has a direction. So each chain runs from the 5' end to the 3' end. And the chains link up in an anti-parallel mode. So you have the 5' and 3' on one side and also at the other end. And the nitrogenous bases form hydrogen bonds with each other.

A binds to T and G binds to C. And this allows the DNA molecule to come together as a double helix, which is the structure we are all familiar with. The AT pair has two hydrogen bonds and the GC pair has three hydrogen bonds. So it is more stable than the AT pair. So the things to note are the directions of the DNA molecules and the AT GC pairing. The other thing to note is that if you look at the double helix shown on the left, you can have one major groove shown and one minor groove. And these are sites where proteins can bind to DNA. And this becomes important when we talk about gene regulation or DNA replication. So the binding of these proteins allows certain transcription factors to turn on or DNA to unwind

and to replicate and things of that kind. So there is a further structural consideration to think about beyond that of the double helix and the base pairing of AT GC.

The same structure is shown here. The reason I have shown this structure is to allow you to appreciate the planar nature of these nitrogenous bases. So these are these ring structures and because they are planar, they give a particular shape to the molecule. And so again we have the two ends, the 5' end and the 3' end and the hydrogen bonding as I spoke about earlier. Another factor to consider is a different kind of bonding which is not normally spoken of but which also confers stability to the double helix and that is the pi-pi stacking interactions between base pairs. So because these are ring structures, you can have a delocalized electron cloud.

You will know this most famously from the ring structure of benzene. And so these pi-pi interactions between two consecutive base pairs can also stabilize the structure of DNA and give it an added stability. And that's shown here in these dots between the base pairs in the structure shown. So then moving on, let's talk about RNA. The two major differences between DNA and RNA is that in RNA you we have a ribose sugar so at the second carbon it has a hydroxyl group and then instead of a thymine we make use of uracil.

And the difference between a thymine and a uracil is that the uracil does not have a metal group but the thymine does. But the formation of RNA is fairly similar to that of DNA. The sugar phosphate link up to in the form of the phosphodiester bond to give nucleic acid chains and then the nitrogenous bases will hang off the carbon at the first position of the ribose to form part of the chain. So again an RNA molecule will also have a direction so it will have a 5' end and a 3' end. Now as students of biology we need to be aware of one thing.

When we are shown a sequence of nucleotides for a DNA molecule and if the direction is not given we have to assume that it runs from the 5' end to the 3' end. And if we are given the sequence for one chain we should be able to write the sequence for its complement and show the direction for that. And then from there if we assume that a particular strand of the DNA molecule serves as the template to form RNA during transcription then we should be able to write the structure of that RNA molecule giving the correct directions for that molecule. So this is something that you should practice on your own. The other thing to note is that RNA can also have structure but over short lengths of the molecule.

So shown three different ways of showing I mean two different ways of showing an RNA molecule. So if we look at A and B. In A we see that it is just one long RNA molecule and at certain points it has folded up over itself and where it can form these base pairs. So you can have the standard GC pair and then instead of the AT pair you will have the AU pair. And because it can base pair over short stretches it can confer stability to the RNA molecule and it can also give it some structure.

The other thing to note is that you can also have non-conventional base pairing. So apart from the GC AU pairing you can have base pairing between different other kinds of bases also. And then the structure shown in C is a 3D rendition of a tRNA molecule and again here you can appreciate that RNA can also fold to form a helix structure over short stretches and the reason for this is the base pairing which can be of a conventional type and a non-conventional type. So moving on to proteins. Proteins are made from amino acids and an amino acid is a carbon which is bonded to four different groups as shown in the image.

So gray are carbon atoms, blue are nitrogen atoms, white are hydrogen atoms and red are oxygen atoms. So a carbon, the central carbon can be bonded to a carboxylic group on one side so COOH and an amine group on the other side so NH2 and then the other two valencies can be satisfied in different ways. One of those valencies is satisfied by hydrogen so H and the other you can have many different groups bonded to that central carbon and so this is known as the R group or the side group. And then depending on what the R group is you can have different kinds of amino acids. So we have 20 amino acids within our body and they can be broadly categorized as non-polar or polar and so this is decided by that R group or the side group.

And then two amino acids can come together in a dehydration reaction, the loss of one water molecule to form a dipeptide. So shown in the figure is the formation of a dipeptide of glycine and alanine. And so again the thing to note here is that apart from the formation of the peptide bond which is a CONH bond, the dipeptide also has a direction so it has a N end and a C end. And so every protein molecule which is built up basically by the combinations of these amino acids will also have an N end and a C end depending on how these amino acids link up. So that forms the primary structure of a protein which is basically just a sequence of amino acids.

Other things to note are that each amino acid can either have a three letter code so alanine will be ALA or just a one letter code A. So you can completely specify the sequence of amino acids in a protein by writing down the letter codes of these amino acids. So after you have an amino acid chain, you can build up the structure because of different kinds of bonding. And so shown here is the secondary structure of an amino acid chain and it can form two kinds of structures. One is an alpha helix which is basically hydrogen bonding between two parts of the chain and these parts are close to each other.

Or you can have a beta sheet which is hydrogen bondings between two different parts of the chain and these parts can be further away from each other. So the chain can fold back on itself and because of the alignment of positive and negative groups, you can have hydrogen bonding between them to form this beta sheet. And so the arrows in the beta sheet shows you the direction of the chain. So this hydrogen bonding to form either an alpha helix or a beta sheet will give stability to the ultimate structure of the protein molecule which is shown in this figure. Here you can have a hydrogen bond between a peptide bond and a side chain of an amino acid or you can have a hydrogen bond between two different amino acid

side chains.

So the secondary structure of the protein is used to form or build up the tertiary structure of the protein. So the protein within a cell will have a specific 3D structure and this structure is closely linked to its function. And the way the three dimensional structure of the protein is built up is via a variety of non-covalent bonds. And these can be of various types. We already spoke about hydrogen bonding to form alpha helix or beta sheet.

You can also have a hydrophobic reaction interaction which is basically two side chains or R groups which are hydrophobic in nature. They will interact with each other to exclude water from that small region of the protein. You can have hydrophilic reaction interactions So charged groups can interact with water. You can have the formation of a salt bridge which is basically a positive positively charged side group interacting with a negatively charged side group.

You can have one kind of covalent bond forming which is for the amino acid cysteine and so it can which has a sulfur side group and these sulfide groups can link up with each other to form a disulfide which is a covalent bond. And so this can again confer extra stability to the structure. Finally you can have a metal ion coordination. So I referred to metals in the early part of my lecture. So if you have a metal ion which forms part of the structure then you can have this multivalent interaction between metal ion and parts of the side chains of the protein.

So all these interactions together will decide what is ultimately the 3D structure of the protein. So this idea is shown here in a simpler form. So if we just consider that these amino acids have polar side chains and nonpolar side chains then the protein can fold up such that the nonpolar side chains form the core of that protein so that they exclude water and the polar side chains are on the outside of that protein molecules and they can interact with water by forming hydrogen bonds. And so this is an easy way to think about proteins folding to give a 3D structure which will ultimately decide their function. So going into just a little bit of detail about the tertiary structure of a protein.

So we can think about the tertiary structure as being built up of many smaller domains and so each of these domains can have its own particular structure. So shown here is a protein called a kinase and it is made up of four domains. For instance if you look at the large kinase domain it has many alpha helices in it. But if you look at the small kinase domain shown in yellow it is mainly made up of beta sheets and you can see that the ATP molecule which is shown there which indicates that that is actually the site of catalysis of this particular protein molecule. So you can have these many domains coming together to ultimately form the 3D structure of the protein.

And the the the domains can link up in a variety of forms. So shown on the right is a fibronectin which is part of the extracellular matrix of cells and here the domains have

linked up in a linear form and each domain is mostly made up of beta sheets.So the idea of modularity is again built into the protein structure itself. The other advantage of having domains is that you can do what is called domain shuffling. So if you have certain proteins with perhaps similar functions then you can have different domains in each of these proteins but you can mix up the domains such that you can form different kinds of molecules just using a specific subset of domains. And so you can form a larger number of proteins just using the same motive of specific domains and then shuffling them to give you kind of a library of protein molecules.

Coming back to the theme of structure and function. So we all have proteases within our digestive system. Proteases are enzymes that are used to break up other protein molecules. So the three proteases named here are Chymotrypsin, Trypsin and Elastase. And they are all serine proteases which means they catalyze a particular kind of reaction and they work at a pH of 8.0. Now shown at the bottom are 3D structures for Elastase and Chymotrypsin. And the part which is shown in green is similar in both these protein molecules and the part which is colorless, the amino acid sequence is different in both these molecules. So you can see that if you look at the amino acid sequence you will find a large dissimilarity between these two proteins. But because they do a similar kind of function within the body the 3D structure of the proteins is maintained. And especially at the catalytic site which is shown in purple you will find the same kinds of folds for both Elastase and Chymotrypsin. So again driving home the point that because the tertiary structure is important for the functioning of the protein that is maintained and you can have a difference in the primary structure of the protein.

This point is shown here again from an evolutionary perspective. So this slide in this slide you can see a comparison between a protein from yeast and a protein from the fruit fly Drosophila melanogaster. So if we look at C what has been given there is the amino acid sequence for both these proteins in yeast and Drosophila and the colored sections of these proteins are the ones which form specific structures, 3D structures in both yeast and Drosophila. Now because this protein performs a very very important function which is it binds to DNA even though yeast and Drosophila have diverged more than a billion years ago because this protein in yeast and Drosophila performs the same function the structure of this protein has been maintained even though the primary sequence of amino acids is now different. And so if you overlay the yeast protein on top of the Drosophila as shown in B you will see that there is a very close match except that the yeast has one extra fold as shown by those three dots and they are also highlighted in the amino acid sequence shown below. And the 3D structure is shown in A so it is formed by three different alpha helices.

Finally you can have one more order of structure which is the quaternary structure that certain proteins form. This is basically the coming together of two protein molecules of the same type and because of certain interactions they can form a quaternary structure which is stabilized mainly by hydrogen bonding and then this quaternary structure is important for the function of that protein. So shown here is the example of the enzyme Neuraminidase and

this is made up of four identical polypeptide chains and if you look within each chain you will see that the beta sheet motive is repeated and if you look at the different colors of the beta sheets you will see that they form different domains within one particular chain. But then all these chains come together and they have the same kind of interaction with each other and they form one large protein which is then made use of for a particular function. The other famous example of course we know of is that of hemoglobin where four globin molecules come together to form a quaternary structure.

I just thought I would show you the different shapes of proteins. This is from the protein database website and you can find on that website many different kinds of protein shapes which have been given. Here all the protein molecules shown are at the same scale. The scale bar is given at the bottom. So we need to think about the fact that proteins occupy a 3D space within the cell and they occupy a particular volume within the cell.

And so we need to know the sizes of a few of these proteins. For instance we can see that the hemoglobin molecule is approximately 5 nanometers large. Then apart from things like enzymes, so shown here is a nucleus, so this will be an enzyme which will cleave DNA. You can have structural proteins like collagen. So this forms long fibers which give strength and flexibility to our cells. Then you can have proteins which are involved in cellular respiration such as alcohol dehydrogenase or aspartate transcarbamoylase.

These are some extremely large protein assemblies. They are made up of many different kinds of protein molecules coming together. So I think we should appreciate the many different kinds of proteins and the different functions that they do within our body to keep our cells going. The same point is made here in this pie chart. So for a human cell or human body, different classes of proteins have been shown which have so far been identified. So if you look at the major category, over 4000 proteins in our body are still unclassified which are approximately one-fourth of all the known proteins within our body.

But if you look at the proteins for which the function is known, then you can look at the various functions. You have receptors, you have storage proteins, you have structural proteins, transcription factors, cell adhesion molecules, signaling molecules, etc. So this wide range of functions that proteins do within our body can be appreciated in this pie chart. So with that I will end the first part of this lecture on biomolecules. In the next lecture I will talk about polysaccharides and lipids. Thank you.