Molecular Biology
Prof. Vishal Trivedi
Department of Biosciences and Bioengineering
Indian Institute of Technology, Guwahati
Module - 03
Basics of Biomolecules
Lecture-14 Amino acids

Hello everyone, this is Dr. Vishal Trivedi from Department of Biosciences and Bioengineering IIT Guwahati and what we were discussing, we were discussing about the biomolecules. In today's lecture, we are going to start discussing about the new biomolecule and the new biomolecule what we are planning to discuss is the proteins. So as you can see that the proteins are actually being considered as the building blocks and these protein molecules are required for maintaining the, you know, the whatever the losses we have right. Suppose there is a loss of protein and we require the protein to replenish or if they will be a damage, then also we require a protein to reconstruct the damaged portion. Apart from that the proteins are also been responsible or also been functional as the enzyme.

So and that these enzymes are required for running the different types of metabolic pathways. I am sure we have already discussed a lot about the role of these enzymes into the carbohydrate metabolism when we were discussing about the glycolysis and the shape cycle. So subsequent to that, let us start discussing about the protein molecules. So as we said, you know, protein is a biomolecule and it is required for the as a building block.

So it is required at the building block, then it is required as an enzyme and then it is also required for running the metabolic reactions. Proteins are even responsible for making the different types of structures in the body, such as like we could be able to stand and you know walk because of the fibers which are made up of the protein. So all these we are going to discuss in this particular module. So let us discussing about the protein molecules. So protein are the polymer of the amino acids and they are joined by a covalent bond which is known as the peptide bond.

Each protein can be broken into the constituent amino acid by a variety of method to study the free amino acids. So what you can see here is that this is a generalized structure of amino acids, what you will see is you have a carbon in the center which is called as the C alpha carbon and then this C alpha carbon is attached to the four groups you have amino group, you have a carboxyl group and then you have the H and on the fourth side we have a side chain. So depending upon the side chain, we can have the different types of amino acids. So there are 20 different types of amino acids which are

present in the proteins and these amino acids are being joined by a covalent bond known as the peptide bond. So what is the structure of the amino acids? So amino acids share the common structures, all the 20 amino acids are the alpha amino acid with the common structures.

Each amino acid has a carboxyl group and the amine group attached to the primary carbon which is called as the alpha carbon, they differ from each other in terms of the side chain or the R group. The side chain varies in the structure, chemical nature and that influence the overall property of that particular amino acid. So what you see is this is the generalized structure of the amino acid where we have the C alpha carbon and that C alpha carbon is attached to one side as the amino group, the other side is called as the carboxyl group and that is why the name suggests is the amino acid. So amino acid means the compound which will have the amino group on one side and the acid group on to the other side and in between it is going to be connected to the H as well as the functional group which means the R group. So depending on the R group, the amino acids could be of different types.

The side chains actually varies in terms of the structures, chemical nature and that has influence on the overall property of the amino acid. Each carbon is attached to the four different groups. This is what we have discussed already. This is the C alpha carbon and it is connected to the four different groups making it chiral center to give the stereoisomers because this C alpha carbon is connected to the four different functional groups like the functional group number one which is the amino group, functional number group number two which is the acid group and then it has the hydrogen which is the third group and then it has a side chain or the R group which is the four side. So because of this four different groups attached to this particular carbon, this carbon is actually is a chiral center.

So it actually gives the stereoisomers. There are two common forms of the stereoisomers called as the enantiomers found in the amino acid. These are the non superimposable mirror image to each other. So this is what you see here is actually a enantiomer. So what you see here is actually a mirror image.

So if you see all the groups what is present on this side is present on to the this side in this particular amino acid. This is the example of the L-amino acid versus the D-amino acids. So compared to these all the amino acids only the glycine is the only amino acid which does not have the chiral center because glycine the R chain is H. So that is why the glycine does not show any kind of the chiral center. So amino acids are classified by the R group as this different amino acids are classified based on the side chain or to the R group.

These 20 amino acids are denoted by the first letter or the three letter codes. So these are the few amino acid what you see here. So we have the 20 different types of amino acid depending upon the side chains. For example we have the arginine, glutamine, phenylalanine, tyrosine, tryptophan and so on. So we have the glycine glycine and all that.

And all these amino acids are denoted by the two conventions. One is the single letter conventions or the third and the second is the three letter conventions. So this is very easy to remember. So even how the people are you can actually be able to remember the single letter or the three letter convention is for the single letter you have to just go with the first alphabet of that particular amino acids. For example in the case of alanine.

So alanine the first alphabet is A. So I can say that the alanine single letter code is A. Whereas if I want to go with the three letter codes I can just take the initial three alphabets. So then the three letter code is actually going to be alanine. So the same way you can actually have that other amino acids also like for example you can have the leucine.

So for the leucine if I have to remember the single letter code it is going to be L and if the three letter codes then it is going to be leu. This is not true for all the amino acid because you can also have the amino acids which is also be starting with the A alphabets for example we have the arginine. So if we have the arginine then it is for the these kind of amino acid you cannot take the A because A is already been alerted for the alanine. So what we can do is we can just go with the sound what is coming from this amino acid. So what I mean sound is coming arginine.

So when you speak the arginine the sound comes as R. So you can actually give the single letter code to R. So that is why these are the exceptions. These are the things what you have to remember. For example the arginine you have to remember that the single letter code for arginine is R and the triple letter code is A R G.

Similarly we have already been alerted the L to the leucine. So you can actually cannot alert the L for the lysine for the lysine the single letter code is K and thus the triple letter code is L by S. Okay the same is true for the other amino acid. For example the glycine you can easily take the single letter code as G and the triple letter code as GL by. So for the few amino acids you might have to remember the single letter code or the triple letter code but all other amino acids you are just going to follow the convention that the first letter of that particular amino acid you are going to take as the single letter code.

The name of the amino acids were trivial or the classical in few cases derived from the food source from which they are been isolated the first. This is a scientific way of giving the name of or the single letter or the triple letter code for these amino acids. But when the people have started discovering these amino acids they have given the name based on the food sources from which they have been isolated for the first time. For example the asparagine asparagine was isolated from the asparagus. And the glutamate is isolated from the wheat gluten whereas the tyrosine is isolated from the cheese and in the Greek the tyrosine is in the Greek the tyros means the cheese that is why the name of the amino acid is tyrosine.

Similarly the glycine, glycine is derived for its name due to its sweet taste because in the grief you have the glycos which is called as the sweet. So these are the conventional name which are been given but later on people have started giving the name based on the systematic scientific way of doing it and that is how they have given these kind of names. As I said you know remembering these single letter code or the triple letter code is very important because when we are going to write the sequence of the amino acids or for a particular protein you cannot write the full name. And that is why the single letter or the triple letter codes are being used in that particular cases. Now the amino acids are as I said you know amino acids are classified based on the R group so we have the 20 different amino acids and all these amino acids are varying in terms of the different types of properties.

So what you see here is I have given you a composite table where I have given you the three letter codes or the single letter code. Then as far as the molecular weight is concerned the molecular weight is also depending on the side chain as well as the other groups. So you can see that the molecular weight is 89 in the case of alanine whereas it molecular weight is 204 in the case of tiptofan. So it actually varies a lot but on a generalized way the amino acid molecular weight is considered to be 110 Dalton. So if I say there will be a question that what will be the how many amino acids are present in a 20 KDA protein.

So if this is a generalized question what you can do is the number of amino acid if you want to calculate then what I will do is I will just take the 20,000 KDA 20,000 Dalton and I will divide that by 110 and then what you are going to get is you are going to get the number of amino acid. So that is a generalized term if I say you should calculate the number of amino acids present in the actin protein then and if I give you a sequence of that amino acid or the sequence of that particular protein then the or if I say I give you a sequence of the amino actin protein or you give me the molecular weight then in that case the situation is going to be different then what you have to do is you have to first count the number of amino acids. For example, if I say number of glycines so suppose

the number of glycines are 10 for example so you can just go with the glycine table and then you say glycines molecular weight is 75 so 75 into 10 is 750 Dalton and that is how you are actually going to do for the calculations you are going to calculate the number of other amino acids like how many aspartates are present how many arginines are present how many lysines are present how many tryptophan are present and then you just keep multiplying that and if you add all those numbers then it is actually going to tell you that what is the molecular weight of the actin proteins. So, this is just a generalized term. So, this is the molecular formula what is being given for each and every amino acid then it is the residue formula what is given and then you have the residue weight which is may like you if you remove the water molecule then this is going to be the molecular weight then it is also giving you the pKa values and if the amino acid has two functional group then it is also going to give you the pKb as well and it is also going to give you the pI value.

So, these pKa values are actually going to calculate used to calculate the charge on that particular amino acids whereas the pI is actually going to calculate the charge. So, it is going to let you to calculate the charge of that particular amino acids at that particular pH. So, in a particular pH so, you know that at that particular pH the amino acid is going to be neutral. So, now the amino acids are classified by the R group. So, you can have the different types of R groups you can have the nonpolar aliphatic R groups.

So, the R group in this amino acids are nonpolar and the hydrophobic examples includes are the alanine, valine, leucine, isolecine, glycine, methionine and proline. Then we have the aromatic R groups which means these are the groups where you going to have a benzene ring. So, these groups are going to be contained the benzene ring whereas in this case it is not going to contain the benzene ring it is going to be a linear chains. So, the R group in these groups are hydrophobic side chains for example, the phalanaline, tyrosine and the tryptophan. So, you might have to remember the structure of these amino acids if you want to understand the functional role or the crucial role going to be played by these amino acids.

Then we have the polar uncharged groups which means these amino acids are not going to have the charge which means they are not going to be positive or negative charges, but they are polar in nature. So, the R group in these amino acids are uncharged and they are more polar than the hydrophobic amino acids example includes are serine, threonine, cysteine, asparagine and the glutamine. Then we have the negatively charged R groups which means they are actually going to be called as the acidic amino acid because they are actually going to impart the negative charge into the protein. So, the R group in these amino acids are acidic with the net negative charge examples are the aspartate and the glutamate. Then we have the positively charged R groups and these amino acids are

called as the basic amino acids.

They are actually going to give you the positive charges and R group in these amino acids are basic with the net positive charges examples includes are the arginine and the lysine. Now when you take a protein and you want to calculate you want to know what will be the amino acids are present what you can do is you can just do acid hydrolysis and if you do the acid hydrolysis is actually going to give you the mixture of amino acids. Now once you have the mixture of amino acid you can resolve these amino acids on to a thin layer chromatography. So, thin layer chromatography is a technique which actually is going to give you the spot for the each and every individual amino acids and depending on the intensity of that particular amino acid and depending on the number of spots what you are going to get from this particular protein you can be able to calculate the number of amino acid as well as the type of amino acids present in that particular mixture. So, that you are going to do when you are going to analyze the amino acids.

So, how you are going to do the analysis of the amino acids? The thin layer chromatography technique is an analytical chromatography to separate and analyze the amino acids from the protein. In this method the silica or the alumina as a stationary factor a stationary phase is coated onto a glass or the aluminum foil as a thin layer and then a sample is allowed to run in the presence of the mobile phase. In comparison to other chromatography technique the mobile phase runs from the bottom to top from the by the diffusion whereas in most of the chromatography technique the mobile phase runs from the top to bottom by the gravity. So, what you are going to do is you are going to take a thin aluminum foil and on or the glass plates and then this aluminum foil you are actually going to make a thin film of the alumina or the silica and then what you are going to do is you are going to apply your amino acid which you want to dissolve and then you are going to keep this into a solvent system. So, once the solvent is going to run in from the bottom to top it is actually going to take up the your amino acid also along with that and then it is actually going to give you the spot as per the and it is going to give you the mixture of the amino acids and that is how it is actually going to dissolve.

Now what you see here is in the TLC you are going to have the moment of the solvent from the bottom to top because of the diffusion compared to that in a conventional chromatography you are going to see always running of the mobile phase from the top to bottom. Like example if you run a column right the column you run from the water from top to bottom whereas in the case of thin layer chromatography the solvent runs from the top to bottom because of the diffusion. When a sample runs along with the mobile phase it gets distributed into the solvent phase and the stationary phase right when it runs it is actually going to be you know going to get distributed so either it will go with the solvent fronts right so you can see that all the several molecules are running right. So

either so if this is the molecule either it will go along with the solvent front or it will actually going to remain with the silica front right and because of that it is actually going to be distributed because that kind of affinity and that kind of differential behavior is going to be different for the different amino acid molecules. The interaction of the sample with the stationary phase retard the movement of the molecule whereas the mobile phase implies and as the effective force.

So what will happen is that when you are going to apply the sample onto this it is actually going to experience for example this spot right. So if you have a spot here it is going to experience two different types of forces it is going to have a driving force right. So driving force is actually going to be exerted by the solvent molecule. So solvent molecule is going to try to push this molecule towards the top side because it is running right but whereas it is actually going to make the interaction with the these silica particles which are present on the this TLC plate and because of that there will be a retardation forces. So that retardation forces is going to be by the matrix molecule and because of that it is going to have the two opposite forces one which is going to be on towards the top side and the other one is going to be onto the bottom side.

So it is going to be distributed so and it is going to be immobilized onto the plate. Now suppose the force caused by the mobile phase is Fm and the retardation force by the stationary phase is Fs. So then the effective force on the molecule will be Fm minus Fs. So this is what I am going to I was trying to explain you right. So if this is the spot on this spot you are going to have the upward forces which is called as Fm or the mobile force by the mobile phase whereas it is actually going to have the retardation forces which is going to be caused by the stationary phase which is called as the Fs.

So this molecule is actually going to run effectively by a force which is going to be Fm minus Fs. Now the Fm minus F1 so where this molecule is going to be stopped right it is going to stop the place where the Fm is going to be equivalent to Fs right and that Fm minus Fs is going to be different for the different molecule and that is why they are actually going to be present at the different places. For example for this molecule the Fm is too big for the Fs which means it is still having a effective charge at this point and that is why this molecule runs for the longer period of time but at this point when it got immobilized again for this one also the Fm is equivalent to Fs. So the molecule is immobilized on the silica gel where the Fm is going to be equivalent to the Fs and the position is controlled by the multiple factors. Nature or the functional group what is present on to the molecule or the analyte.

So if the molecule is going to interact with or suppose it has the functional group and it is going to interact with the silica particle then it is actually going to increase the Fs. If it

is going to increase the Fs then the Fm minus Fs is going to be small right and if this is the case then it is actually going to be immobilized towards the spotting points. Then it is also going to be depend on the nature on the composition of the mobile phase. So depending on the mobile phase also it is actually going to be different right if the mobile if the molecule is very soluble into the mobile phase then the Fm the value of Fm forces are actually going to go up right and that is why it is actually going to run very far away from the spotting place right. Then it also depends on to the thickness of the stationary phase because that also is going to have the effect on to the retardation forces and then it also has the functional group.

So functional group what is present on to the stationary phase. So apart from the functional group what is present on to the analyte molecule the functional group if the functional groups are also present on to the silica particles right. For example if you take the silica particle or if you take the functional silica particle they may have the higher affinity for this particular molecule and as a result the Fs will actually go up and if the Fs will go up it is going the Fm minus Fs is going to be small and that is why it is actually going to immobilized very soon and it is going to be closer to the spotting points. So you can imagine that if I want to see the differences like what I can do is or if I want to know where the molecule is going to immobilize what I can do is once it got immobilized I can take a distance from this molecule to the origin.

So this is the origin point. So for example at this point I have started putting the spot of the mixture and then the solvent started running right. So when the solvent reached to the end of the plate or it reaches to at least the 75% then what I can do is I can just stop this I can develop this spot and then I will calculate the distances what is run by the solvent and I can also calculate the distances run by the individual spot. So for example this is the spot number 1 this is the spot number 2 this is the spot number 3 and the distance of the spot number 1 is d1 plus distance of the spot number 2 is d2 and the distance of the spot number 3 is d3. So what I can do is I can just calculate the Rf values and Rf value what is the formula is that the distance of the analyte right distance of the analyte spot which is like d in this case so dx divided by the distance of the solvent. So distance of solvent like the in this case ds okay and this is going to be fractional.

So the maximum Rf what is possible is 1 and it going to be the fraction of 1 right because this is the ratio right and this Rf value is going to be dependent on to the solvent system right depending on the solvent what you have taken depending on the matrix material and depending on the conditions in which it is run. So if you are going to maintain all the 3 constant right if you maintain the same solvent system if you maintain the same matrix and if you run it under the identical conditions the Rf value is not going to be changed even if you run it for for example if I run it for 50 centimeter it is going to

be distributed accordingly. So that for example if I have Rf value of 0.5 okay. So if I run it for 50 centimeter the spot is going to be formed as 20 centimeter 25 centimeter if I run it or 100 centimeter right then the spot is going to be formed at 50 centimeter because it is going to maintain the ratio of Rf is equal to 0.

5. So if Rf is 0.5 which means the ratio of dx to ds is going to be 0.5. So that does not depends on the how much length you are going to run the TLC plates it is going to be always be immobilized at the 50 percent distance right if the Rf value is 0.5. So this Rf value is constant and that side the Rf value can be used to characterize the different types of amino acids.

Now the question comes how you can be able to determine the Rf values you can actually be able to run the TLC plate that you can be able to run the thin layer chromatography and that is how you can be able to calculate. So how you can actually be able to run the thin layer chromatography. So these steps are required to perform a thin layer chromatography to analyze the complex samples these preparative and operational steps are as follows. So in the step number one you are actually going to make the spotting but before that you are actually first going to take a thin layer plate.

So what you see here is this is the thin TLC plate right. So what you are going to do is what you take the TLC plate you cut the TLC plate as per the number of sample what you are going to place right on the width wise and its height is also going to be as per the chamber of your TLC plate right. So you are going to run it in a chamber right and then what you are going to do is you are going to take a scale and you are going to put a line at this line should be above to the solvent front right because it is going to dip right ultimately you are going to dip this. So then you put a line and then on this line you are actually going to put the spots. So the line is drawn with a pencil little away from the bottom the sample is taken into the capillary tube or in a pipette. So what you can do is just take the sample into a capillary tube the capillary is touched onto the silica plate and sample is allowed to dispense.

So what will happen is when you touch the silica when you touch the capillary to the that particular silica it is actually going to suck the sample automatically by the action of diffusion right. So it is important that the depending on the thickness of the layer a suitable volume should be applied spot is allowed to dry in air or a hairdryer can be used instead. Then we have the running of the TLC. So once a spot is dried it is placed in the TLC chamber in such a way that the spot should not be below to the solvent level, solvent level front is allowed to move until the end of the plate.

So what you can going to do is just keep it into a TLC chamber. So you can actually be

able to develop a TLC chamber into a beaker or into a thin small chambers depending on the type of the solvent right. So then you have to cover this with a some you know through the lid right rather you can just put a cap right and why it is important because so that the solvent what you have put is actually going to form the vapor and that vapor should be condensed otherwise it is not going to give you a upward moment. After the so then you are going to place the TLC into this right. So you are going to plate the TLC plate and make sure that this line should be above to the solvent front so that it is not going to get dissolved into this solvent before running right. The analysis of the chromatography plate, the plate is taken out from the chamber and air dried if the compound is colored it forms the spot and these substances there is a no additional staining required.

There are two method of developing the chromatogram. So what you can do is later on you can just take out this plate and then you air dry. So if the analytes like the amino acids are colored then actually they are going to give you a spot and then you can directly take the all the sort of measurements so you can actually know that this is the like the ds right. So this is the solvent what you have run and then you can just take the calculation of this and it is actually going to be dx and then you can actually be able to calculate the RF value by dx by the ds. So for this the position of the spot is very important right. So if it is colored compound there is no need to have the any kind of additional staining procedures or any kind of procedures you can actually be able to do this right.

But if it is not then you have to go with the staining procedures. In the staining procedure the TLC plate is sprayed with a staining reagent to stain the functional group what is present into the compound. For example the ninhydrin is used to stain the amino acids. So if it is not then you can actually be able to use the staining procedures. If you want you can actually go with the non staining procedure as well.

So you can use the non staining procedure. In the non staining procedure spot can be identified by the following method. You can use the autoradiography a TLC plate can be placed along with the x-ray foam for 48 to 72 hours. Exposure time depends on the time and the concentration of the radioactivity and then the x-ray film is processed. So you can actually have the radioactive amino acids and that is actually going to be exposed to the x-ray film and then it is actually going to give you the spot onto the x-ray film and then you can do all the calculations from this particular spot. Like for example you can do the RF calculations you can actually do the dx and ds and that is why you are actually going to get the RF value and that RF value can use for identification of that particular unknown spots.

The second is you can do the fluorescence. So several by heterocyclic compound gives the fluorescence in UV due to the presence of conjugate double bond system. TLC plate can be visualized in the UV chamber to identify the spot. So what you see here is this is the typical UV chamber where you have the UV bulbs and this is the chamber right this is the lid of that particular chamber. So what you can do is just open from here and place the plates under this and then you can just turn on the UV lights. So you can have the two different types of UV lights which you can use as per the wavelength and then from this size because the UV light is dangerous for the eye.

So that is why you can actually be able to observe not directly but through this particular observing window and what you see here is that this all the spots are visible right. So under the UV and then you can actually be able to use the camera or some other acquisition system and that is how you can actually be able to capture the image. Apart from that you can also use the iodine. So you can also use the iodine staining right. So you can actually be able to incubate the TLC plate into the iodine chamber and this iodine chamber is actually going to stain the spots as well.

So the proteins are as we discussed that proteins are the polymer of the amino acid they are joined by the covalent bond known as the peptide bond. A peptide bond is formed between the carboxyl group of the first and the amino group of the second amino acid with the release of the water molecule. So this is you will see that this is the amino acid number 1 and this is the amino acid number 2. So when they will go by with the condensation reactions so what will happen is that the OH of this acid and the H from this is actually going to combine and that is why there will be a bond which is going to be formed between the two amino acids and that is how there will be a loss of water and this is a dehydration reaction.

So it is a dehydration synthesis or the condensation reactions. The peptide bond has partial double bond character due to the resonance and the CN bond is not free to rotate. But the bond between the N to alpha like to N to the alpha and the C to C alpha can be able to rotate through a dry handle angle designated by the phi and psi. So what you can so this peptide bond is rigid it is not allowed to rotate but the bond between the N to alpha and the C to C alpha can be able to rotate through the dry handle angle which are designated as the phi and psi and these angle can be able to rotate from the minus 180 to plus 180 with the few restrictions. To exploit this particular type of phenomena, the Indian scientist G N Ramchandran has determined the possible phi and psi angles for a particular amino acid by synthesizing the tripeptide with the amino acid of interest in the middle.

So what he has done is he has actually synthesized a tripeptide. So for example, if he

wants to calculate the phi and psi angle for A he has made a tripeptide with C and B. So you can keep changing these tripeptide and that is how you can be able to calculate under different conditions what will be the different psi phi angles are possible and that is how you can be able to make a map between the psi versus phi. So you can actually be able to make a angle between the map between the psi and phi and then you can calculate that under different combinations how much these psi and phi angle are going to vary for the A molecule and that is how you can say that okay A will go from this psi angle to this phi angle. And based on that he has actually developed a map which is called as the Ramchandran plot and that Ramchandran plot is used to define the region of the allowed rotation for the amino acid present in a protein structure and he what he proposed is that he you can use this particular type of plot to say whether a solved protein structure is correct or wrong because if it is incorrect then the psi and phi angle are not going to be present in that particular defined regions. So that is what you see here is that you are actually going to see the different regions and it is going to be what you see here is the plot between the between the psi and phi and that shows the location of the different types of structures what is present in the protein structures and so on.

And that is how it is actually going to give you the distribution of that particular amino acid in the protein structures and how much its phi and psi angle are going to vary. So with this brief discussion about the Ramchandran plot and as well as the brief discussion about the amino acid I would like to conclude my lecture here. In our subsequent lecture we are going to discuss about the some more properties structural properties of the proteins. So what we have discussed so far we have discussed about the biochemical properties of the amino acids and we have also discussed in detail about how you can be able to analyze the amino acid using the thin layer chromatography. So with this I would like to conclude my lecture here. Thank you.