Organic Chemistry In Biology And Drug Development
Prof. Amit Basak
Department of Chemistry
Indian Institute of Technology, Kharagpur

Lecture – 01 (NOTE FORM)
A Brief Introduction to Molecules of Life: Structure of Amino Acids and their Various Charged Forms

Welcome, to this course NPTEL online certification course on Organic Chemistry in Biology and Drug Design. Myself, professor Amit Basak, Department of Chemistry and School of Bioscience, IIT Kharagpur will be the chief instructor along with my research scholars who will be assisting this course to conduct the tutorials and to answer the queries and questions that you will have during the during this course and also at the end of this course.

Now, let me first give you the background of this course or the genesis that why we are trying to enter into a field which is the field of biology and also drug discovery process as an organic chemist. Now, you know organic chemistry has started when Wohler synthesized urea from ammonium cyanate which is a landmark discovery because prior to that, it was believed that organic molecules exist only in living systems. But, Wohler showed that an inorganic material, ammonium cyanate could be converted into a substance which is urea that is present in the living systems. So, that basically is the onset of organic chemistry.

For the two next centuries, the organic chemistry mainly concentrated on isolation of molecules from natural sources as well as unnatural sources and their structure elucidation to understand what is their constitution. And then came the subject of stereochemistry; that means, the 3-dimensional orientation of the different groups in the molecule. Along with this, the organic chemists also started to look at the reactivity of molecules; on the parameters that this reactivity depends.

So, it grew in a very rapid pace and as a primary motto of organic chemistry,a big emphasis was put on how to make molecules which is called the branch of organic synthesis. Today virtually, we have different kinds of reactions which allow us to make extremely complicated molecules. So, organic synthesis has reached its pinnacle of success today.

Now, the question is that what after this when organic chemistry has developed so much?. Alongside with chemistry or organic chemistry, another subject was growing at a parallel rate and that was the subject of biology. And, the domain of biology basically comprised of understanding the life processes, what are the different organelles, different cells, different tissues, different organs that are present in the living system and how do they cooperate, what are their functions.

But, the point was that the biologists were looking at the living system at the macro level; that means, what are the components and what are their functions. The more inticrate details that what are the fundamental components that are present in these compartments like the cells or the organs or precisely what are the molecules that are present which cooperate with each other and sustain what is called life.

So, it took a little bit longer time when Linus Pauling gave the structure of the complex or biological molecule called the proteins and that basically started the era of looking at biological system from molecular perspective. Linus Pauling precisely gave the structure of proteins and he gave the α helical concept that proteins can fold in a helical manner and then subsequently Watson and Crick they revealed their landmark discoveries on the double helix structure of the double-stranded DNA.

So, that was basically the starting point of a new type of subject when people felt that yes, we have to look at the biological system from molecular perspective and when you talk about molecules, these molecules are present in the living system and when these molecules are present in the living systems; that means, they are basically the organic compounds; that means, they are carbon based compounds.

Along with carbon we know that they have hydrogen, phosphorus, nitrogen, oxygen (these are the main elements that are present). So, that was the start of the looking at biological system from molecular perspective and when you talk about molecules, then organic chemistry has a big role to play.

So, that is why as today's chemists, the new brand of chemists, we have to generate one who must know, not only the basics of organic chemistry (mainly the structure, their function, the stereochemistry, the reactivity, the synthesis, the bond forming processes they have to know all these things), but alongside, the new brand of chemists should also know how biological processes take place in the living system.

Remember, interestingly these reactions, which are taking place in biology or the molecules which are there present in biological system, in a living system, are not, different because they are basically organic molecules and the chemistry that takes place also they are nothing different, but organic chemistry involving established organic reaction pathways. Only difference lies in the kind of environment that is present in biology as compared to the organic chemistry done in the laboratory- that is the basic difference.

We do organic reactions in test tubes or in flasks and so that is a very isolated system. But, in the biological system, the biological reactions take place alongside with the surroundings; so many molecules, so many catalysts, so many macromolecules are present in the system living system, but still we can have only one reaction in presence of so many reactive molecules.

Now, these reactions are done by what we can call the carpenters of biology; that means, they do the synthesis or they do the breakdown of big molecules. So, they are basically the biological carpenters and they work in perfect harmony. The harmony; (harmony means one after another) is maintained, i.e, it is a harmonized process; and once the harmony breaks down, then what happens? Then we suffer from lot of problems; what are called the diseases. So, the diseases are basically related to this harmony that takes place in the biological system.

So, I think now it is amply clear that today the chemists have to know the basics of biology and what type of chemistry that goes on in the biological system. At the same time the biologists now should look at the biological system from the standpoint of molecules; that what are the exact microscopic details that take place when something is synthesized or when something is broken down or when a signal has to pass from one cell to the another, ok? So, that is basically the genesis of this course.

So, I realized that this type of course, is extremely essential for today's chemists and biologists. Just one point to add, that around 1990s, world renowned universities, they have changed the name of their department from department of chemistry. Harvard University renamed their department to Department of Chemistry and Chemical Biology. So, they realized at the time that this new brand of chemists should be called as chemical biologists. I think that justifies that why this type of course will be important and I will try my level best to show you organic chemistry in biological systems.

And, then after we do the biological system study from organic chemistry point of view, that will generate a platform from where we can utilize our knowledge and then show that how different drugs have been designed, synthesized and then now they are extremely useful. And, so, you have a clear perspective of the whole process that from chemistry to organic chemistry and then organic chemistry to biological processes and from biological processes we will go to the drug design. So, there is a chain of events that will happen in this course or during the deliberation of this course.

Now, let us talk about the biological system as I said we will look from the molecular perspective. Now, biological molecules are the molecules which are also called the molecules of life; they are mainly comprised of four types. These four types of molecules which are present in all living systems (with few exceptions possibly) are proteins, nucleic acids, lipids and carbohydrates. Now, most of these are called biological macromolecules. However, I can say that macromolecules are basically made up of small molecules. So, small molecules join and finally, give the macromolecules. So, obviously, when we talk about these macromolecules, we should talk about the building blocks of these systems.

Now, all are not macro molecules per se like if I talk about lipids. The lipids are basically molecules which are having molecular weight less than say 1000. So, they cannot be called macromolecules. But proteins, nucleic acids and carbohydrates are all basically macromolecules. There could be small carbohydrates like glucose, fructose or sucrose; or there could be large branch of carbohydrate molecules tied together like what is present in starch or cellulose.

Now, apart from these four basic classes of molecules, there could be small molecules, there could be medium sized molecules. Their role is also extremely important in maintaining the coordination between the cells; coordination means it is called a process which is signal transduction; that means, one cell gives a signal that signal goes to another cell.

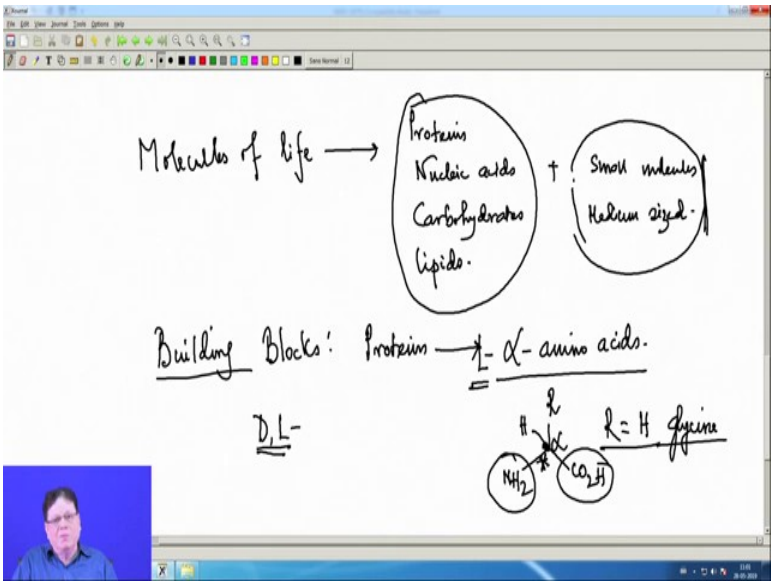
Now, who actually creates these signals and also transmits these signals? So, in that case, there are several small molecules which play a role, but these small molecules are mainly generated from some of the other molecules which are present in the biological system. Now, when we say that a small molecule is generated in a biological system, we have a specific term for that process that is called biosynthesis of molecules.

Biosynthesis is nothing, but synthesis; but synthesis taking place in a biological system. And, the term synthesis is specifically reserved for organic chemists to do their reactions. So, we will cover all the aspects, all these four class of molecules, along with several biomolecules which are small, which have small molecular weight.

Now, let us come to the building blocks of the molecules of life. Today, it will be just a preliminary aspect of all these molecules and then we will go deeper and deeper on the structure and function of these molecules. Now, these molecules of life, the first one I told you about are the proteins. Now, proteins you already know what are proteins? We know that their basic building blocks are amino acids.

Now, these amino acids can be of various types; when we say amino acids, amino acids are basically compounds where there is an amine and a carboxylic acid group. But, these amino acids which are present in the biological system to make the proteins, they have a special character. They are called α amino acids. α amino acids means the amine is present α to the carboxyl group.

(Refer Slide Time: 17:53)



Basically we have the following molecules of life: Proteins, nucleic acids, carbohydrates and lipids. But, remember these are the large macromolecules. Alongside we have small molecules and also medium sized molecules. It will be very interesting to know that some of these small molecules are also extremely essential to maintain the life processes.

Now, we were talking about the building blocks of these large molecules; and the first one we took up was proteins. Now, we know that the proteins are made up of say only amino acids that are not enough; you have to say that they are made up of α amino acids. Again that is not sufficient, α amino acids, I can draw the structure is basically an R and then NH_2 and CO_2H. Why it is α? Because this carbon is the α carbon. However, if you look at this molecule, you see that there is a centre, there is a hydrogen here also. So, this carbon is what is called a stereogenic centre or the classical nomenclature is a chiral centre.

So that means, this can exist in two forms, two enantiomeric forms and interestingly in all α amino acids which are present in the living system to make the proteins, they belong to the class of what is called L-amino acids. So, they have a configuration which is L-amino acid. So, what are the components of the proteins' building blocks? You have to say that they are made up of L-α amino acids.

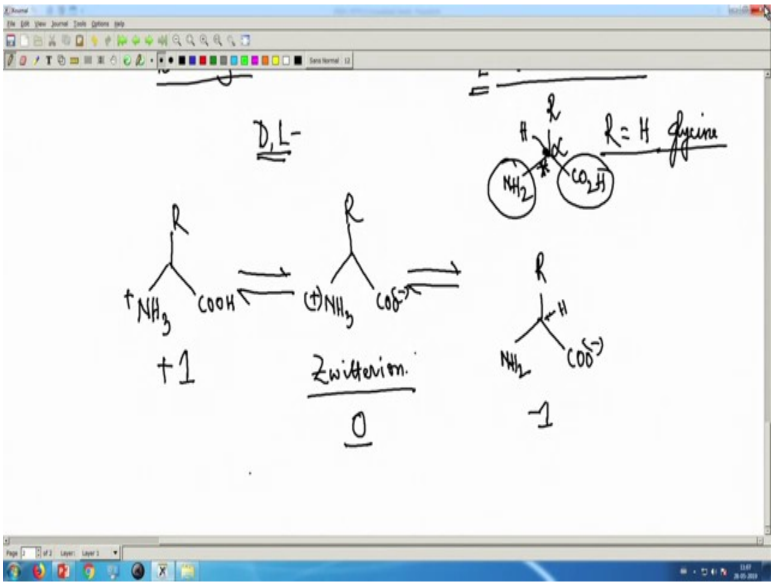
Now, you possibly know that the number of amino acids that are present in proteins are 20; 20 α amino acids we know. Except glycine where R is equal to H, that is your glycine. This glycine because of the lack of R; R means some alkyl or aryl groups what you have is hydrogen. So, that makes it a non stereogenic carbon and if it is non stereogenic so, there is no question of stereochemistry here.

So, glycine is the only achiral amino acid that is present in proteins. All other amino acids are chiral and they have the L configuration. Now, this D, L nomenclature for assigning absolute configuration is little bit older version of telling the absolute configuration. Now, this has been replaced by R, S nomenclature in organic chemistry. However, biologists still prefer this D, L nomenclature in organic to say the absolute configuration of the bio molecules which are chiral.

Now, this L configuration can be translated into R, S nomenclature and if you do that you will find that this 19 amino acids which are chiral; these 19 amino acids except cysteine, (there is one amino acid which is a sulphur containing amino acid, named cysteine), all other amino acids are having S configuration. So, L amino acid means S amino acid, but remember except cysteine and glycine which does not have any chiral centre. So, this is the building block of proteins- α amino acids.

Now, we know that the amine is basic functionality and carboxy is an acid functionality. So, in a molecule you have both acid and base present. So, that presents an interesting situation. So, there can be internal hydrogen transfer between them, and so NH_2 can be present as NH_3 plus and CO_2H will be present as CO_2 minus, ok.

(Refer Slide Time: 23:53)



So, now, if you take an amino acid and put it in an acidic medium suppose a pH say 1; at pH 1 aqueous solution, then what we have? Then the NH_2 will basically be protonated because this is an amine and that will be protonated and so, that will be NH_3 plus. So, if an amino acid is taken in an acidic medium suppose which has got a pH of 1, then what will happen this NH_2 group? Being a basic group, that will be present as NH_3 plus.

The carboxylic acid will be present as CO_2H because at this pH, the carboxylate ion cannot be present. Because carboxylic acid has its own pK_a value, own pK_a means the pH at which the dissociation takes place. So, below the pK_a, it will be present as CO_2H and then suppose you have this R group. So, at acidic pH, this is the situation and as you raise the pH; that means, you are adding alkali. So, what will happen? So, now, the question is you have two acidic hydrogen this is the CO_2H, the hydrogen is acidic and this NH_3 plus, this hydrogen is also acidic. So, one of these will first come out and obviously, this CO_2H has a lower pK_a value, so that will come out that will be abstracted by the base and so that will be present as CO_2 minus and this will be the structure of the amino acid.

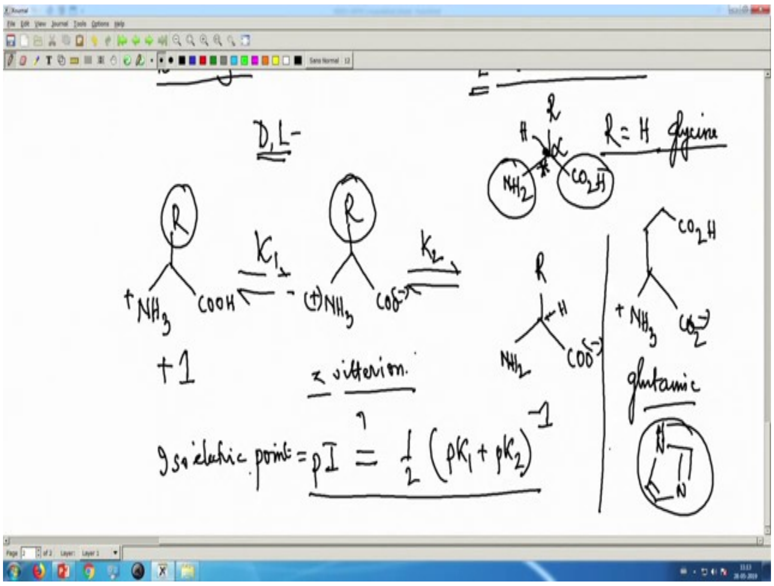
So, now you have a molecule where plus and minus both are present. This is what is called zwitterion, so that is the zwitterion structure of amino acid. Remember this zwitterion will be present at a particular pH and then if you increase the pH further what will happen? Next proton that can be abstracted from the NH_3 plus. So, that will be abstracted and in the process and you will get the NH_2, now that the hydrogen has been abstracted by the base and that will be CO_2 minus.

So, this is the end, after that you do not have any hydrogen to pick up. There is a hydrogen here, but its pK_a is much way above which cannot be abstracted by a simple base as such, ok. So, these are the two ionizable hydrogens that are present in the system.

Now, this dissociation from the protonated form (where the total charge of the system is plus 1), will give the speciese where the total charge of the system is 0; and after further deprotonation, the total charge of the system is minus 1. So, it is fluctuating from plus 1 to minus 1, but going via a zwitterion which has got a 0 net Now, in case of amino acid, if you start titrating it from pH 1. Now, at some pH what will happen? It will be present in the form in the zwitterion form and at that point or at that pH, the whole system is electrically neutral. So, if you want to put this amino acid suppose in a gel and then you try to start applying a voltage, a plus on one side and minus on one side, and put the amino acid in the middle then what will happen? Depending on the charge there whether it is a negative charge or positive charge, it will start moving in one direction or the other.

But if it is present in the zwitterion form, then what will happen, it will not move, it will stay there. Now, what will be the pH at which this zwitterion is present? We can deduce the pH later on, there is a formula for that.

(Refer Slide Time: 28:53)



The pH is basically pI, this is what is called isoelectric point. pI becomes half of pK_1 plus pK_2. What is pK_1? pK_1 is this one the first dissociation this is K_1 this is K_2. Remember these are not rate constants, these are equilibrium constant between these two species. So, that is the general relation of pI. Remember, what is pI? pI is called isoelectric point. Isoelectric point is is not a point, this is a pH.

The pH at which it is electrically neutral it does not move on any side when you apply voltage on this molecule which is now applied on a matrix in which it can move, ok. Now, this is a very important parameter of amino acids. Here I have taken the amino acid where there is only the α amine present and the carboxy present.

Now, what about this R? Depending on this R, we have different classes of amino acids; What type of different Rs are present in amino acids? R can be uncharged, a non-polar group and these amino acids will be called non-polar amino acids. So, the amino acids are classified according to the nature of the R. Some amino acids are called non polar amino acids and that means, the R is non-polar say a methyl which is alanine and if it is isopropyl that is called valine, if it is CH_2Ph that is called phenylalanine.

So, different types are there, but they are all non-polar amino acids. Then, that means, this side chain is such that they do not like water to be surrounding them, ok. So, they are hydrophobic in nature. If R is a group which is polar group, but not charged polar group;

polar group means that there is a dipole moment present there. And, then that can form hydrogen bonds with surrounding molecules which can have donor hydrogens present.

And, this R could be CH_2OH having a hydroxy group, the R could be CH_2SH having a SH group. So, or it could be $CH(CH_3)OH$ that is what is called threonine and these OH groups or SH groups have the ability to participate, first of all they are polar groups because we know that sulphur oxygen they are more electronegative than carbon. So, that generates a polarity and this generation of polarity creates, or it gives an ability to the atoms like oxygen or the sulphur to form, to participate in hydrogen bonds. So, they are the amino acids with polar groups, but uncharged.

Then, you have amino acids where this R group contains another acidic group and like you have NH_3 plus say CO_2 minus. I am talking about just writing the zwitterionic form and you can have another acid group a CO_2H group, so an extra carboxy group. Now, this amino particular amino acid has 1, 2, 3, 4, 5 carbons. So, that is called glutamic acid. So, they are amino acids containing acidic functionality. Acidic functionality means in the side chain. That is called the side chain of the amino acid and you can have slightly a one carbon less also, the lower homologue of glutamic acid which is called aspartic acid. So, these are the acidic amino acids.

Then you have just the reverse; the basic amino acids where instead of carboxy group you have other different amino acids which contain either NH_2 or it could be what is called imidazole group. Imidazole group is functionality. This is called the imidazole functionality. So, this is a basic group. So, now, depending on whether there is NH_2 present or imidazole present or it could be arginine; that is another functionality arginine group; that is called a basic amino group. So, these are called the basic amino acids. So, the basic amino acids are mainly lysine and arginine.

There is another amino acid which is called histidine which is much less basic than arginine or lysine. So, the strongest basic amino acids are your lysine and arginine. So, these classifications are there for amino acids. So, now, what we have learnt that first of all there are four building blocks of the life which are called the proteins, the nucleic acids, the carbohydrates and the lipids. And, then we said that alongside these large molecules or these four major class of molecules, you have small molecules which are also extremely important and they are generated inside the body biosynthesized. Or some other are taken also I forgot to mention that some of these molecules come from our diet. They are extremely essential again to maintain the life.

And, then we talked about the building blocks- the monomeric systems which makes this biological macromolecules like proteins or nucleic acids or carbohydrates; we have only talked about proteins. Proteins are made up of combination of L-α amino acids except glycine; glycine does not have a chirality and they are joined together through bonds which we will talk about they are joined together by amide bonds and then give different varieties. You have 20 amino acids, so there are different permutations, combinations that you can make and that will make all these different diverse array of proteins that are present in the body.

And they have different functions. Some proteins have catalytic power and they are called enzymes, some proteins could be structural like the collagen, which maintains the muscles or some proteins can be hormones also, like insulin; some proteins could be structural like they are present in the membrane surrounding the different cell contents. But discussion on functions will come later; first let us focus on the structure of these proteins.

So, the most important point of amino acids is the isoelectric point because isoelectric point will ultimately decide the charge when different amino acids are present; the charge of a protein at a certain pH that can be calculated only on the basis of the isoelectric point. What is the definition? The definition is that isoelectric point is a pH when the amino acids are electrically neutral and do not move in an electric field. So, this is the first half of the lecture on amino acids, we will talk about the other issues of amino acids in the second lecture.

Thank you.