Essentials of Biomolecules: Nucleic Acids, Peptides and Carbohydrates

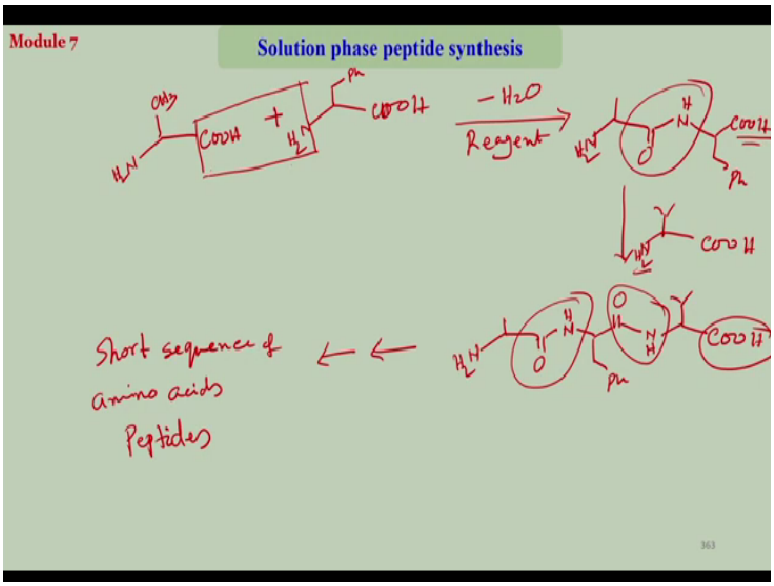
Prof. Dr. Lal Mohan Kundu
Department of Chemistry
Indian Institute of Technology-Guwahati

Lecture-26
Peptide Synthesis and Therapeutics

Hello, everybody and welcome back to the lectures. So, we have been discussing about the protein sequencing, and that was in module 6. So, actually the title of the module 6 was protein sequencing and solid phase peptide synthesis. We have completed the protein sequencing and I think it is a good idea to break this module into 2. So, today we will start with module 7 and that is regarding the synthesis of the peptide including the solid phase peptide synthesis.

So, module 7 will be about chemical synthesis of peptides and how the peptides could be used as therapeutics. So, if you have 2 amino acids, how a peptide bond be synthesized.

(Refer Slide Time: 01:28)



Let us take the example of maybe an alanine which has the CH 3 group. This is your NH 2. This is your carboxylic acid. Of course it exists in zwitterionic form, but I am writing the neutral 1. And you take another amino acid, let us say phenylalanine acid and then H 2. So if you combines them together, it will release 1 molecule of water. Specifically reaction between these carbonyl and this amine acid base reaction, eliminating 1 molecule of water.

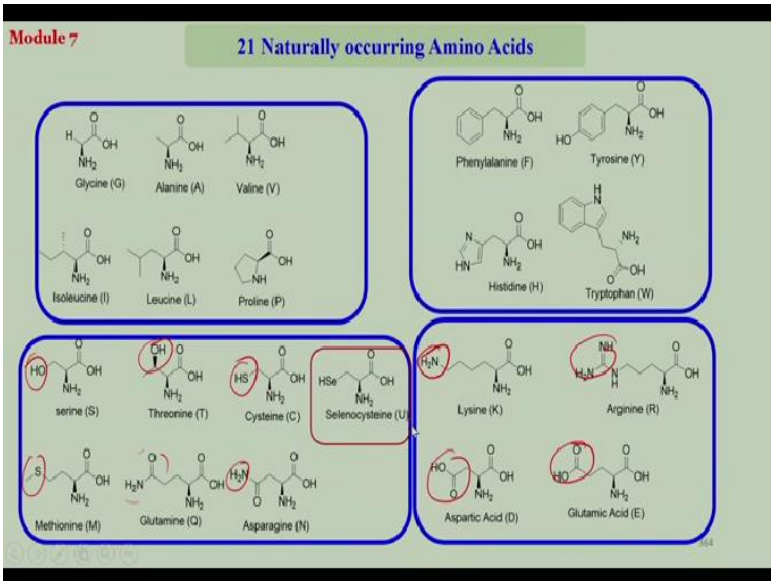
But this reaction does not happen easily. If you just mix the 2 amino acids together, they simply do not react. Amino acids are not very reactive. So you need certain reagents to make them reactive. So you need a reagent for this reaction. I will talk about the reagent just a little bit later. So if you fuse these 2 amino acid you can expect to get a peptide bond. This is alanine. This is phenylalanine.

So you get a dipeptide, dipeptide means it is a single peptide bond, which has 2 amino acids. And then you have a free carboxylic acid group. Now again, if you add another amino acid to these, let us say valine acid group, amine, then again, you can expect a condensation here. Phenylalanine this acid will become a peptide bond and you will have valine with another free carboxylic acid group.

I know there are problems here. I am not talking about those right now, if you go on so, you have 3 amino acids linked together now, if you go on like this, then you can expect to synthesize a short sequence of peptide amino acids which we call peptides. So, this is a chemical synthesis and nowadays we can synthesize a 30 40 50 60 to 70 amino acids long peptides can be synthesized in the laboratory.

And I will tell you later that most of those long chain synthesis we usually do in solid phase that is called the solid phase peptide synthesis. I will come to that later. So that is how I am this is a rough sketch scheme, how to synthesize the peptides in the laboratory. So, the question is, why do we need to synthesize the peptides.

(Refer Slide Time: 05:37)



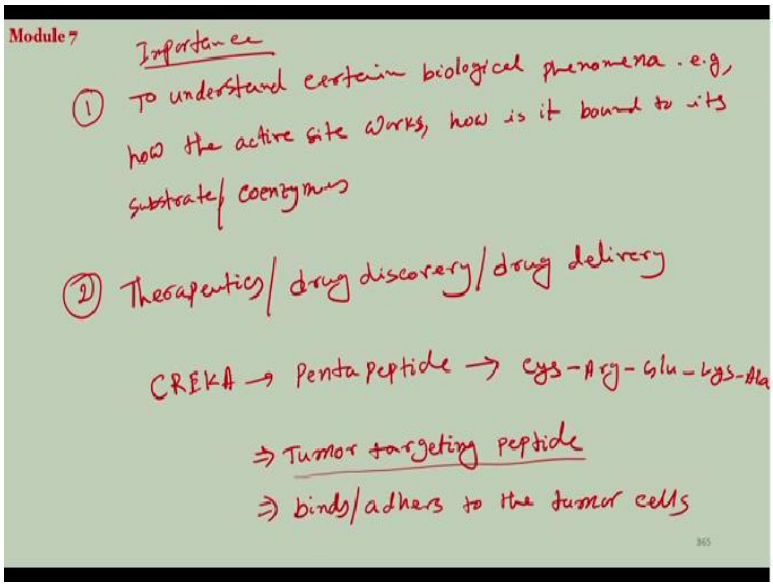
What is the need for the short peptide synthesis okay, before that. So, your target is to synthesize a peptide bond and that occurs by the fusion or by condensation often carboxylic acid group and the amine group. Now if you look at the structure of all the amino acids that are present, you can see a lot of variations in their functional groups or in their structures. Of course, mainly in the side chain here R group varies a lot and there are plenty of different functional groups present in different amino acids.

Peptide synthesis is very different compared to the DNA synthesis we have done. For DNA the nucleobases itself, they do not have many functional groups, those lot of variations exist in amino acids. There are basic side chains, there are acidic side chains. Here also amide bonds are there. Sulfur groups are present, which are reactive, the alcohols, OH functional groups and so on. So, whenever you are doing a peptide synthesis, you also have to take care of these functional groups of the side chains.

So, that is one chemistry aspect in the perspective of organic synthesis, that whenever your N is to synthesize only the peptide bond, you have to be very careful about the side chains. So, that they do not react or they do not interfere into your synthesis. But, on the other hand, this also gives you immense opportunity to vary the modifications to vary the structure of peptides. You can use these functional groups to make new kinds of peptides in a branched way or whatever way you like you have the way out to make mold to modify your peptides.

So that you can synthesize new molecular probes or new molecules. So, the question is why do you need to synthesize the smaller version of peptides, the shorter sequences of the amino acids, they are not proteins.

(Refer Slide Time: 08:18)



So, the first importance of it, the first one is that of course, if you want to study the biological property of a protein then you can use instead of the whole protein, you can use the short sequence of the area of interest of the protein and to use it and use that to study the or to understand the biological properties or the biological functions of the protein. For example, you want to know how the active side of the protein is bound to its substrate.

How it is bound to its cofactors, what is the chemistry behind the catalysis of the within the active site of the protein and the substrates that are coming in. So, in order to know that many times if you take the whole protein, the things becomes very complicated, the chemistry becomes very complicated, the analysis of the data will becomes very complicated. So, if you use instead of the whole protein.

If you use the peptide which constitutes the active site of the protein, then you can use that only the short sequence of it and to study the binding properties or the other physiological properties that you want to know. And data becomes very easy to analyze because you have a short

sequence is things get easier for handling purpose also. So, that is one way, one aspect or one importance of the smaller peptides that to understand certain biological phenomena.

For example, how the active site, this is just one example of it, there can be plenty of other biological functions, how active site works or how is it bound to it is substrate or coenzymes, all these are very essential to understand of course, that how the catalysis works, what are the properties of the binding between the substrate and the peptide and the protein active site, which kind of coenzymes are involved.

How the enzymes are stacked on to the pocket of the protein. So, in order to understand that, you can use these things and another very important aspect of using the short peptides is that you can do crystallography very easily, rather easily compared to the whole protein. Whole protein is very, very tough to create crystallize. So, if you want to see actually the structure of the active side substrate combination in a frozen form, then you have to go to the crystallography.

You have to form the crystals and then the short sequence of the peptides helps a lot in getting a good crystals. So, that is one aspect of it. Second is there are many, many, many short peptide sequences that we know of nowadays that have therapeutic purpose that can be used to cure as well as to take other molecules into the different targeted sites of the disease cells. So, the second aspect is therapeutics.

Not only therapeutics if you want to do a certain imaging you want to take the fluorescent molecule or other markers of a certain part in the cell, then sometimes peptides also help you to take those molecules to the specific positions. So, there are many peptides nowadays, which are known and which have tremendous applications in drug discovery and drug delivery. Therapeutic drug discovery, drug delivery.

For example Creka C r e k a. This is a peptide sequence and of course, you can see it is do you remember the alphabets which amino acids they represent you to look at the here you can find out actually, here, all these alphabets are written, symbols are written. So, it is a good practice to

know or to remember all the alphabets which represent the amino acid, because that is how usually a protein sequence is written, proteins is too long.

So, you cannot write the individual amino acids name, it will take too long time. So, it is represented usually if you see a protein structure or protein, primary sequence somewhere, you will see these kinds of alphabets are written. Of course, they represent the amino acids. So Creka 1 2 3 4 5 is a pentapeptide. What is the sequence C for Cystine, R is arginine, E is glutamic acid, K is lysine and A is alanine. So, this is the sequence of this peptide pentapeptide.

And this peptide has found enormous importance. This is actually a tumor targeting peptide. You can write tumor or tomour, tomour targeting peptides. That means this peptide can go and adhere to a tumor cell because tomour cells have receptors to receive where tomour cells have certain receptors, where this peptide can go and bind. So the normal cells does not have them, specifically, tumor cells have them.

So, if you have this peptide and if you try to send it then this peptide will specifically go and be bound to tumor cells. So this will be this peptide binds or adheres to the tumor cells, because the tumor cells have the receptors to dock this pentapeptides and of course, you can find now that how important this would be, it is usually very, very, very difficult to differentiate between a normal cell and a tomour cell.

So, therefore, whenever you send it try to send a drug to cure the tumor cells or to kill the tumor cells, they also kill the normal cells. So, of course, our aim all the time has been, how to make a difference, how to take the drug only to the tumor cells, which will not affect the normal cells. That has been still going on. I mean there are plenty of that is the major problem associated with tumors or associated with cancer.

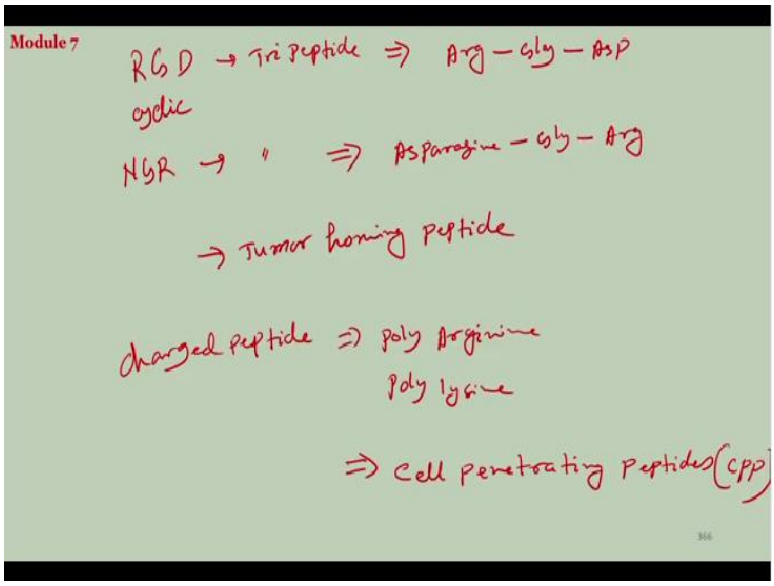
That all the drugs, all the therapies, all the chemotherapies, all the radiations that are there, all those treatment techniques that are available. They are also harmful, very harmful, in fact, for the normal cells and that is a reason that the tumor cells cannot be selectively killed. So, if you as I

was saying, if you want to take a drug to specifically to tumor cells, there are not many ways. It also goes to the normal cells and kill the normal cells.

So now, if this peptide only goes or binds to the tumor cell, then the advantage is, if you conjugate a drug here, or if you attach a drug along with this, then what will happen. The drug along with the peptide can now go specifically to the tumor cells. So, that is called the targeted drug delivery. So, such kind of short peptide sequence because of its tumor targeting property can be very useful in therapeutics.

Not only therapeutics for our academic purposes whenever you want to understand something, you want to study a phenomena, you want to study a physical property or biological property that is going on into the tumor cells. And you want to take certain other molecule, for example, an inhibitor or a fluorescence marker molecule to it, you can always attach that here and this will go specifically to the tumor cells to give you the reports.

(Refer Slide Time: 19:29)



Similarly, another peptide which is a tri peptide RGD is a tri peptide. What is the sequence, R is arginine, G is glycine and D is aspartic acid, so it is a tripeptide, but often this exists in a cyclic form. RGD peptides often exist in a cyclic form and it can have a polymorphic structure, also polymorphic cyclic, maybe a tri 3 member of RGD can be present. So, this peptide RGD or

another one called NGR is also a tripeptide which has the sequence of any aspergin writing the full form here to differentiate between aspartic acid.

G is your glycine and R is arginine. So, very close structures actually. Very close sequences. This as arginine, this as arginine, this as a glycine, this as glycine, only aspartic acid and aspergin are different, these 2 peptides are known as tumor forming peptide almost similar to Creka. These peptides also have receptors in tumor cells. So they can go and specifically bind to the tumor cells.

Tumor forming peptide means they can go and stay with the tumor cells. So, again, similarly, this peptide sequences can also be applied or can also be very, very useful in understanding the biological functions that is going on in the tumor cells, as well as to carry drug molecules to the tumor cells and do a treatment. These kinds of peptides I have talked, there are other examples also. These are the most well studied peptides that are used to target the tumor cells.

Now, another kind of peptide sequences mostly the charged peptides, which has multiple positive charge, for example, poly arginine sequence or poly lysine sequences, arginine and lysine both have amino acid chains. So, obviously, they will exist as in the zwitterionic form, they will exist as H 3 plus that has positive charge. If you have a polymer of argentine, if you have polymer of lysine, they will impose a lot of positive charge in your peptide sequence.

And those charged polar peptides are very good in penetrating the cells. So, they are known as cell penetrating peptides. In short, they are called CPP. So those kinds of molecules with high positive charge and polar side chains can cross the cell barriers can penetrate into the cells. And that is very important because our cell walls are pretty protective, pretty strong, usually, cells do not allow any foreign bodies or any foreign entities to cross into its membrane, because they are outsiders, and they can harm the cells.

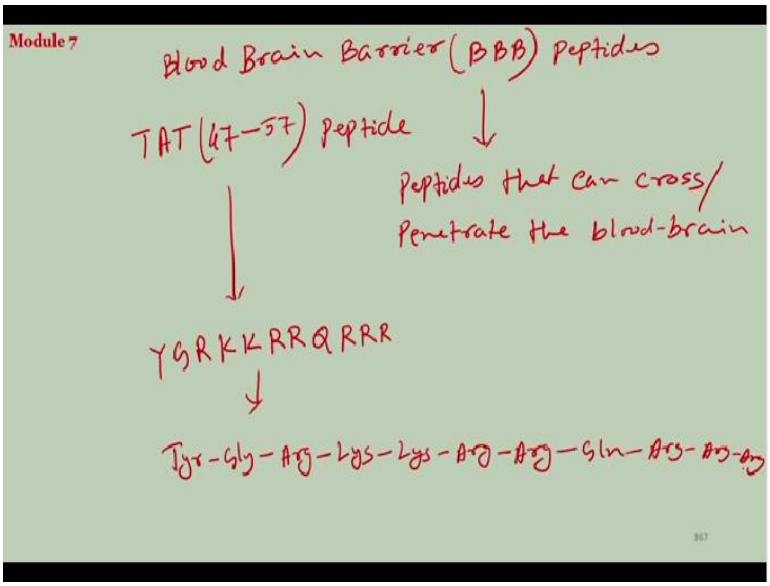
So, cells are very protective, not to allow any foreign any unknown bodies or any unknown molecules to penetrate into the cell. And that is where most of the drugs that you use do not go and do not get through the cells. So, only a certain percentage of the drugs that we actually use

can penetrate into the cells and can do its functions, rest of the drug are actually eluted out or will be coming out excluded out of your body, and they do not go into the cell.

So in order to increase the bioavailability, in order to increase the amount of your drug inside the cells you need something that will help you to cross into the cell barriers. And that is why these peptides are for CPP. So, if you attach drugs here, if you take other molecules that you want to take into the cell, this is a good idea that you can combine the cell penetrating peptide along with those molecules.

And then it can be together they can get into the cells. And this thing self nutrition becomes much more harder when you think of brain cells. So our brain cells brain is the most important part of our body. So brain cells are over protective not to take any molecules from outside. It is very hard to send drug molecules or medicine into the brain cells, they simply get rejected. So, there are blood brain barrier peptides.

(Refer Slide Time: 25:33)



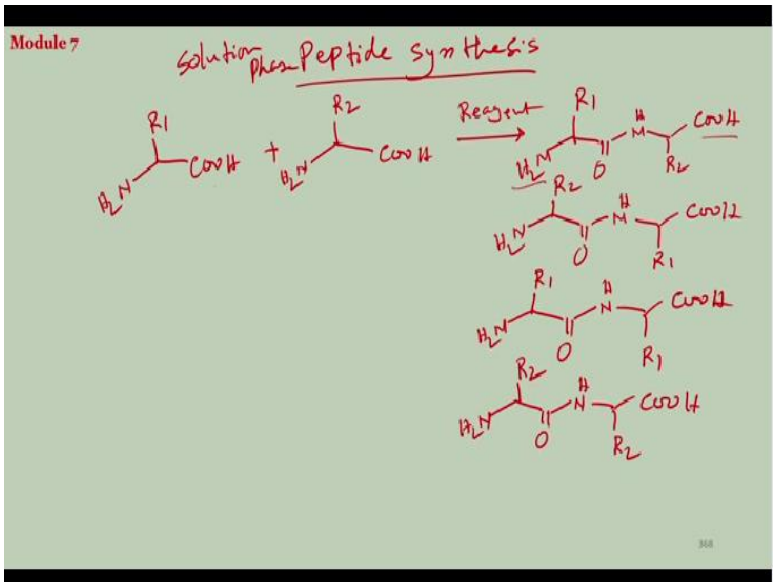
In short, they are called BBB peptides. So, those peptides can cross the blood cells and therefore, obviously, they have really huge importance in biology, in drug delivery, in pharmaceuticals. So, again there are many sequences that are known nowadays, some of them are very active, some of them are not that active. So, one of these is a TAT peptide, TAT is actually a whole sequence of a protein.

And particular part of it the sequence 47 to 57 this 10 more sequence I think in the whole TAT sequence. So, I call it TAT peptide that is a blood brain barrier peptide, it can cross the blood brain. So, blood brain barrier peptides are peptides, that can cross or penetrate the brain cells or blood brain that can actually penetrate the brain cells. So this peptide, a certain sequence of TAT peptides is I will write down this sequence of it, YGRKKRRQRRR.

So you can see here a lot of R argentine. So as I was saying that any self nutrition you need more positive charged, positively charged amino acids. So therefore, the number of arginine should be more polar side chains. What is the full form Y is your tyrosine, glycine, arginine, K is lysine, so it is full of lysine and arginine. Again K is lysine, arginine, arginine. Q is your glutamine GLN glutamine, arginine, arginine, arginine 1 2 3 4 5 6 7 8 9 10 11.

So it is 11 more peptide sequence that will allow you to cross the blood brain cells. So, these are some of the important reasons why we need to synthesize peptides that I have talked about. There are many others. So, now coming back to the synthesis.

(Refer Slide Time: 29:24)



So, peptide synthesis, as I have said, if you want to synthesize a long sequence of peptides, then you go for solid phase peptide synthesis, but in the short sequences of the peptides for example, 3 4 5 up to 10 even, we can synthesize in solution phase itself. So, let us start with the solution

phase peptide synthesis first. So, they should be solution phase peptide synthesis for the time being. If you do it in solution, if you take 2 amino acids again let us say you have R 1 here and R 2.

We had reagent, what is your product going to be in solution, you have both the molecules. So, obviously there will be condensation here R 2, this is 1 is the second possibility. So, in this case, the acid of R 1 is reacting with the amine of R 2, what can happen the other way around the acid of the R 2 can react with the amine of R 1. So, in this case you will have R 2 R 1. This is a different molecule, this is a different molecule. So, we have already 2 possibilities.

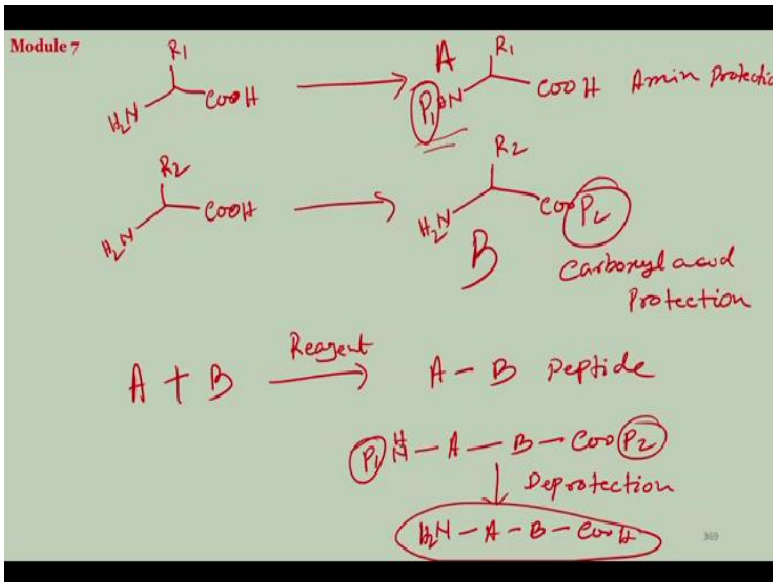
Is there any other possibility of course they are, they can have, they can go self condensation. So one molecule of R 1 can react with the other molecule of R 1 itself to give you 1 peptide R 1, so it is a homodimer, so both are R 1. This is also possible. Similarly for the R 2. Both are R 2. So you have at least 4 possibilities. If you do a solution phase reaction between 2 amino acid, I am just talking about 2 amino acid.

It is to synthesize a dipeptide and all 4 will be formed in your solution. Now, we have to separate them out and take and then of course, you can see that separation will be a very complicated factor, because all of these are dipeptide. So, lengthwise they are same, they have same number of free acid and same number of free amine groups. So, they are polarity will be roughly the same. If the side chains are not very different, then the separation would be very difficult actually, if you have that many possibilities of product formations.

So, what is the way out, if you synthesize if you want to do it in solution phase itself, just dipeptide first, what is the way out. Of course, the way out is to do selective protection. For example, if I want to have only R 1 R 2 this molecule then I have to inactivate this amine and I have to inactivate this acid. So, selective production of the amine group here, selective production of the acid group here.

And then mix them together with the reagent you can expect to get your desired product. So first how to do the amine protection.

(Refer Slide Time: 34:29)

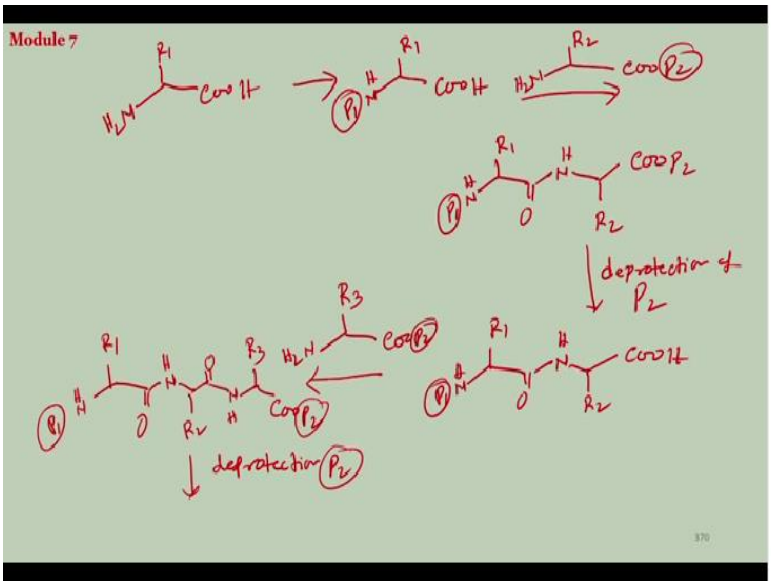


So, if you have so this is our desired product therefore, what you have to do is keep the acid group free and protect the amine. I am terming it as P 1. This is amine protection that you had to do for the first amino acid, what about R 2. For R 2 you need the amine group here. Because there is acid free here an amine should be free. And your carboxylate should be protected. So that it does not react, so this should be carboxyl acid protection.

So, 2 different amino acids should be protected selectively. And now, let us say this is your A and I termed this as B. Then if you now mix A and B in presence of the reagent then you can expect A-B peptide and of course in this A there should be acid protection, there would be debase protection and then deprotection. So, these are the additional synthetic steps that you have to use for the deprotection of this first and then the deprotection of these 2 different regions obviously would be required.

Then you will get back your desired peptide, this is for dipeptide. Now, if you want to go for the longer chain peptide sequences then what are the possibilities to do them.

(Refer Slide Time: 37:44)



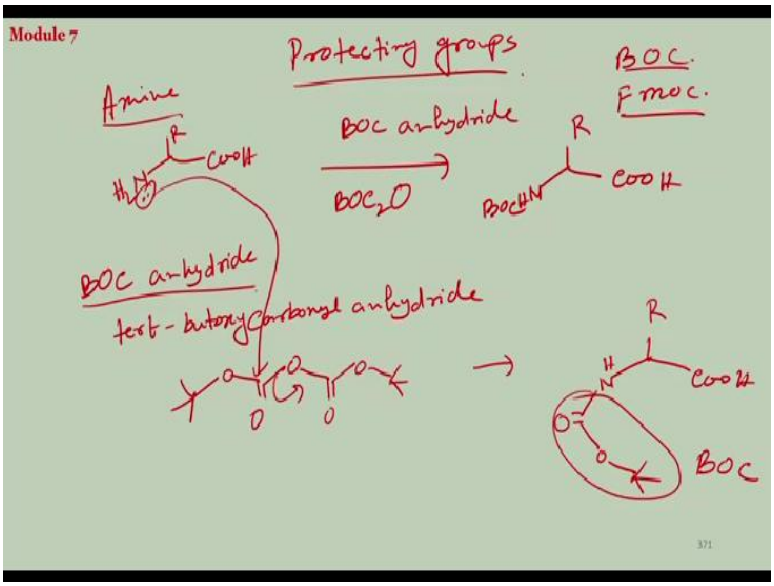
You have R1 acid NH 2, so first step what you can do of course, you have to protect the amine leave these on this is the first and then you add to it with the second amino acid which will have a free amine group and a protected acid as we have done before. Then you get NH R 2 OOP 2. Now what you have to do, if you want to carry on the synthesis, you have to make this acid free. So therefore, deprotection of P 2 selective deprotection of P 2, one way of it.

This can still be protected R 2, this will give you a free carboxylic acid. Now, if you want to carry on, bring in 1 R 3, what you need is a free amine here P 2 protected carboxylic acid. So, you can see after the first one all the subsequent amino acids that you will be needing does not require protection of the amine group. If you follow this kind of protocol, it only requires the protection of the carboxylate.

And it can go on like this NH P 1 R 2 R 3 P 2, then again deprotection of P 2, it will give you another carboxylic free amine can move on like that. So, this is one way of it, it can be done other way around also, where you start with the protected carboxylic acid and do the reaction on the amine part. So, this I think, this the other way around, where we will have protection on the amine not on the acid and that is mostly used, we use it for especially for solid phase peptide synthesis, we use the protection of the amine group, amine side chains that we will see.

So, this is for the protection of the acid group here and before I have talked that we can protect the amine group also. So both protections sometimes would be required.

(Refer Slide Time: 41:52)



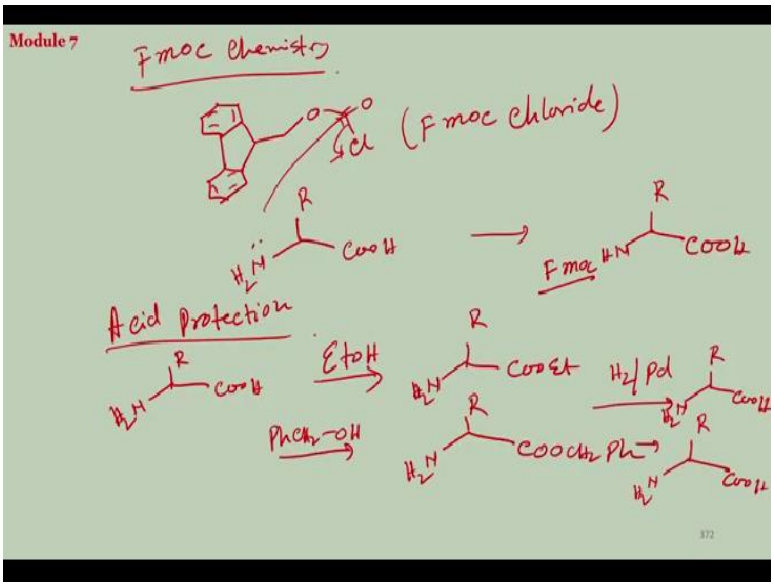
Now, the question is what kind of protecting groups you will be using. So protection of protecting groups. Let us start with the amine, how to selectively protect the amine group here. There are many reagents. The reagents which we use most are 2 of them. One is called BOC. And another is called FMOC. These 2 are the different protecting groups or protecting group chemistry that we use for amine protections of the amino acid.

So first is with the BOC, BOC and hydride is the reagent you can write BOC and BOC anhydride or you can also write BOC 2 O that represents BOC anhydride, then you get R COOH NH BOC, that is a selective protection of the amine group. Now, how does it work, what is the structure of BOC. BOC is basically, this is BOC anhydride, tert-butoxycarbonyl anhydride.

What is the structure, this O COO this tertiary butyl group linked to oxygen here is the basically as SL bond here and then this is anhydride BOC, how does it work. Obviously, now we can see and the chemistry that is supposed to go on, this is electron deficient carbon, this is nucleophilic center obviously carboxylic acid would not react here. So, it will react **to** at this and it will come back, it will come here and go back there. Ultimately it will cleave this bond right.

So you will have NH COO, this we call as BOC and obviously, it decreases the nucleophilicity of the nitrogen. So the electron lone pair will be dispersed here. And that is the reason that this amine is not reactive anymore. It is BOC chemistry is usually been used a lot when you do peptide synthesis in solution phase.

(Refer Slide Time: 45:52)



Another is the FMOC chemistry, FMOC chemistry is mostly used when you do solid phase peptide synthesis, this is FMOC chloride. This is (()) (46:44) that is connected with the intervening CH 2 group with a acetyl chloride. So again the same reaction if you have R acid chloride obviously, chloride would be out NH, the rest of it we call a FMOC. So, FMOC protection group is used to protect the amine group of the amino acid.

FMOC chemistry is a little bit difficult to do in solution phase because the separation or the separation and dry up becomes a little bit difficult. So that is why we prefer BOC chemistry for solution phase, FMOC chemistry for solid phase. So that is about the amine protection. Now comes the acid protection. Usually, to get a good protection of the acids, we will use ester formations because ester bond are very strong bones.

So, we will synthesize ester it can be done with ethanol with ethylester that will make the acid pretty much unreactive or you can use the benzyl ester PhCH 2 OH even the other ones R COO CH 2 Ph, these are the shortened protecting groups of the acids. Now, of course, the deprotection

chemistry I have not talked about after the reaction is over you need to deprotect them all right. So, deprotection for the esters is usually done so little bit harsh condition actually, H 2 palladium.

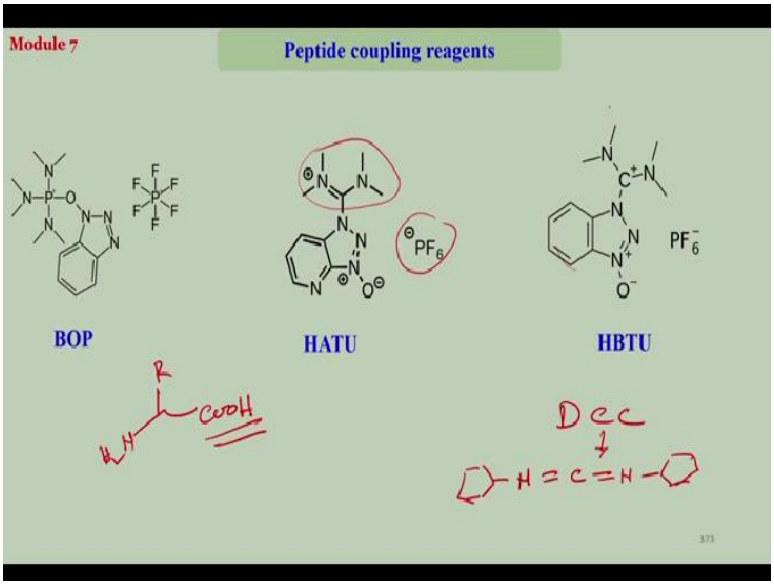
So, that is why these are usually doing here also the same thing, same reagent H2 palladium or H2 platinum can be used to deprotect the acid groups. So, that is actually problematic if you have to do if you want to synthesize the peptides in solution phase every time you are bringing in a reaction for example, here, you have to deprotect the P 2 here. So hydrogen, palladium again to deprotect the acid, then you have to go to the next step again deprotection with hydrogen palladium, that takes time.

It is also a little bit harsh conditions. So, that is on disadvantage of doing the solution phase peptide chemistry. The BOC and FMOC can be deprotected by acids. That I will show you when we do the solid phase peptide synthesis. Now, we have talked about the protections. Now remember, I was always talking about a reagent that whenever you want to synthesize a dipeptide or you want to make a new peptide bond formation, you always need a reagent.

Because the amino acids are not very reactive. So another thing that I otherwise will forget to mention, I also said that the side chains we have seen that pretty much different functional groups are present in the side chain of the amino acid. So, whenever you want to do the peptide chemistry, you also have to protect the side chains accordingly appropriately and then deprotect it also. Sometimes it can be the amine groups has to be protected.

The acid groups have to protected but you have to use different protection groups for those the side chain amino acids side chain amines or the side chain acids compared to these otherwise while deprotecting those will be deprotected at every step also that you do not want okay. So now the coupling reagent the reagent that is used to coupling the acid free carboxylic acid with the free amine are called coupling reagents.

(Refer Slide Time: 52:01)



And there are many, many different types of coupling reagents are present. Without the coupling reagents peptide synthesis does not happen because the both the carboxyl group and amine group in the amino acids are not very deactivate. So you need to activate, these are some of the structures of the coupling reagents. There are many more, I will include one more here DCC. So, DCC is another one.

DCC has the structure carbodiimide right and double bond C double bond N cyclizing. Di cyclizing carbodynamite that is DCC. This is also a very famous coupling reagent. In fact, DCC was one of the first coupling reagent that was used for peptide coupling. So, all these coupling agents are mostly used to activate the carboxylic group. This carboxylate are usually targeted to be activated.

So, once you activate these then the free amine can react. So, you can see the complicated structure. This is BOP, which has first atom here with a positive charge and then fused with a 6 membered and a 5 membered ring and the 5 membered ring is with 3 nitrogen ring. So it is aromatic kind of compound. This is BOP, this is HATU, HATU has again this kind of structure which makes with electron deficient nitrogen here.

Obviously, there are charged species, all these are charged species most of them, this here the charge separation is neutralized, here you have a positive charge. So you need a salt, PF 6 – is a

stable salt. Here also, this is HBTU, HBTU has again the fused structure within NNN triazine. It is a cyclic triazine actually. And then with this kind of this, you will see this is actually our derivative of urea that you can get as a side product.

And then again PF 6 minus, and DCC. So, next lecture, we will see how the mechanism works, how the coupling reagents are involved in activating the carboxylic acid, and how the peptide bond formation actually works in presence of this coupling reagents and then we will also move on to the solid phase peptide synthesis. Thank you.