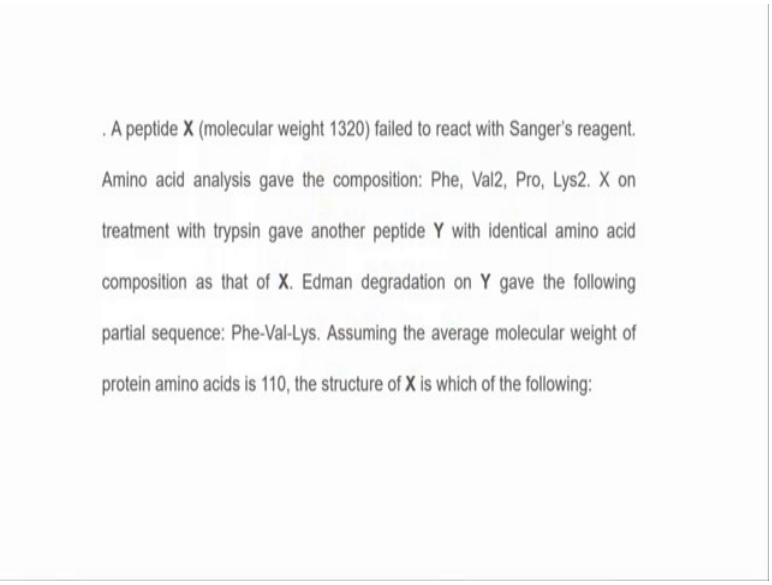
Organic Chemistry in Biology and Drug Development
Prof. Amit Basak
Department of Chemistry
Indian Institute of Technology, Kharagpur

Lecture - 07
Peptide Synthesis: Protecting Groups for Amine and Carboxyl Functionality

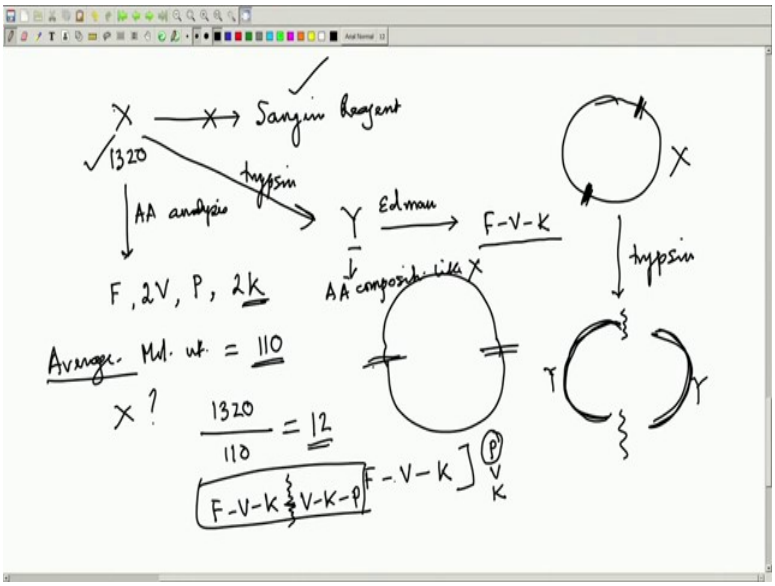
Let us do one problem on the concepts that we have learnt so far. The problem is given here in the slide.

(Refer Slide Time: 00:36)



There is a peptide X of molecular weight 1320; it failed to react with Sanger's reagent that is the first observation. The second is the amino acid analysis, now what is meant by amino acid analysis? That means, if you have a protein or peptide; what are the amino acids that are present in that peptide, sequence will not be investigated, only the amino acids composition which is there in the protein and then the following questions will be addressed: What is the number and what is their ratio? That means, you basically hydrolyze the entire protein and then try to see; what is the concentration; whether there is lysine present or whether there is proline present and what is their molar ratio? So this is called amino acid analysis. So, amino acid analysis will give you the amino acids that are present. So, amino acid analysis gave the following composition of the peptide: 1 molecule of phenylalanine, 2 molecules of valine, there is 1 molecule of proline, 2 molecules of lysine. So, I write it here.

(Refer Slide Time: 01:57)



Some other information is also provided to us. First of all, it says that X, does not react with Sanger's reagent. The second is that its molecular weight is 1320; that will be required and then you already have the reports from amino acid analysis. So, if you have amino acid analysis then you see that there is one phenylalanine (F), (now we can write the one letter code), there are 2 valines (V), there is 1 proline (P) and there are 2 lysines (K).

What are the other observations? If you hydrolyze this molecule X on treatment with trypsin; so tryptic hydrolysis gives another peptide Y. So; that means, if X is hydrolyzed by trypsin; you get another peptide Y and for this Y, one can do the partial sequencing. Partial sequencing means from N-terminus you can determine one after another.

And it says that if you do Edman degradation on this; you get a sequence and the sequence is as follows: First is phenylalanine, second is valine and the third is lysine. So, these are the first three amino acids in the peptide Y; the peptide Y is basically a derivative of peptide X by the action of trypsin.

We have to assume the average molecular weight of an amino acid to 110. Once if you take all the 20 amino acids and do an average, you will see that the number comes close to 110; thus the average molecular weight is 110. So, what is the structure of X? How to solve this? First of all let us find out how many amino acids it has. Now the analysis

revealed that it has got 2 lysines; that means, 2 amino acid here, then 3, then 5, then 6; 6 amino acids are present according to amino acid analysis.

Remember amino acid analysis will only reveal the identity of the amino acids and the ratio. It is not absolute value, which is like empirical formula versus molecular formula. So, these amino acids are in these ratios: 1 : 2 : 1 : 2; that does not mean that this is the absolute value which is present in X. Number of amino acids in X is given by 1320 is the molecular weight.

So, you divide by 110 (average number average molecular weight of amino acids) and that gives you a number of 12; that means, this is a dodecapeptide; that means, 12 amino acids are present in X. Now, it does not react with Sanger's reagent. So, this X has to be a cyclic peptide because there is no N-terminus, if there is an N terminus present in the peptide, Sanger's reagent would have reacted.

We have not written any sequence here, but we have noted that this is just basically the ratio of amino acids. The absolute values actually of X will be 4 Ks, 2 Ps, 4Vs and 2Fs; that will make 12 amino acids.

So, this will be the schematic structure of X and now when you treat with trypsin; the cyclic peptide will be cleaved. This is the empirical formula and the molecular formula becomes double of this; that means, this molecule whatever sequence is there; that is going to duplicate like this and then they are joined together and to give the X.

So, when you treat with trypsin; trypsin recognizes basic amino acids and there is a basic amino acid that is lysine. It is definitely cleaved somewhere. since there is the doubling of the whole thing, so there will be at least 2 cleavages in this cyclic dodecapeptide.

And if there are 2 cleavages in the whole circle; what you get? You get 2 pieces. If there are more number of cleavages? Although there are 4 lysines in X, there cannot be 4 cleavages. Because 4 cleavages cannot give the same sequence of the amino acid composition of Y. said it was mentioned earlier that peptide Y has same amino acid composition like X.

That means, it has also got F and then 2V then proline and then 2K. So, if both the fragments have the same amino acid composition; then the peptide X must be made up of

2 units of same amino acid composition. It must have been cleaved like this to give 2 pieces which have got same amino acid composition; that means, they look the same ok. They have the same amino acid, (Refer Time: 08:27) composition. Now the question is that whether they have same sequence or not?

Y is the only one (exclusive) product. You are not getting two products that means, they have the same amino acid analysis results, as well as they have the same sequence; otherwise you would have got two products here by adding trypsin, but that is not obtained. Now let us start drawing the structure of Y. You have got a partial sequence that is as follows: F is there; then valine, then lysine, up to that point is there.

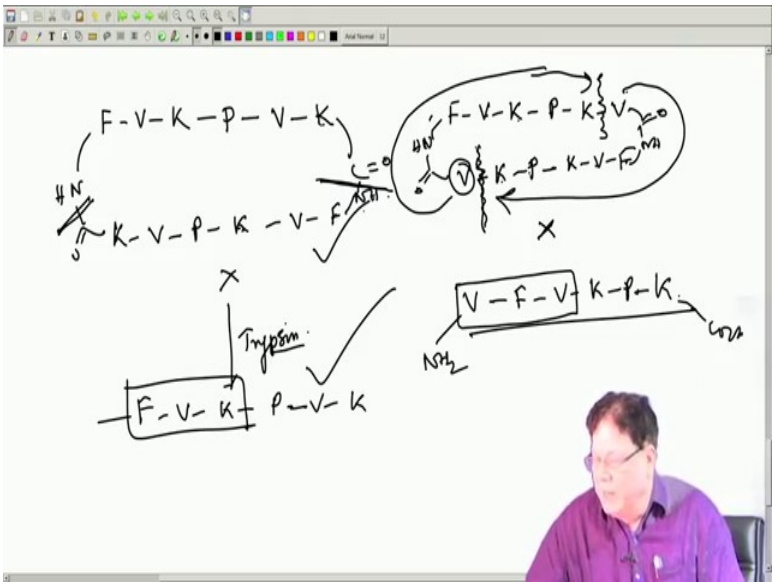
What are left now? You have already taken care of F, you have already taken care of 1 valine and you have already taken care of a lysine. Now what is remaining? You are left with a proline, a valine and a lysine. Now I have different options here; the sequence F V K is fixed because that is the partial sequence.

And then if it is followed by say a valine and then a proline then a lysine, if that may be the case; there are a lot of permutation combinations now with P V K. But if somebody is very intelligent, he can immediately say that there must be a proline because there are 2 lysines here in Y, so the 2 lysines are expected to be cleaved by trypsin. But that is not happening because Y is the only cleaved piece obtained.

So, if the Y is cleaved again by the lysine; since there are 2 lysines present in Y, still you are getting only 1 cleavage on two sides. So, if that be the case, then there must be a proline which is after the lysine because only if that happens; then you have got only another lysine which can induce the cleavage only once. In other way, if you want from the very beginning; ab initio, then there is a lot of wastage of time because you can start writing many sequences with that F V K; suppose after this I put V K P. So, this is Y; that is one possibility.

Now whether this will be correct or not that is a question; in this case what will happen is that there will be a cleavage here because this is the lysine; however, there will not be any cleavage at this position. Then you can have another sequence F-V-K, then K-V-P all these things are there, but you can continue with this but then it will be a very time consuming route.

(Refer Slide Time: 11:54)



But apply your brain that there is F, V and K; that was there and then you try to put in proline to stop the breakage within Y. If you do not put proline here and if you put a valine here; that means, there should have been cleavage here at that point, but trypsin is not doing that. Trypsin is giving this entire 6 amino acids containing fragment; so that is the trick here. Now, you are left with only two amino acids, one is valine and the other is a lysine; now we have two possibilities: Valine lysine that is one possibility, the other possibility is F V K P and then K and V. Remember, this is your N-terminus and this is your C terminus. Now, what we have to do? To dimerize it; that it will be just the opposite way you write that this will be K, this will be V, that will be P, that will be K, that will be V that will be F and then this will be CO and that will be your NH; so, this is one possible structure.

The other possible structure is coming from this dimerization or cyclic structure of this. So, you write the opposites one that might be better; might be easier to understand. First write from the opposite side V K P and then K V F and then this is your N H, this is your CO and on this side this is your CO and that is your NH. So, this N could have been the N terminus, however there is no terminus in that sense because it is a cyclic one. So, it is going like this; we started with F thinking that the NH_2 is on this side and the carboxyl on the other side.

So, either this is X or either that is X; now which one is correct? If you give trypsin then where are the cleavages? This is a lysine; remember trypsin recognizes basic amino acids and it cleaves on the carboxyl end of the basic amino acid; these are the two important points.

There is a basic amino acid, but after this there is a proline; I told you if there is a proline, hydrolysis will not take place. Then there is a valine and then followed by a lysine; now this is the one, after this there is a phenylalanine. So, here there is no problem since there is no proline after this lysine; so there will be a cleavage at this point; the carboxyl end of the lysine.

And again you come back to this lysine, that cannot be cleaved because there is a proline here, but this lysine is connected to phenylalanine. So, now, there will be a cleavage here, and now you see that this actually matches. The sequence here is F V K P V K and the sequence here is the same F V K P V K; remember you have to read from this side because this is the N terminus end of the lower chain. On the other hand, suppose this is correct, then what will happen? This will not be hydrolyzed because there is a proline, but this will be hydrolyzed.

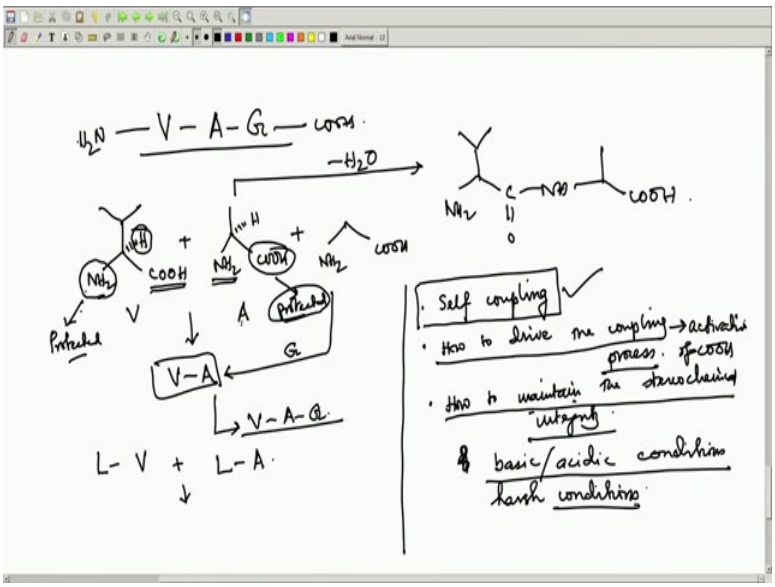
And where is the other hydrolysis point? The other hydrolysis point is this; again you get this the same two parts here: V F V K P K and on this side you get again V F V K P K; they seem the same, but where is the difference? Because our results say that the first three amino acids are F V K. But in this case, the amino acids that will come out will be V, then F, then V, then K, then P then K. I hope you understand that this is NH_2 and this is CO_2H because now the breakage has taken place here.

So, this is the first amino acid from this side and that goes up to this side and this is the other way- V F V K here. So, only difference is that in this case, the partial sequence will be this V F V, but in this case, the partial sequence will be F V K P V K; so the partial sequence matches, so this is the correct answer. So this is a very tricky problem a cyclic one which is divided into equal parts, having the same amino acid composition and then with trypsin. Remember trypsin does not discriminate between cyclic or acyclic; that also we have learned from here, but Sanger's or the carboxyl peptidase will distinguish between the C terminus, whether there is a free C terminus or whether there is any free N terminus.

So we just learned how to know the primary structure and we have also sorted out one problem and that is a more difficult problem; there are even simpler problems like the other ones. And in the weekly assignments, we will give different types of problems based on the primary structure of proteins; for the first week,

Next if you have a protein, the first thing that you want to know is to know the amino acid sequence. Now, you just reverse the problem that I want a protein or I want a peptide which has a particular sequence. Suppose, I want to make a peptide which has got V and then A and then G; this is a tripeptide valinyl alanyl glycine.

(Refer Slide Time: 18:09)



So, it is now the reverse problem; earlier I had a peptide I wanted to know what are the amino acids sequences, it is the now the reverse problem, that I want to make a peptide with a defined sequence. I take a very simple a tripeptide, I do not want to make it really big so I chose a very small tripeptide; because the historical reports on how peptide synthesis took place reveals that it was the dipeptide that was synthesized first.

Now, I will first highlight what are the problems associated with this. What is valine? Valine is having a side chain which is isopropyl; this is valine. And what is alanine? Alanine has only a methyl side chain and what is glycine? Glycine does not have any

side chain. So, when I have written this V-A-G; that means, NH_2 of this V is free and the CO_2H of G is free.

So, basically we have to combine these three; now you cannot combine all three in one go; that is not possible; so you have to go step by step. There are two possibilities now; what can I do? I can combine V and A so that I make a dipeptide V-A and then I add the G and to make the complete the synthesis as V-A-G. That could be my strategy; or the alternative strategy could be combine A and G; so I will make A-G and then I will add the V and to make V-A-G; that is also possible; from both directions you can synthesize the molecule.

Suppose we decided that first we will make V-A. So, we have to join V and A, valine and alanine and these joining is nothing, but expulsion of a water molecule and that should give a peptide bond CONH and then the alanine with free COOH. I told you that the peptide bond formation is thermodynamically very favorable; its ΔG is highly negative. However, there is an enormous kinetic barrier to do this reaction; that is one problem.

You cannot just mix these two and you cannot just remove the water from this; that is number one the problem. The second problem is, what is the guarantee that this NH_2 will react with this carboxyl because this NH_2 can react with another molecule of alanine. So, that alanine and this alanine can combine to form a peptide alaninyl alanine. valine can also undergo similar self-coupling. So that is the second problem.

And there is the third problem; the third problem is that these molecules are chiral. When you make a peptide, we are talking about making from L amino acids. If you require the amino acid for some purpose, you can also start with D-isomer, but the question is that when you are joining these two, whether your chiral integrity is retained or not; that means, whether there is any percentage of racemization or not.

Because, if there is any racemization, then your number of peptides will increase L configuration of V plus L configuration of alanine. And if these two are joined if there is any racemization L going to D then you have a new peptide; L configured V with a D configured A or the vice versa, D configured V, and the L configured A or D configured V and D configured A.

So there are basically three problems; one is self coupling, how to avoid that? Number 2 is, how to drive the coupling reaction because it is; although thermodynamically favorable, it is kinetically unfavorable, to drive the coupling process. By coupling process, we are meaning the formation of the peptide bond and the third is how to maintain the maintenance of stereochemical integrity...

So, these are the three challenges; now to avoid self coupling, what you have to do? Since you want only the reaction to happen between this carboxyl of V and this NH_2 of A; and you do not want any reaction between this carboxyl of A with this NH_2 of A; or this carboxyl of A with this NH_2 of V; because that will give A V; not V A. So, if that be the case then what happens? So, what you have to do now? You have to do what is called the protection chemistry see you have to protect this amine so that it cannot participate when you are coupling this with this amine. So, this amine of V needs to be protected.

And at the same time this carboxyl of A needs to be protected; then once you have these protected things, (N-protected valine) and the C-protected alanine); then if you do the reaction under the appropriate condition, then only you can have the combination of the desired peptide bond. Thus to avoid this self coupling you have to protect these amino acids.

Another good part of a protection is that these amino acids are actually water soluble, and they are insoluble in organic solvents. When you do synthesis, it is very difficult to do synthesis in water; today everybody is searching for methods, which can be done in water; that is what is called green chemistry that is in the domain of green chemistry; remember enzymes work in water.

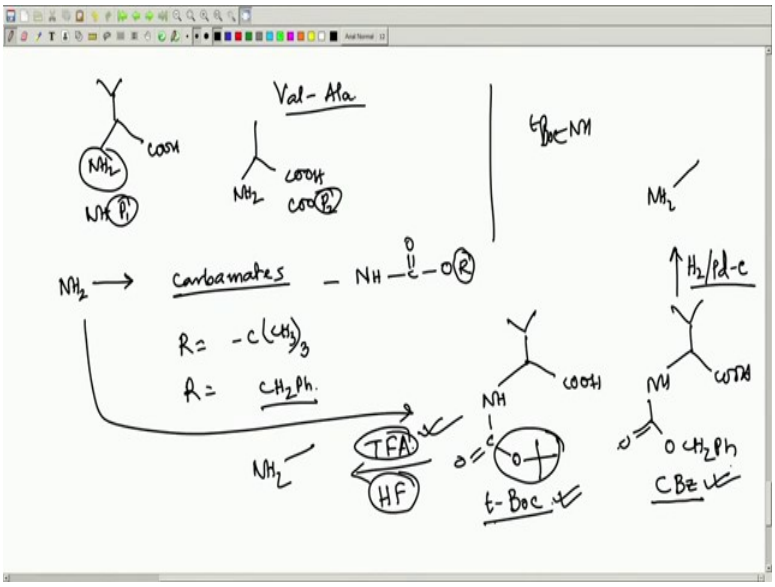
That is why people are trying to mimic enzymes or use enzymes to do synthetic chemistry. But when you do peptide chemistry, this coupling which involves the removal of water, you cannot do the reaction in water. If something is removed which is already present, you know by simple (Refer Time: 25:46) Le Chatelier's principle that will not work. So, you cannot do it in water; by protection these compounds become soluble in organic solvent. So, your chemistry can be done in organic solvent; so that is another advantage.

The second one, how to drive the coupling process? So what will you do when you want to couple NH_2 and CO_2H? Now, in the laboratory what we do? In specially the organic

chemistry laboratory, if you are asked to r make an amide bond from a carboxylic acid; you know the way to do it is that you convert the carboxylic acid into an acid chloride by adding PCl_5 or thionyl chloride or oxalyl chloride. So, that will become COCl; that means, you activate the carboxylic acid and since Cl is a very good leaving group..

So, now the NH_2 can attack this carbon of the COCl and the chloride will leave. So, basically what you need to do drive this coupling process is activation of CO_2 H group. However, in amino acid chemistry you cannot use PCl_5 or $SOCl_2$; I will tell you why. Third problem is how to maintain the stereochemical integrity; you know the stereochemistry comes from this hydrogen. Epimerization of this hydrogen or a racemization of this hydrogen involves removal of the hydrogen, because this is quite acidic and then again putting back the hydrogen into the same carbon and while putting back, you may actually invert the configuration.

So, one should not use highly basic or highly acidic conditions; mild acid or base may be used. But one should avoid the bases like sodium hydroxide or potassium hydroxide or acids like HCL; one should avoid those type of reagents in the entire process. That means, one should avoid harsh conditions because racemization can take place under harsh conditions; even if there is almost neutral pH, but if you heat it strongly during the coupling process or any of these chemistry that we are going to describe, then also there could be racemization. So, these are the three major problems during peptide coupling.



So, first what it says that I told you that our first task is to make the dipeptide between valine and alanine; accordingly I need to protect the amine of valine, keeping the carboxyl end free. Then I protect the alanine carboxylic acid and keep the amine free. So, that should be COO tethered to some protection group here and that should be NH bonded to some protection group these are different; so is P1 and that is P2. Now what are these protecting groups? Throughout the process, harsh conditions have to be avoided in order to minimize or completely remove the racemization; so, you have to use very mild processes.

Now there are many protecting groups, but I will just use as I show you the very common ones. NH_2 protection is done as these are usually converted into called carbamates; what are carbamates? Carbamates are NH, then CO, then O and then R; this comes from carbamic acid; you know what is carbamic acid? It is NH_2CO_2H; which is very unstable and hence it immediately loses ammonia and carbon dioxide; but carbamates are stable.

Now, so this amine has to be converted into a carbamate and this R should be such that this can be taken off very easily. So, based on this you have R which is equal to ᵗbutyl (tertiary butyl group); R can be CH_2P h; that means, the amine basically is converted into NH; suppose it is CO then O, then t butyl is written like this and this is your valine; so the carboxylic acid remains there. So, this is protected and this is called the ᵗbutyloxycarbonyl. So, the amine has been protected as ᵗbutyloxy carbonyl.

So, this is commonly known as ᵗBoc group (tertiary butyloxycarbonyl). Since all the time you cannot write the full structure, so you write in the abbreviated form. Usually t is in the superscript. So, similar to this protection, you can also use the benzyl instead of the tertiary butyl, you can also have benzyl and then that will be called NHCOO and then CH_2Ph.

Earlier when it was discovered that this it is a good protecting group for amine functionality, it was named benzyloxycarbonyl. However it is better known as carbobenzyl group; that is why this is called Cbz. You could have called it benzyloxycarbonyl, but they do not call it benzyloxycarbonyl because that will match with benzyloxycarbonyl; so very difficult to differentiate between these two.

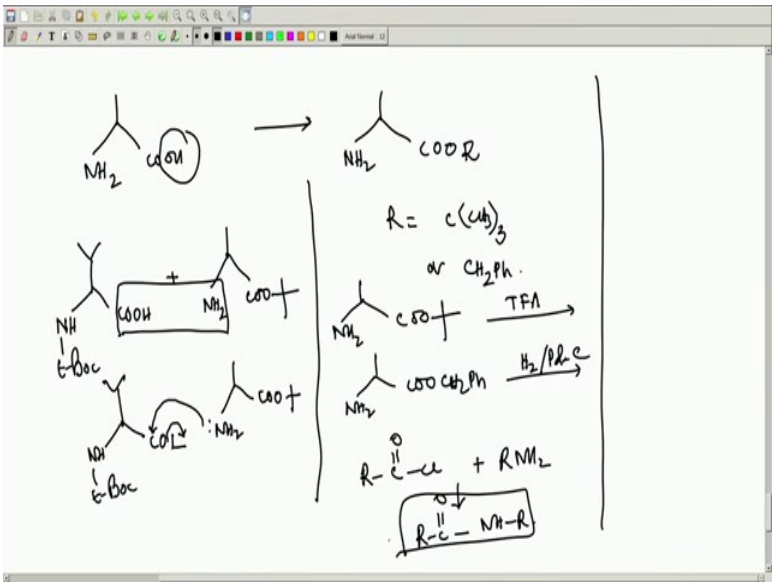
So, that is why this is called carbobenzyl Cbz and this is called ^{t}Boc. One thing that one must remember is that in peptide synthesis, the first step is the protection of amine of one amino acid and carboxyl of the other amino acid; then you couple the two. But what is the third step? In the third step, you finally take these protecting groups off because you do not want the protected peptides, so you have to deprotect this.

Under deprotection, The N-protected end deprotects into the free amine again by organic acid like TFA; TFA is an organic acid. I said that strong acid and strong base should be avoided, but by strong acid base, I am usually talking about NaOH or HCl; those types of things are to be avoided, organic acids are no problem. And also anhydrous HF which is actually a very weak acid in the pure form... You can also use HF; that is allowed as long as water is not there. So, these are the t-butyl group comes off by treating with TFA usually TFA is easier because for HF, you need special apparatus; HF etches the glass.

So, you cannot do HF reaction in glass vessels; you have to do it in teflon coated glass vessels; and that is costly. So, usually in the laboratory normally we use TFA; as TFA is much easy to use. Thus this t-butyl comes off with TFA and the Cbz (this whole group) comes off, if you treat that with hydrogen using palladium charcoal; then this Cbz group comes off; that means, one comes off under very neutral condition, another comes off with TFA.

So, in a peptide if you have both these groups present; say a tripeptide where Cbz and ^{t}Boc are present together; then you can selectively deprotect one over the other. If we use hydrogen, cbz will be deprotected; so it will form the free amine and if you add TFA; then this ^{t}Boc-NH will form free amine, these are very important; that means, these two processes are orthogonal to each other. So, you can selectively deprotect one keeping the other intact ok. So, these are the two protecting groups for the amine.

(Refer Slide Time: 35:21)



So, next what are the protecting grouse for the carboxyl? In our case I have to protect the carboxylic group of the alanine. I have to use the alanine where the carboxyl is protected. So, what are the protecting groups that are used for carboxyl?

As a protecting group here, COOH is converted into an ester, but not any ester; R cannot be methyl or ethyl and those things. Because you know methyl or ethyl esters have to be hydrolyzed to get the free acid using alkali or concentrated acid.

So, this R cannot be methyl or ethyl; then what is R? R is again either t-butyl (CMe_3) or it could be again benzyl (CH_2Ph). That means, the carboxyl is protected either as t-butyl or as the benzyl ester. Remember there are lots of other protecting groups that are used, but I am saying you the most commonly used protecting groups. So, either t-butyl or it could be benzyl. Now, again they have to be deprotected; this is the protected form; if you want to deprotect at the end of the day. So, what are you going to do? TFA will take care of this t-butyl ester and hydrogen will take care of the benzyl ester. You will see the utility of this orthogonality of protecting groups, when you have peptides of higher chain length; that means, higher number of amino acids, and then you have to constantly switch back and forth of different protecting groups ok.

So, I am not saying how these protecting groups are incorporated; that is possibly the content for my next lecture I will quickly tell at that time that how the protecting groups are incorporated because that is the first step; in the last step after all the coupling is

done, you have to deprotect the terminals. So, those mechanisms we can discuss in the next class, but right now; now we have a protecting group, suppose we have put a protecting group here NH ^{t}Boc ok. So, our valine is protected as a ^{t}Boc and the alanine should be protected as the carboxyl should be protected; the question is which protecting group I should use?

Now, in this scheme I see the valine has been protected as ^{t}Boc which comes off with TFA. Now, if I protect the alanine with benzyl; I can do that previously and then do this coupling reaction, but then I will get a protected peptide where to take the benzyl off, I have to use a hydrogen.

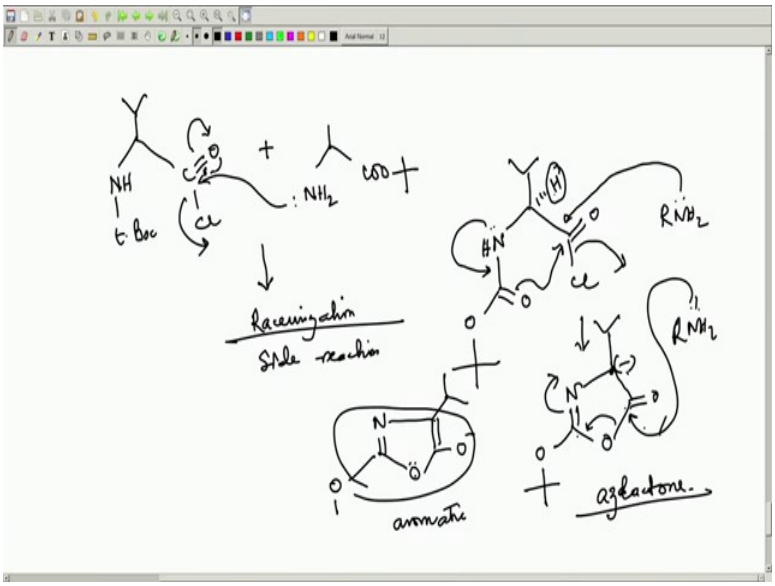
So, that means, I am adding more number of steps into my synthesis which is not desirable. So, what we should do then? We should not use a CH_2Ph here; it will be better that if we use t-butyl here; then what will happen; that after the coupling reaction here; I can take both the protecting groups off at the last step; so with the single step; I can deprotect the molecule.

So, during the peptide synthesis, the protecting groups are selected in such a way that last step becomes a single step deprotection of all the protecting groups. So, at that time, you should have protecting groups which are not orthogonal; which are of similar type, which comes off under the same condition. So, the next part is the elimination of water leading to the formation of the peptide bond.

So, what to do? I said that what you need is basically to activate this carboxyl. So, this carboxyl must be converted into something which will be called a COL; L is a leaving group and this NH_2 remains as NH_2, that is the nucleophile and now this should be the reaction where the leaving group leaves.

So, you get CONH bond. Organic chemists make an amide by reacting RCOCl and RNH_2 and the reaction is very spontaneous, very vigorous. Those who have done this amide preparation, they know that immediately it forms the amide bond ok. So, initially it was tried to couple this amino acids via the above mentioned strategy where they convert this carboxylic acid into your RCOCl.

(Refer Slide Time: 41:06)



Let us consider t-Boc protected valine with COCl group, and you have this amino (NH_2) group free alanine with its C-terminal protected as t butyl. You could have the other possibilities of using Cbz protecting group here. So that last step of deprotection in case of Cbz is hydrogenolysis. Now, this if you want to couple this; this is going to attack here and that is going to come out. Initially, the negative charge goes on the oxygen; then it comes back and the chloride is kicked out.

So, you get the peptide, but the problem is here lot of racemization happens because of side reactions. What are those side reactions? That when you have this valine; NH_2 being protected a NH'Boc; you have COCl and what you want is an external nucleophile; that is the amino group from the second amino acid is going to attack here and furnish what you are expecting.But alongside, something else will also happen and that is before the external nucleophile attacks; just like your methyl in cyanogen bromide reaction; this now, comes to this bond; the oxygen is activated and that is going to attack the carbonyl and the chloride leaves. So, what is formed is N double bond C bonded to O and then it is a lactone and this is your valine and this is the product O-t-butyl. Now this is called an azlactone.The azlactone is also not very stable hence it can again undergo further reaction.

So, that will come here and it will again go back to the t-butyl, that is possible; but this coupling reaction has got two mechanisms. One is it can directly attack the acid chlorides and form the amide bond; if that happens no problem, no racemezation.

But if it in the meantime; if it goes to the azlactone and then the amine from the other amino acid comes an act and attacks the azlactone and then form the amide bond; then there is a problem. The problem is in the azlactone, this hydrogen which was there earlier in valine, that is where the stereochemistry comes now. This hydrogen becomes very susceptible as soon; as it forms the azlactone, it is much more acidic than this hydrogen when it is in the cyclic form.

And why is that? Because if you remove this hydrogen; and you make a negative charge in the conjugate base, you see this negative charge will delocalize with the carbonyl. So, you have a double bond here that will be O minus and what is the situation then? Then the whole thing becomes N double bond O and then O minus and then there is a double bond and then you have the isopropyl part of valine.

Now, this is an aromatic system; oxygen lone pair and these two double bonds that make its aromatic. So; that means, this hydrogen becomes extremely acidic and by the time this RNH_2 attacks this carbonyl, already there will be some racemization that will take place before the attack of the external nucleophile.

So, azlactone formation is a serious drawback of activating the acid group as the acid chloride. So, we should find a very milder version of coupling reaction and that we will talk in our next lecture. So, what we have learnt is basically that when you want to make a peptide of desired amino acids, then the first step is protection, the second step is coupling and the third step is deprotection and then you get the designed peptide.

So, I have just highlighted some of these protecting groups, some coupling reagent what was initially tried; I told you about this azlactone formation. I have told you the methods to deprotect; we have also told about some of these protecting groups ok, thank you. So, we will go deeper in the next lecture about all these coupling reagent and the coupling issues.

Thank you.