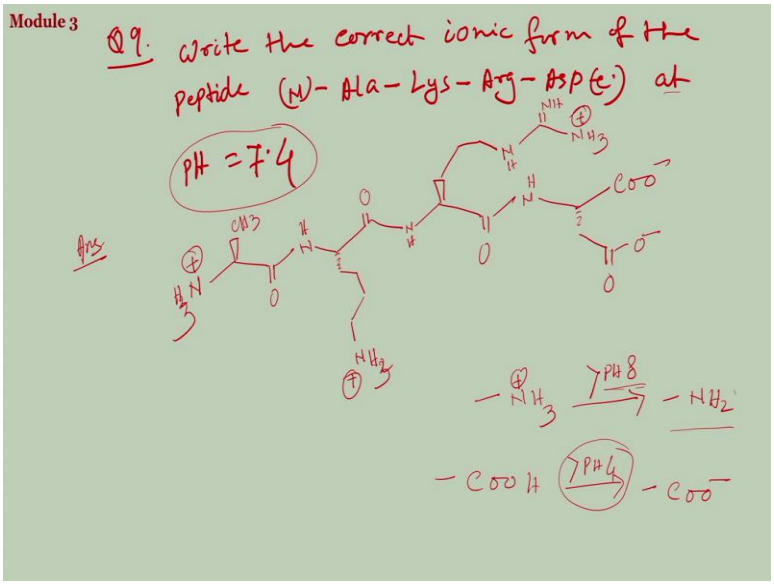
Essentials of Biomolecules:
Nuclic Acids, Peptides and Carbohydrates
Prof. Dr. Lal Mohan Kundu
Department of Chemistry
Indian Institute of Technology-Guwahati

Lecture No. 16
Sugar Chemistry

Hello, everybody and welcome back. So, since the last lecture, we are discussing about some problems, new medical problems we have talked about some problems and the solutions of the modules what we have so far covered. That means from starting with the model 1 up to module 3. So, these problems and solutions are in combination of both numerical problems as well as today will come to the synthetic problems will talk about synthesis of some components.

(Refer Slide Time: 01:11)



So, last class last lecture, I had left you with this problem that is correct ionic form of the peptide and this is the sequence starting with N terminal alanine, lysine, arginine and aspartic acid that is C terminal, what would be the ionic form of this peptide at pH 7.4. This is important of which pH you are you have to write the structure. So, in this case its pH 7.4. Now, if you have to write the structure it start so this is alanine in N terminal, which means it has the amine group.

So, I am first writing the skeleton and then we will look at that what are the ions or which ionic form they might alanine is CH3 here. And then you have CO NH and then next is lysine is 4 CH2, 1234 and then amine after that so this would be CO NH sorry this would be the down the plain 123, 1234 they should be off the plain archeline has 123 carbon and then you have NH now one is NH and here it is NH2 I guess that is the structure of origin and you can verify.

CO NH and this aspartic acid that is aspartic acid is CH2 OH and here this is the carboxyl terminal. So this is the raw skeleton without going to any charge formation or without going to any ionic characters. Now let us see what happens at pH 7.4. So, if you remember the PK values or if you look at the table of the PK values that that was given for all amino acids we will see a general trend. The trend is for amine group pH the NH2 of the amino acids itself or be the mean of the side chains, that trainees, all of them have the peak A value which is more than 8.

At least more than 8, most of them are above 9, but at least they are more than 8. Which means that above pK of or above pH of 8, they will be deprotonated. Below that pH or below the pH 8 it will stay as protonated form so, this form will stay as NH2 and NH3+ will be pure NH2 only at least above pH of 8. Most of the times it is above 9. If you look at the table you will see but in general at least it is more than 8.

So, if you want to get the deprotonated version that your pH of the solution has to be more than 8. So, this is will come to the pure NH2 or below that pH. This will always exist as ionic form NH 3+ protonated form. So, there you go if you do it at pH 7.4, which is the physiological pH so, this will stay not as NH2 this will be as NH3+ and similarly here NH3+ so is here NH3+ they will all exist as the protonated form or the positive form.

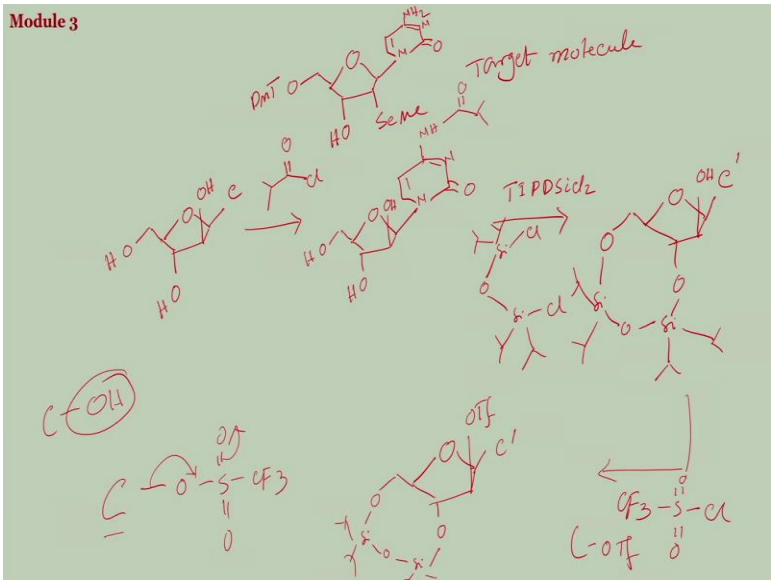
Now what about the other side chains? Alanine does not have acidic or basic side chain so no problem here. This is lysine, this is arginine. And this is aspartic acid has a carboxylic acid as the side chain. Now the other trend is for the carboxylic acids they have the pK value off, of course below 7, and much below 7. At most, I think pK value of the carboxylic acids in the amino acids were close to at least 4. So below that is the most that is the highest pK value actually of the carboxylic acid of the amino acids I guess.

So which means above, above that pH, pH4 for they will exist as the deprotonated form. So this above the pH4 will exist as C double minus below that pH, it will be protonated not for depends on the amino acids which amino acids were talking about you will see that many of them have the pK of the carboxylic acids, which are close to 1.8 to something like that so in those cases, so above 2 or above 3 itself, they will be deprotonated.

But the general that at least above pH everything would be deprotonated. So and we are doing it at pH 7.4. So, here in that pH, your acid would stay not as acid, but an ionic form. So this would be the actual ionic structure of the peptide that all the image will be in protonated form. All carboxylate would be in the carboxylic acids would be the carboxylate amine form of course, now, you can see that if you change the pH, this will vary.

If you reduce the pH2 maybe 2, then this will all change if you increase the pH upto 10 or 11 maybe they will, they will also you will also see different kinds of structures possible. Now, you have to deduce how to practice what will be the forms and different pHs so, likewise, so far I have shown you a few problems in numericals and as well as based on some structures, so, you could frame your own problems, which is closer to this area and then can try to solve this so, I am just giving you a few examples of which can have many very essence.

(Refer Slide Time: 10:56)



So today we will talk about some synthetic problems and we will start with a target molecule NH2 this is cytosine and here selenium methyl OH here ODMT. This is your target molecule question is how can you synthesis this molecule the change is for the ribose sugar you have OH here in this case, we have our seme, selenium methyl, selenium group basically, how can you synthesize this compound, you can start with your favorite molecule or your molecule of choice.

You can choose your own molecule where you want to start from. So, this molecule or synthesized the reason is such there are plenty of various sounds like there, I mean, there are many types of variations that have been synthesized and they show different kinds of properties actually. So, this particular molecule or this particular kind was synthesized to improve crystallization property of DNA.

If you synthesize a DNA out of this, what was seen that presence of selenium that actually increases the chance of getting a good crystals. Usually, the crystallization of DNA is very hard actually. And most of the times it fails, it is quite hard to get a proper crystals of DNA, if you want to study the structural pattern of a DNA, so, introduction of selenium was seen that it will increase the tendency to form the crystals.

So, there you go, this is the target molecule. Obviously, only one thing you are doing is replacing the hydroxyl group with a selino methyl and both of these are down the plain. So what they have done is they have started with this molecule writing cytosine. And since you need the selino methyl in the down the plain, they have started with the OH which was off the plain because then it makes it easier. And then the as usual I think that that is only the clever part probably.

Because the rest of these the others you can easily do so, first thing you have to do as we have talked that if you have an equally ways so and in order to make sure that they do not participate in any reactions that you intend to do on the sugars, the new protect the functional groups are the reactive functional groups of the basis this case you have a free I mean in the cytosine and therefore, you definitely need a suitable production of the amine group.

So, you can do it with the acetyl production or they do it with this kind of productive group, if you use this of course, you get the elimination nucleophilic substitution of the chloride and what you have is here in NH CO isopropyl amide link is basically which is less reactive and it will be stable I am sorry, I made a mistake in the target. This would be this, the double bond here. So this is the double bond. This is the target the double bond that is the first step.

Second step is that you need to give this OH as it is, and you need to protect the other 2, so that they do not interfere in your future nucleophilic substitution reaction. So, and again you can see this is a 1 3 dial, so, you can protect it by the silyl group that is the TIP Dsicl2 and of course, you need that suitable solvent and all and all this if you do that, so, the structure of this was I think I had mentioned once CL see this is TIP dCL actually, so this is the structure of it.

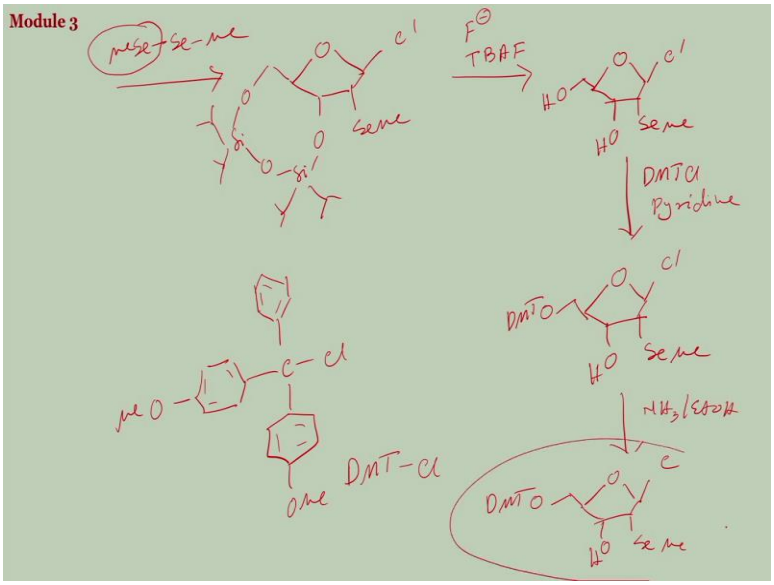
So if you do that, then what you get is so I am writing this as C prime, because this will be all along their production, this is not going to react was mostly once will react si o si and this would be the production group once you are protected the 1 and 3. Now you can do reactions on this center and in nucleophilic substitution you do will happen here because the cytosine will not react this is not going react only position is left is the C2 prime you have the OH group and that you want to replace.

And we know that ethyl bond or the alkaline bonds are CO bonds basically in this bond CO bond is quite strong, it does not get eliminated easily, if you remember OH or the OR ether bond or the alkaline bonds are very hard to break because the OH group or the OR groups, they are not good living groups, they have very bad living groups. That is why you need to use it usually very strong acid to if you want to make it protonated and then only it can be eliminated.

So, either you have to use a very strong acid which will make problems here or you have to find other ways to make this wage a better living room. And that is done here actually that this OH made into a better living group by OTF. So, if you use CF3 so, s to CL this is in short this is called O Triflate, OTF you have electron withdrawing capacity. Here also they are all electron withdrawing in nature. So, and it will do the production first here and wants to do that.

So, OTF it is a very good living group and very widely used in organic chemistry. Let me here the other 2 are there So, this becomes quite a good living group because you see if you write O Triflate is this so if this is eliminated, this comes this goes there and also tracks the electron density. So, this overall this will pull the electron density so bonded electron with the carbon for example, would be towards this and it can leave easily with it nucleophilic proper nucleophilic substitution even if you do not have a very strong nucleophile that rate.

(Refer Slide Time: 21:40)



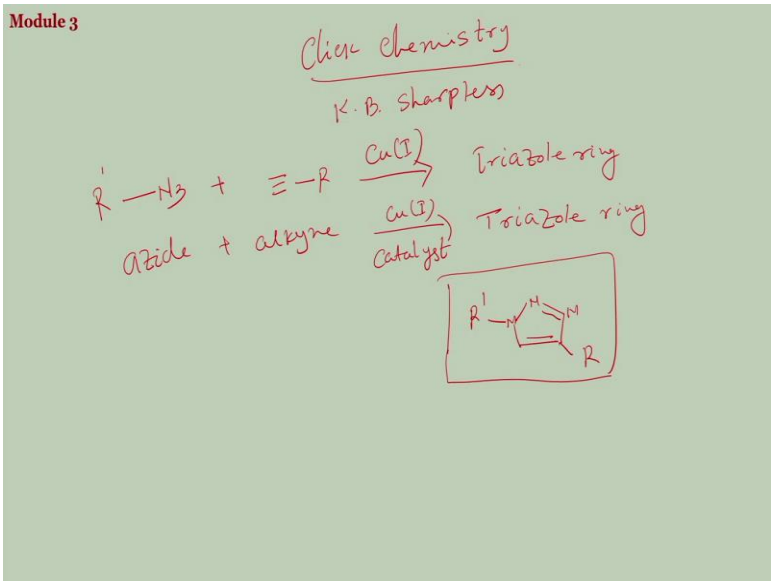
So now mese se me can be for this effectively, mese is the nucleophile. So this is the nucleophile there to latter the center eliminating the triflate, and of course, since the triflate is on is down the plain since this is a, I am sorry, since the OTF is up the plain, the nucleophile has to attack from the down the plain it is a pure SN2 reaction. So therefore, you are seme comes down the plain as your target is now all you have to do is to cleave this production which you can do with the fluoride and source as have talked usually use TBAF.

You can find the structure of TBAF that is the fluoride components ammonium fluoride basically tetra butyl ammonium fluoride, F- this is ionic form, which has fluoride minus so little bit fluoride that will de protect this and will bring back C prime OH, bring back OH then you do the DMT production DMT chloride you need period in a period in is a solvent that is usually used for DMT production, because the DMT production is very sensitive chemistry.

Presence of slight amount of acid or slight amount of water will make this will kill this reaction. So, dry pyridine is usually used seme which and you have O DMT, DMT chloride because DMT has a larger size that only reacts at the 5 prime position. I think I have shown the structure of DMT before. So, the DMT is O Methoxy 2 benzene rings will have a para O methoxy groups and 1 benzene is acetone benzene. So, this is your DMT chloride, dimethoxytrityl, CL is chloride.

This is because of this large size it goes actually here only and then if you want to protect this, if you want to go for the DNA synthesis of course, he will keep this on as it is but if you want to do this, you can simply use the ammonia ethanol or other suitable DE protecting agents that will bring back your C, O DMT. So, here is your target compound.

(Refer Slide Time: 25:48)



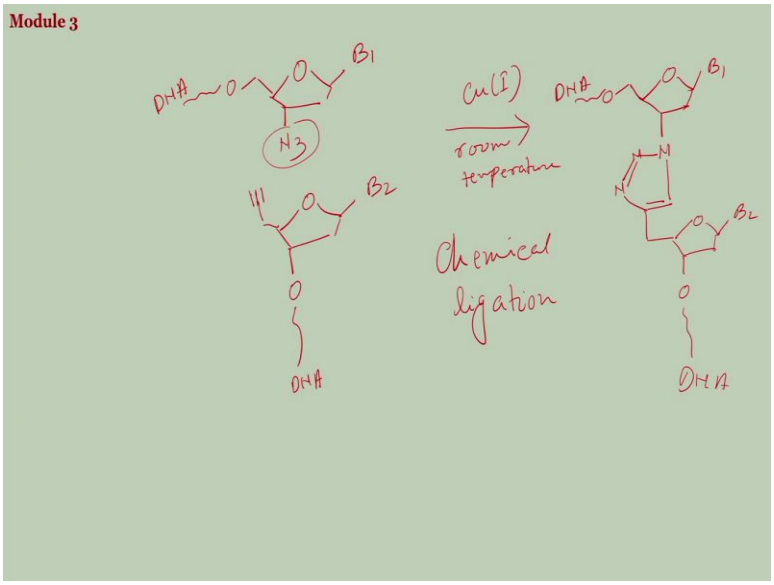
Now, I like to give you a brief introduction to a chemistry which is known as I think maybe many of you are aware of it already, which is known as click chemistry it become very, popular because of its applications in chemistry and Chemical Biology. So, and for that this was developed by Barry Sharpless, K. B. Sharpless, for which he was he has received the Nobel Prize, because it is a worst array of applications in, especially in biology, as well as in chemistry of course.

It is basically a coupling reaction metal coupling reaction that you have seen in many, name reactions. They have they are involved metal, metal coupling reactions are involved in those so, here this also is kind of a coupling reactions between and as I said, if you have a N3 group as I plus if you have an alkyne then in presence of copper 1 catalyst, they will be coupled together and they form a 5 membered ring.

So, 5 membered triazole ring and that chemistry is known as the click chemistry. So, it is a reaction between azide plus alkyne in copper one, using copper one catalyst gives you triazole ring, which is in this case, if you do, let us, if you call R N3, and if you take this as R prime, this is R prime and this is R then product to be here is R double bond. This carbon here is 1 N that is R prime. Here is another end double bond, this is the compound triazole compound.

So, these R can R prime can be anything in this direction this R can be anything in the other direction and you can like get them together you can join them together by this triazole ring formations, very easy chemistry. And this can all happen at room temperature, which makes it really a good way for its applications in biology because you cannot use higher temperature there.

(Refer Slide Time: 29:44)



So in this case, we will see an example of doing click chemistry for DNA. So if you have in general B1 base, if you do N3 for example, here and you have a DNA where the last base is

modified with N3 and you have another DNA B2. Next this is protect this is another DNA. The other part of it said 1 kind of DNA. This is other DNA, DNA 1 and DNA 2 and here if you have 5 prime 121, then triple bond there.

Copper 1 room temperature then what will be the product? Of course product is simple as you have seen just that these as I will couple with this alkyne gave you the 5 memory. So, you will basically connect in this way you will basically connect the 1 DNA with other DNA. This is your 1 DNA and what you will have is in another in here this should be double bond here CH2 something like this the angle is not proper, but you can make it better.

This is what you will get instead of usual now, if you look where the difference between the normal DNA with this molecule is that this is also a DNA only difference is here, the backbone skeleton, normal DNA has a phosphate bond, phosphate group here you have a try so mighty in this case without a phosphate. This kind of chemistry in biological terms is known as chemical ligation and that is why this is very much for because you can like you have seen how to like it to DNA strands using the enzyme DNA ligase.

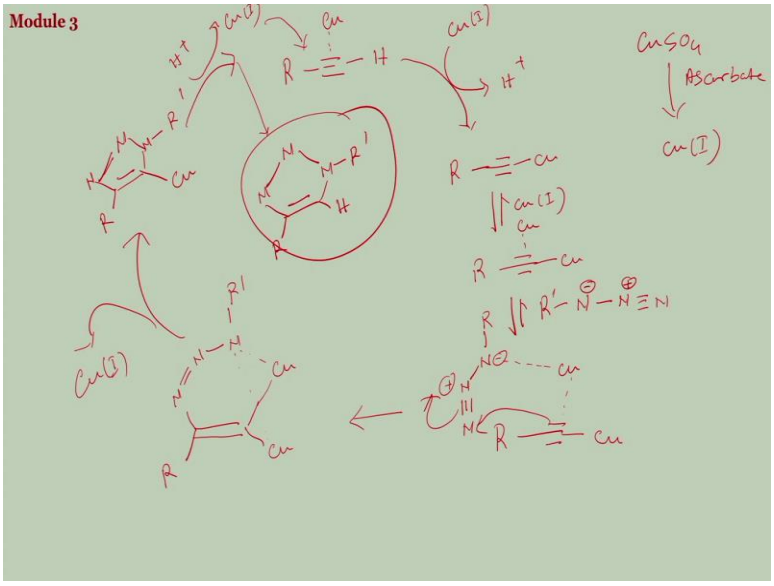
So, in this case, you can ligate it to DNA is a chemical method. That is why this is known as chemical ligations. And it has found literally a lot of applications in biology. If you want to connect 1 gene with the other or one modified DNA that if you have that you want to use it as a probe with the target DNA, which you can again modify then you can use the method of chemical ligation in situ to gate or to obtain variation or to obtain longer version of the DNA for various of course, depending upon your aim and your applications.

Similarly, you want to for example, make a covalent bond between a protein and substrate, which can be a drug or which can be any other bio molecule can be DNA can be the smaller small molecules, for example, vitamins, antibiotics anything. So, in those cases what you can also do this has been applied really well, they are, so your proteins contains many side chains that has been amine groups, NH2, which you can easily convert into N3.

And then, if you are the other molecule, which you want to connect it to, for example, the antibiotic or for example, the other substrate for example, or drug, if you modify that with an alkyne, then you can easily connect them together and study the protein functions or what the protein actually does in the leaving cells. So, this is called the chemical ligations. And of course, the other advantage is not advantage other difference is that you have now, when the backbone becomes triazole ring, this is hugely reach in pi electrons also in hetero atoms.

So, this is a planar system first is not a planar system this becomes a planar system. So, this will give you a different orientation of your DNA and also it will increase the pie stacking interactions as well. So, they are used as a modified versions of DNA as well. So, all you need to do is convert the normal functional groups such as weight here into this. Similarly here there are many ways to do that so, the mechanism of this reaction. You can find out from click Chemistry is quite popular I will show you here briefly.

(Refer Slide Time: 36:03)



Now let us start with a general mechanism. If you have alkaline RC triple bond C8. Then the first thing is this binds with the copper. So this is copper 1 so I will not write copper 1 again I will just simply write copper, which means copper 1. So copper is coming in here first in the pi electron, then comes in here, it eliminates proton and takes the place of the proton. So this becomes a copper now, another copper comes in next forms in a complex here that tracks the

electron density and here comes you R prime N3 as I now N3 can be written as this so, neutral, this become plus this becomes minus.

So that is what is your as N3 always exist in of course, in the resonating structures we can write other resonating structures as well, but this is one of the form. Now, if this happens, then what will go rewrite this into proper form you have the offer and then R prime in R prime in triple bond N so, this is plus this is your minus now, you have the this copper here that makes a complex session this next complex session. Now copper is electro positive.

So, it wants to push the electron density towards this, which increases the electron density on this carbon that can act as a nucleophile. This will be coming there so, what you are getting now is a double bond R in R prime with copper and here with the copper this and they have this is originally surplus has shown that this nitrogen forms a complex session in this kind of thing this dotted, but nevertheless this is the raw skeleton that comes here.

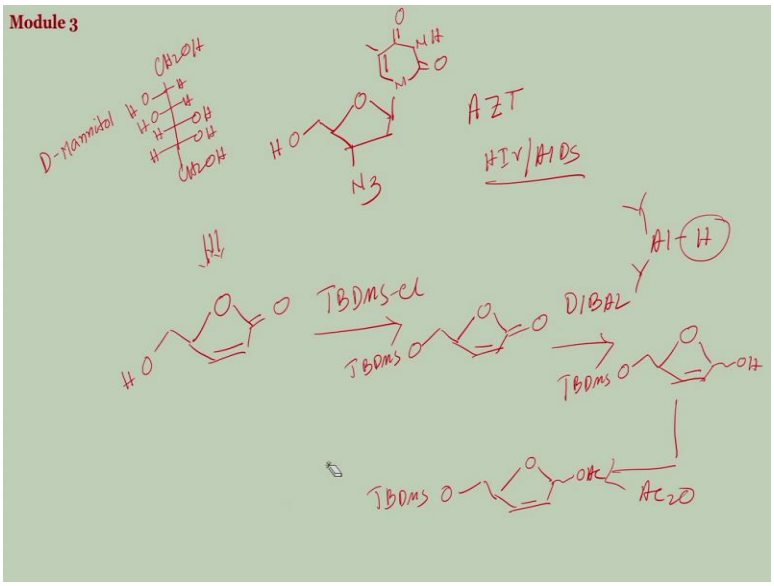
So, next is of course the elimination of 1 copper. And then you have here the copper was covalently bonded, you know, the double bond here is R and there is this nitrogen here. Now it is a double bond and this is nitrogen with R prime yes, this is what you have now, this copper has to be eliminated with replaced by the H. So this is what it will happen that H+ that you have released here will come and copper 1 will be released just the reverse of this and that will give write it here that will give you here the H the central compound.

And this copper is eliminated which can come back to this cycle. So, the cycle can continue that is why copper 1 acts as a catalyst. So, this is what the proposed mechanism of the reaction Sharpless reaction now, when you were you know practically when you are using the copper 1 as the catalyst, you do not have to use really copper 1 as a key compound. You can use also the copper 2 like CuSo4 which is copper 2 or any other salt because they are cheaper and if you want to convert it into you have to convert it into copper 1 because that is the active catalyst.

So, you hope to use some they use ascorbate or ascorbic acid so that will reduce this into copper 1. So this is what is the reaction of the click chemistry so, I am giving you this I am talking about

the click chemistry just to make you familiar with you that in future if you are thinking of making any covalent bonds if you are thinking of connecting to conjugating different kinds of biomolecules, then this is one of the quite popular method to use.

(Refer Slide Time: 42:55)



So, now we will move on and I have talked about that if this is your thymine, this is your thymine NH and this is the sugar N3 here and OH there. So, as I basically how can you synthesize this molecule you can start again with the molecule of your choice. So this molecule is actually our famous drug famous medicine. This is called A Zed T. This is one of the first HIV agents, HIV AIDS and very popular medicine actually very, popular drug. So, we have to synthesize the A Zed T molecule, you can find out the full A Zed T.

As, I tell you, you find it out. It is a very famous molecule. So here they originally they have synthesized it starting with the sugar. But the key material from where it actually starts is this actually comes from the sugar this is derived from the sugar which is D Mannitol multi steps, reactions and it comes down to this. So, from here if you look from here, how do you get into the compound? So first thing that you that you need is the production over here, which you can do by DMT or you can do by salalations.

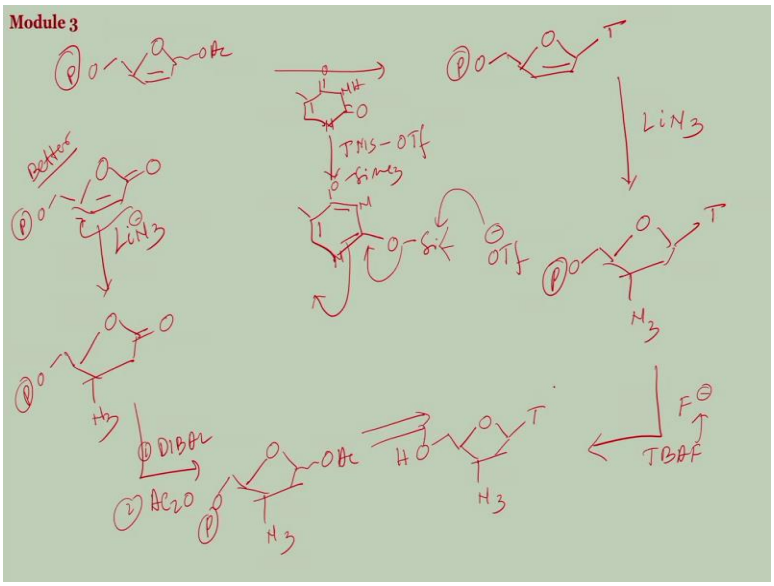
So, if you use the site, you can use it by salalations also, TBDMS chloride, if you it was that then you can get O, TVDMS you can do it by DMT as all DMT is little bit fragile. So if you want to

do multiple reactions after DMT it is better to ever DMT at the beginning. Now you to bring in some functional roof here you can do that or how to bring in your nuclear base in this solutions. So they have forced attested with the nuclear waste.

So, if you want to address the nuclear base, so, the first thing you have to do is to reduce this OH into alcohol into the OH TBDMS. Now for this reduction in how to use the reducing agent. Usual reducing agent is sodium mono hydrate, lithium aluminum hydride and all this. But of course, you know, since this is only our single carbon reduction, you want to avoid the multiple H minus the reagents which has multiple H minus for deductions.

So you use Dibal here that is single hydrogen or single hydride transferring agent basically it allows a single hydrate transfer the structure is it has a single one hydride base die isopropyl, lithium aluminum hydride basically. So, it has only one H. So that so, it will not affect the other part of the reactions. So, you have OH which you can convert it into to do the acetylation make a better living group little bit better living group at least OAC, OTB, and DMS. And from here you can make it further this a better living group by reacting it with a triflate OTF. Now, if you directly do this chemistry I will go to the next page better.

(Refer Slide Time: 49:08)



Starting here OAC I am just writing P for protection you need a thymine. So, thymine if you if you start with thymine if you treat it with TMS because we need we have talked about this when

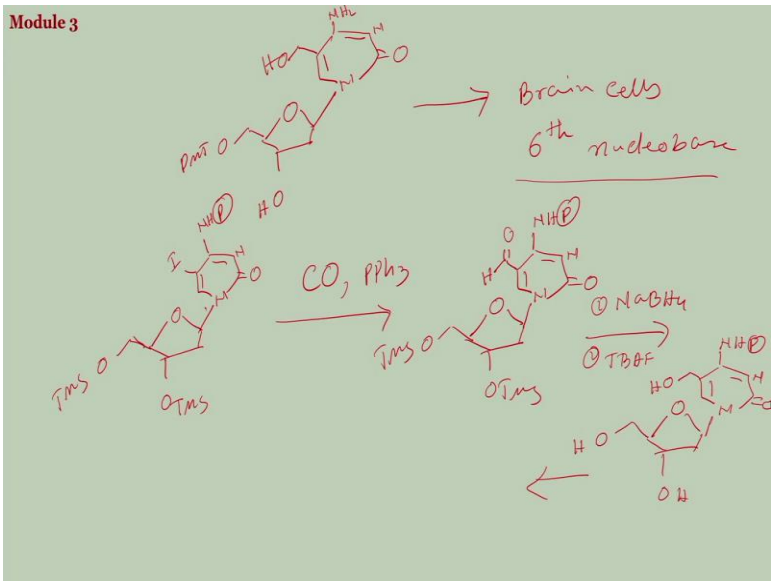
we have done the nucleotide synthesis, that the nucleophilicity here of this carbon is not very good reactive, the all the sugars are not very that much reactive, so you have to increase the nucleophilicity of this of your base basically so, and we can do that by converting it.

So, you can use TMS O triflate or TMS chloride both will do in this will give you O SiMe3 and this basically increases the nucleophilicity here how here or there you have the OTF minus that was released here. So, this will come immediately this will go and that is how you are in nucleophilicity of through this nitrogen is increased. So this will react here eliminating the OH OSe and you can have are now writing T then OP.

Now this is reacted with LiN3 nitride along with the solvents so as I do have T then this will include in N3 over here OP it can go either they are also so it can have you can have a mixture of them and then of nitrogen TBAF if that is the nitrogen and get this back is one way. Other way is that I will do it quickly OP you had first you can do with the LiN3 and this is even better actually right here better procedure.

So, because of this conjugation, your nucleophilic substitution will probably mainly happened in this case. So, you have this, the audience you will come here, selectively almost selectively. This is better. This is for me, this is better and then you can do the reduction and other way dibal reduction followed by one followed by acetic anhydride to that will lead you N3, Op and OSe and then from there you can add the thymine over there. This is a better way.

(Refer Slide Time: 53:37)



And lastly the compounds that we are doing here that compounds that I am writing the structures of our all kinds of all important kinds of components which are either biologically found or they are active they have properties in biological systems. Let us see O DMT and here is I mean so, this is the cysteine. Now cysteine if you have aldehyde group here, this is called from here it will come. This was actually the CH2OH, this derivative of cysteine this is actually frowned as nuclear ways in many brain cells.

And this has been invented quite recently, I think, 6, 7 years back was found a different nuclear ways, not in present in all genome but present in certain kinds of genes or cells. Brain Cells are the genes of tenting the brain cells content, a certain quantity of this variation, and this is known as the sixth nuclear base. So, ATGC and eurosil, these are the 5 nucleobase is present in the cells living cells, and this is the sixth nuclear waste that is sometimes present.

So, how to synthesize this molecule I will just throw briefly if you want to make this molecule in large quantity in laboratory wait. This I mean to protect with NHP with the switchable production group you can use that we have done in and here you start with the compound from here using our metal coupling reaction that is Carbonyl basically phosphine and the other the solvents, these are the key ingredient CO is the key ingredient in here.

From reaction basically in NHP, then you will have a CO inside and eventually it would not be replaced and it will have the H so it will come aldehyde I forward to do one thing when you start this you have to start with the production, so if you start with the sugar then you can do the production. Either you can do TMS or as we have seen, you can use also the breezed 1 the TIPDSTL and a TPDMS chloride.

With the dice sialylation which goes to 1 and 5, 3 positions that also you can do or if you treat it with TMS both will be protected. So, here the TMS for example, here TMS and then you have the insertion. This you can reduce with the sodium borohydride or maybe a dibal also sodium borohydride is fine here followed by opening of the so if this is one, your second would be your TBAF, NHP and this would be reduced into the weight so that is why the reducing agent is 4.

And then you can convert it into the TMT so, this is a naturally occurring of course, nucleobase which you can see this find a synthetic route to make more amount of in your laboratory. So, these are some of the synthetic problems that I intended to solve and to show that you can play around. So, you can find again if you have a target molecule, you can figure out in and around of what we have got that how to synthesize those kinds of molecules, especially the very essence in the nucleic acids. Thank you.