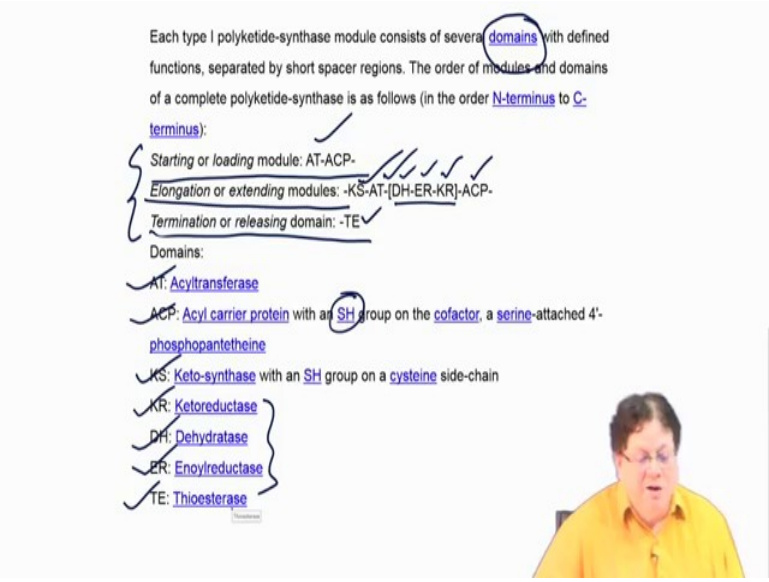
Organic Chemistry In Biology And Drug Development
Prof. Amit Basak
Department of Chemistry
Indian Institute of Technology, Kharagpur

Lecture – 55
Biosynthesis of Macrolide Polyketides and Introduction to Virus

Welcome back to this course on Organic Chemistry in Biology and Drug Development. We have finished the biochemistry part; organic chemistry in biology means basically biochemistry. And, in the drug development area, we have also studied a number of topics. And, the last topic that I had taken up was the polyketides.

(Refer Slide Time: 00:58)



Because, we are studying the antimicrobial agents and under the heading of antimicrobial agents, we have different antibiotics and then we have done how these antibiotics are made, specifically the biosynthesis of penicillin. We also have done how the bacteria acquire resistance against these antibiotics or antibacterial agents. And, then we went on to a very important class of compounds which are called polyketides.

The characteristic feature of polyketides or secondary metabolites, which follow the polyketide biosynthetic pathway, are that they are having either alternating $COCH_2$ group or they could be having a double bond like CH double bond CH or they could be having all saturated carbon framework, but all these 2 carbon units are derived from a keto-methylene.

So, basically it is keto-methylene which is present alternately; but the status of oxidation of the carbon bearing the carbonyl can vary, it can be carbonyl, it can be hydroxyl, it can be double bond or it can be even saturated one. Now, I told you about the different types of polyketide synthases; one is PKS type I and then PKS type II, which is the major classification; PKS type III is also possible.

PKS type I is basically a single enzyme containing different modules and each module has got different domains of activity. So, first be very clear, that it is a single enzyme with different modules. Basically, it is like the different the compartments of a train. So, each compartment is basically a module and each seat inside the compartment is basically the domain, where you sit.

So, in polyketide synthase I, you have this modular system; one particular multifunctional enzyme. And, then you have different modules; each module has different domains. Then there is type II where you have different proteins individually; that means, these enzymes are not attached with each other to start with, but at the time of synthesis they come together and then react in a cooperative manner to synthesize the polyketide.

But, we are basically discussing the PKS type I, which involves the modular type of synthesis. Here there should be a loading module to start with. Loading module is the module where the building blocks get attached first and then that is transferred to the first module. In the loading module, you have an acyltransferase and you have an acyl carrier protein.

The acyltransferase puts the starter unit, because there are now 2 units that react with each other; one is a starter unit, which only participates only once, that is in the beginning. And there is the extension unit, that extension unit repeats itself over and over again depending on the number of carbons that the microorganism or whatever system we are talking about that wants to put.

So, depending on that you will have these number of extension units. See this will be a little bit repetition of the last session's topic, because it is little bit complicated and it is entirely new. Usually it is entirely new to the students. So, I am spending more time on this.

Type I polyketide synthase module consists of several domains; I have already explained what domains are. And, then the important thing is starting or loading module that is an acyltransferase and an ACP. Then you have elongation or extending modules, sometimes it is called elongation unit; that means, you extend that chain that is why it is called extension unit or you elongate the chain so, it is called elongation unit.

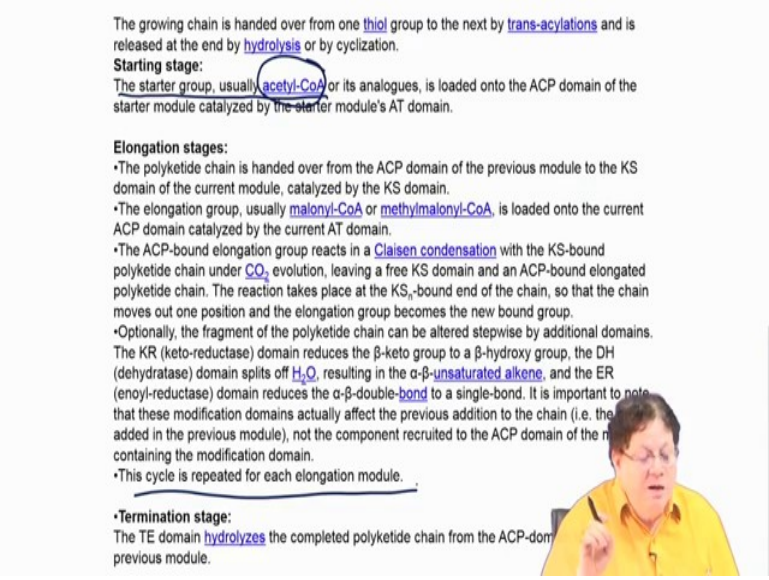
So, that elongation or extending modules comprise ketosynthase; that is obligatory, that has to be there; and acyltransferase and an ACP. These 3 have to be present in the module, apart from that, within third bracket it is written ketoreductase is there, enoylreductase and dehydratase. Again, I repeat, you should not worry about the sequence how it is written. The important thing is that what type of functionality or domain is present.

So, these may be present or they may be completely absent or one or two may be present. So, ketosynthase, acyltransferase, ACP are the obligatory domains in a particular module; apart from that you can have ketoreductase, you can have enoylreductase or dehydratase. And, then this is repeated over and over again; then it goes to module 2, then module 3, module 4. Again, I repeat depending on the number of extension units that the microorganism wants to put, the living system wants to incorporate.

And, then at the end, when the desired length of the polyketide has been made then what happens? There is a termination domain, because you have to terminate the process and that involves thioesterase. Because, at the end of it, the product is a sulfur in the form of a thioester, which is linked to the ACP. And, then you have to release it from the ACP (acyl carrier protein) by hydrolysis. So, it ends up as a carboxylic acid or it could be a cyclisation that is intramolecular transacylation which can result in the formation of cyclic system.

What are the domain activities? Acyltransferase, I already told you, acyl carrier protein that has got an SH to hold up the extension or the starter unit. Then, ketosynthase; that is the major reaction, which carries out the Claisen type condensation *via a* decarboxylation process, then you have a ketoreductase, dehydratase, enoylreductase, these are not obligatory; depends on the nature of the product that the living organism wants to make and then there is a thioesterase. So, this repeats and then finally, the thioesterase comes.

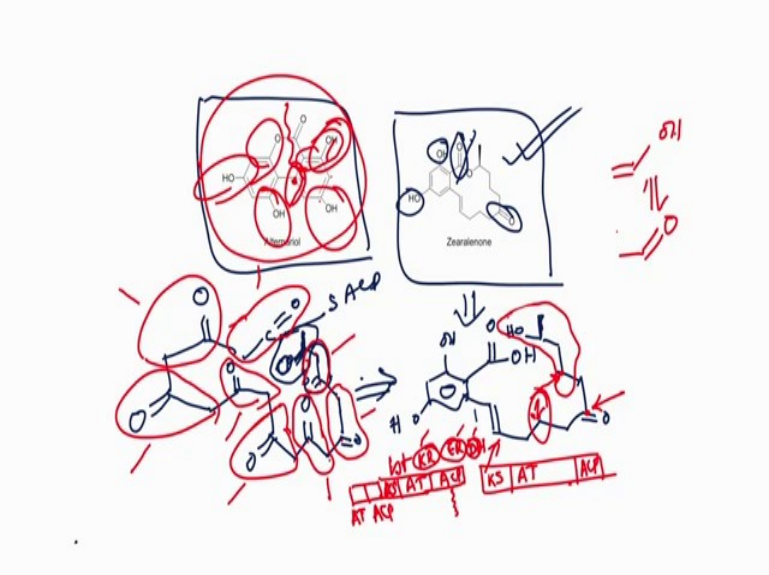
(Refer Slide Time: 09:09)



So, it is not difficult at all. You can actually do retro-biosynthesis; if I give a natural product, you can do a retro biosynthesis. This is basically repetition, the starter group is usually acetyl coenzyme A, but again that may vary also; it could be propionyl coenzyme A as you have seen in case of in case of Erythromycin biosynthesis.

The starter unit was propionyl-CoA and the extension unit was methylmalonyl-CoA. And, so, that is the AT and the ACP domain; then elongation stages occurred. And, finally, this cycle is repeated for each elongation module and the termination stage comes when the desired length is obtained.

(Refer Slide Time: 10:07)



So, now let us do some exercise; let us do retro-biosynthesis. You know what retrosynthesis is. Now, this is what is called retro-biosynthesis. Now, there are the 2 polyketides; they are natural products and they are biologically quite active. Basically, they are some of the toxic agents that are present in many of the cereals or the staple food that we take like wheat, maize or rice; all these things. So, they have toxic effects. So, these are important because your cereal should not have any of these toxic agents.

There is one more very toxic agent that is called aflatoxin, but anyway we are not taking aflatoxin which is a bigger molecule and also a polyketide, but we should take a jump step by step. Now, if I want to first take this one. This is a polyketide, how do I know? Is there any signature that after writing the natural product you see that whether it is a polyketide? Usually polyketides have lot of oxygens in it.

There is no hard and fast rule, but it is basically an intuition that there are lot of oxygens that will be present in the molecule. And, these oxygens are usually having 1,3 relationship. So, I see there are a lot of oxygens here, but there is lack of oxygen in many of these carbon frame works.

Now, in the polyketide synthesis, if you again analyze carefully, the terminal thioester which is hooked up with the ACP that ends up either as carboxylic acid or that ends up in the form of a lactone. So, you try to find out where is this carboxylic acid, where is the

carboxylate function. So, you see the carboxylate function is here because this is an ester or a lactone. So, you can break at this point. So, if you break it there, that is the starting point.

So, if you break that what you have is CO OH, a carboxylic acid and then you have this double bond, then double bond O and this is the methyl and you have OH. Now, as there are two OHs here OH and OH, so now you work backwards. What will happen? See this is your basically COS-ACP, the final before the hydrolysis. And, then you start from there and then work backwards.

So, you have a CH_2, you have a CO; just put whatever is the redox level; if you see OH put a carbonyl here for the time being and then there is another oxygen here put as a carbonyl and then there should be a carbonyl here by rule that 1,3 position there will be oxygen. So, there must be a carbonyl that was there, but somehow ultimately through some condensation reaction that carbonyl is no longer there and then you continue. So, there should be a carbonyl here; then a carbonyl is already present.

So, there must be a carbonyl here in 1,3 position to match that and then again after the methylene, you should have a carbonyl here and then you have methyl and OH. So, OH means carbonyl. So, there is a methyl here, this is a S-ACP and this is the methyl that is present here.

So, what is the starter unit? Starter unit is this acetyl coenzyme A, then you have how many extension unit? This is one extension unit, this is the second one, this is the third one, this is the fourth one, this is the fifth one, this is the sixth one and this is the seventh.

Let us see, this is CH_2CO, then you have CH_2CO, then you have CH_2CO, then you have CH_2CO and then you have CH_2CO.

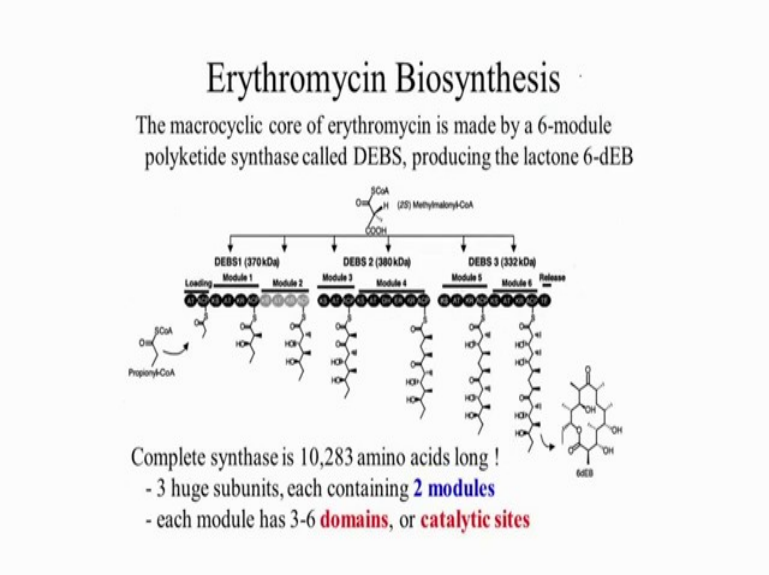
So, how many extension units are there? There are 8 condensation reactions that will happen and if it is synthesized by a type I polyketide synthase, then you can write the modules. I will just write a few of them; AT and then ACP, that is the loading module, then the second module, remember it all starts from here. The reaction starts from here so, the second module means you have gone up to this point.

The first module is the loading module. I said that each module should have a ketosynthase right after the ACP, then you should have AT. And, the question is that whether you should have other ones or not, because this ends up as a CH_2; that means, it should have ketoreductase right here, you should have a dehydratase and then enoylreductase. Again the sequence does not mean much; it is just the activities that are important. So, this should be the first module. The first module should have all the possible domains that are required, because it ends up as a saturated system.

So, it should have reduced the carbonyl. So, initially it will be carbonyl, but then it will be reduced, then the alcohol is dehydrated and then the enoyl compound is reduced by enoylreductase. And, all along it is held up by the ACP. ACP is holding the whole chain with the carbonyl at the β-position and then first it shows it to the ketosynthase and then it is reduced; then it takes to the dehydratase, so it is dehydrated then it takes the molecule and shows it to the enoylreductase and then the enoylreductase reduces the double bond, then the job of the first module is over.

So, now the ACP will transfer that whole chain into the second module. Look at the modules of erythromycin, I can show you.

(Refer Slide Time: 18:59)



Each module starts with the ketosynthase and ends up with the ACP. In between there are these acyltransferase, then you know these ketoreductase; all these things are present.

So, what happens? First the loading module transfers it to the ketosynthase, reaction takes place between the ACP and the ketosynthase.

The whole chain is held up here and this is OH because there is a ketoreductase here, then it is transferred to the next ketosynthase. So, all modules end up with ACP and starts with a ketosynthase; they are side by side. So, that the ACP can transfer it immediately to the ketosynthase of the next module.

So, I will not proceed any further; I will not complete that, by this way you can tell what is the second module. In the second module, check what the status of the carbonyl is and you see the carbon is having a carbonyl; that means, second module has a ketosynthase, then acyl transfer is there and last there should be an ACP; is there any other domain? No, there is no other domain, because it is the carbonyl.

And, then the next module; the next module should have again all 3, because it is CH_2. And, after that the module that is the first module, this is the second module, that is the third module, the fourth module now, there is this double bond.

So, what will happen now? That means, ketoreductase is present and dehydratase is present. So, by that way you can complete the entire biosynthetic gene sequences and along with the entire biosynthetic gene and then also assign the protein that is actually translated from them.

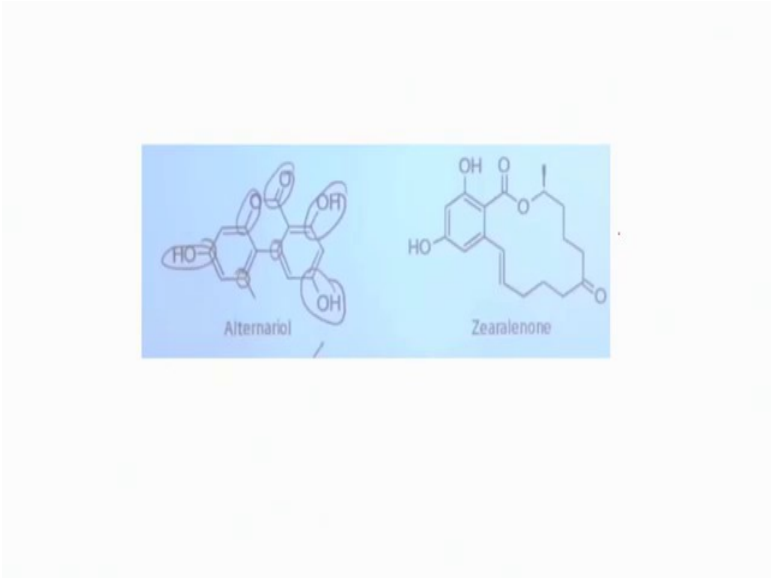
So, this is the way you do it. If you want to do for this compound, this is another secondary metabolite, again you find out where is the carboxy; again there is a lactone here; that means, there is a carboxy here. So, that will be a carbonyl; this will be another carbonyl and then it goes like this. So, then the next carbon is this one. So, that is there; this is the other one. So, by this way then there is OH, then there is this one, CH_2CO like this is the CO SCO ACP.

And, then you have CH_2CO, then CH_2CO, then CH_2CO, CH_2CO, CH_2CO, CH_2CO. Only you have to find out what is the status of this; what we are assigning carbonyl ultimately in the metabolite what is their status. I could see that if you break these bonds; actually they are all carbonyls, because phenol is an enolic group. Enol means actually that is the other form of the carbonyl generated through tautomerization.

So, when you see an aromatic ring with a OH; that means, actually that is the ketone functionality, that is not a reduced carbonyl functionality, because phenols means actually enols and enols means it is the tautomeric form of the ketone. So, this way you can work out the biosynthesis of a natural product; if some natural product is given to you. The first thing you see is that what are the oxygenation patterns and then if there are a lot of oxygens. Accordingly you might think of the polyketide and then if it is a polyketide, then try to put the oxygens at the proper position; identify the starter unit and the extension unit. And, then you can finally work backwards and find what are the starter and the extension units.

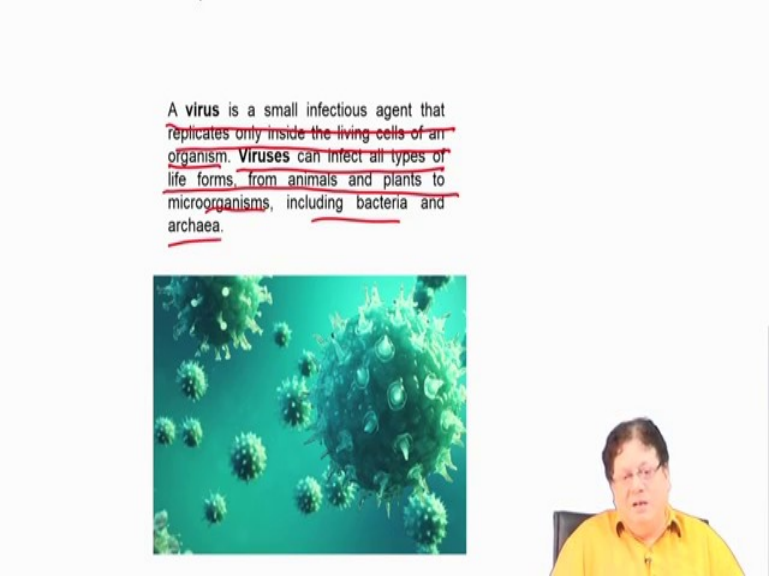
How to prove that is the extension unit and that is the starter unit? They put ^{13}C-label starter and then find out where it ends up or you find ^{13}C-label extension unit, and then do by NMR study of the secondary metabolite; that whether that extension unit has been incorporated or not. That is just by simple NMR technique, but you have to use isotopically labeled starter units or extension units.

(Refer Slide Time: 24:00)



So, that is all about the polyketide synthesis; I think that is the end of the antimicrobial agents. Now, we will go to another topic that is that not microbes, because microbes are organisms which can amplify by themselves, which does cell division and then which can grow by themselves provided sufficient food is added from outside or it can extract the food from outside.

(Refer Slide Time: 24:44)



Another very important and tiny entity, which also infects, cause infection in the body is the virus. Microbes do that, microbes include bacteria include fungi, and parasites are included in that. But apart from that, you have something which is called a virus; you all know this virus and we get viral infection quite often. And, virus actually do not fall into the group of microbes, because microbes by themselves are capable of multiplying whereas, virus cannot multiply by themselves, they need a living organism to replicate and ultimately do multiplication or amplify to new virus particles.

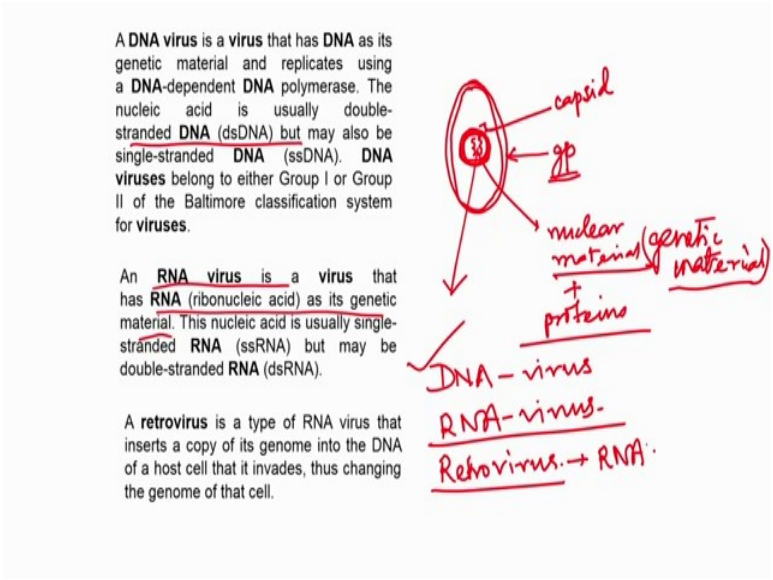
So, a virus is nothing, but a small infectious agent that replicates only inside a living cell of an organism. So, there is a difference between microbes and this virus. Viruses are non-living as such outside, when they are left alone, but if they enter inside the body, they utilize the machinery of the host and then try to grow. So, virus can infect all types of life forms, they can infect animals, even plants are susceptible to viral infection and plants to microorganisms including bacteria, archaea, humans. We suffer from more of viral infection than bacterial infection to be honest to you.

Now, the question is how do you design antiviral agents? Obviously, virus's life cycle will be entirely different from the microbes. I have shown you that there are different targets for bacteria; you have the cell wall, you have the membrane, you have the nucleic acid, you have the proteins, you have the metabolites; all these are there. You have bacterial transcription, so you can stop the translation process; all these things are there;

but in virus, those type of mechanisms are basically hijacked, those mechanisms by which things multiply they belong to the mostly belong to the host.

So, if you try to stop those processes; basically you are interfering with the hosts own mechanism to multiply. So, that is the problem and that caused much delay in developing antiviral compounds. So, in our next session, we will talk about how to develop antiviral agents, but now before we end this session, we would like to know what the different types of virus are? What is their life cycle? How they multiply, when they get inside an organism?.

(Refer Slide Time: 28:01)



Viruses are different types, but the one common thing that all viruses have is an envelope (outside coating) made up of glycoproteins and then there is a circular thing, which is called capsid. Inside the capsid, you have the nuclear material. I am not saying it is DNA; you have the nuclear material, which helps it to multiply by hijacking the host organism.

Now, these are glycoprotein, these are very important. Because, the glycoprotein actually recognizes; if a virus enters into my body, the virus particles are first recognized by my cell, the cells that are present inside my body. And, these glycoproteins help in this recognition process. Our cells are all having receptors; we know now what receptors are;

these are actually basically present on the membrane of the cell and with the receptor moiety pointing outwards to the extracellular side.

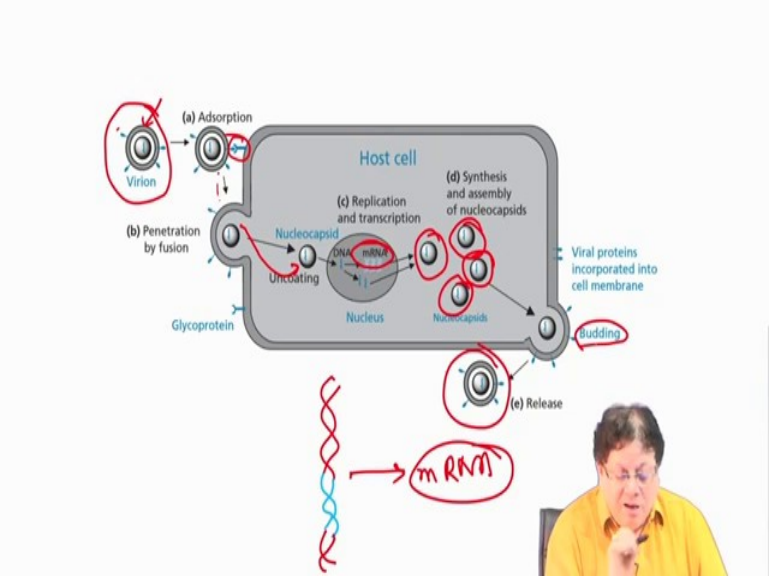
So, there are different types of interactions that can go with this receptor site. So, the virus particle comes and sits here, but the recognition is through the glycoprotein. So, envelope is very important. And, then inside there is again another circular coating thing, which is called capsid and the capsid is enclosing the nuclear material along with some proteins. So, inside there is this nuclear material plus some, not many, proteins which are essential for viral replication.

Now, depending on these nuclear materials, viruses are classified. Viruses are classified like if the nuclear material is DNA then that is called a DNA virus; if the nuclear material is RNA, then that is called an RNA virus. The RNA virus has a special type, which is called retrovirus; I will come to that retrovirus, but let us try to find out what is the DNA virus. So, DNA virus means as the name suggests that it will be having DNA as the nuclear material.

Now, the DNA may be present in double strand (ds) or single strand (ss), both are possible. RNA virus usually have a single strand and RNA is the genetic material. So, I can say nuclear material means basically this is the genetic material. Without it, the virus will not be able to do anything.

So, if the genetic material is DNA that is the DNA virus; if it is RNA then it is RNA virus and in retrovirus, the genetic material is RNA, but I will say what the difference between RNA virus and retrovirus is. In the DNA virus what happens?

(Refer Slide Time: 31:57)



This entire virus particle is called virion. The virion is having this envelop and with some receptor sites for the recognition and with which it binds; see this is the receptor on the cell this is the entire host cell. So, this is recognized by this glycoprotein. And, then by a process called endocytosis, the virus enters; not the whole virus enters, it is only the capsid portion that enters into the cell.

And, now if it is a DNA virus, then that capsid goes off, the DNA is injected into the nucleus of the host cell. Though there are exceptions, usually that viral DNA now becomes integrated with the DNA of the host.

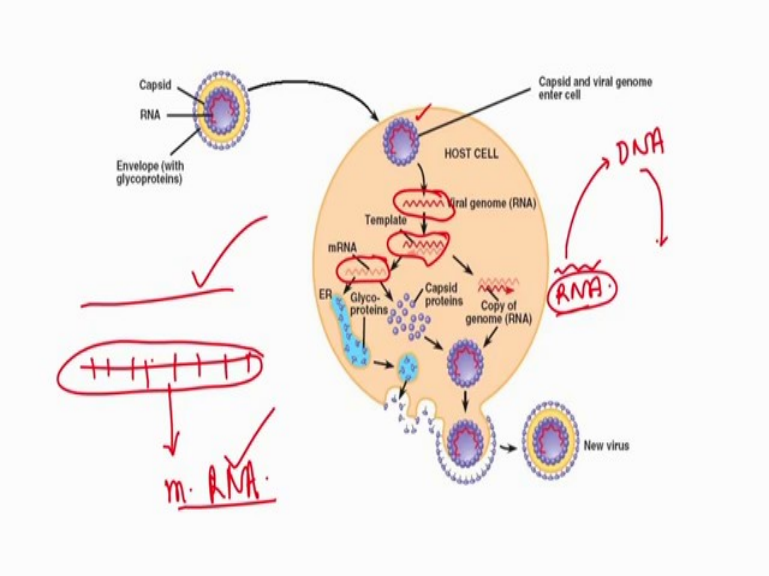
The other option is that it does not integrate, but it utilizes the host's DNA polymerase to have the replication of the DNA. And, then it can form the mRNA by utilizing the host's RNA polymerase. So, basically the virus is recognized by the cell, the envelope falls off and it goes inside with the capsid. The capsid falls off, the DNA is injected inside the nucleus and then there are 2 possibilities. Either that DNA now does not integrate with the host DNA or it can integrate; these are the 2 possibilities.

If it integrates from that the normal process, like replication and transcription, that will take place. So, the transcription means you are making the mRNA. And, mRNA comes out. That means what is being made will contain some mRNA corresponding to the virus

DNA and they will make the required proteins and then basically form new virus capsids and finally, this comes out and an envelope is also formed around the virus.

So, for DNA virus, basically it is a direct way that DNA goes in, it can integrate or it may not integrate; remain in the nucleus and then by the nuclear machinery; that means, the DNA polymerase and the RNA polymerase, the mRNA of the virus is synthesized that comes out, it is encapsulated by the capsid, formation of all capsid means the required proteins, that is made from the mRNA and then it comes out; this process is what is called the budding.

(Refer Slide Time: 35:38)



In RNA virus; this is the envelope and this is the capsid and the genetic material is now RNA. So, it gets recognized and then goes inside. So, that is the capsid along with the RNA and then what happens? The capsid falls off. So, this is now the viral genome, but this is RNA.

You know that central dogma of biology is that always information flows from DNA to RNA to protein. After the discovery of the RNA virus and the retrovirus, that notion or what is called the central dogma that is becoming ineffective here. This does not hold true in case of retrovirus and RNA virus, because in RNA virus the genetic material is RNA. Now RNA cannot be integrated into the DNA, that is not possible, because the

RNA has to be taken to the DNA and then DNA follows the typical replication transcription and translation pathway.

So, the virus has RNA here. So, first the RNA is acting as a template. So, another mRNA is made directly from here. See, you do not have to go to the DNA now. Remember that when the double stranded DNA is copied one of this strand remains silent, that is the coding strand; that means, the ultimate sequence of the mRNA, will be same as the coding strand. The template strand is also called the anti-sense strand.

The template strand has the complementary bases here. So, now, the question is when the virus has the RNA, what is the status of this RNA? Whether it is the RNA corresponding to the coding strand or the non-coding strand; that is important. So, you have two types of RNA viruses; positive strand RNA virus, negative strand RNA virus. Negative means anti-sense, means non-coding, if you remember one strand which does not act as a template that is called coding strand; that is also called sense strand, now that is called positive strand.

Negative strand is the non-coding strand that is also called the template strand. So, depending on this what type of RNA is present, positive strand of RNA or negative strand of RNA, that will decide what type of life cycle the virus will be having. I think we will discuss this classification in the next session. For now on, what we know that viruses are classified into DNA, RNA virus and retrovirus. So, we will come back in the next class.

Thank you.