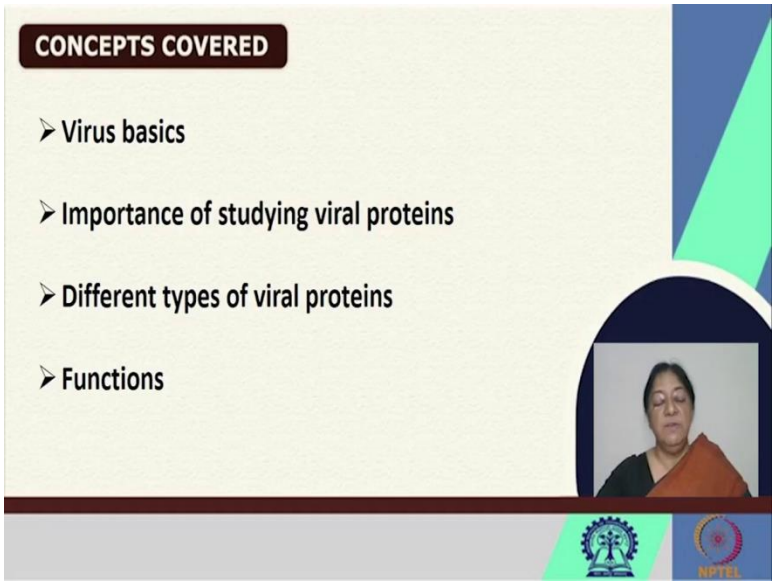
Fundamentals of Protein Chemistry
Prof. Swagata Dasgupta
Department of Chemistry
Indian Institute of Technology, Kharagpur

Module - 12
Special Topics in Protein Chemistry
Lecture - 59
Viral Proteins

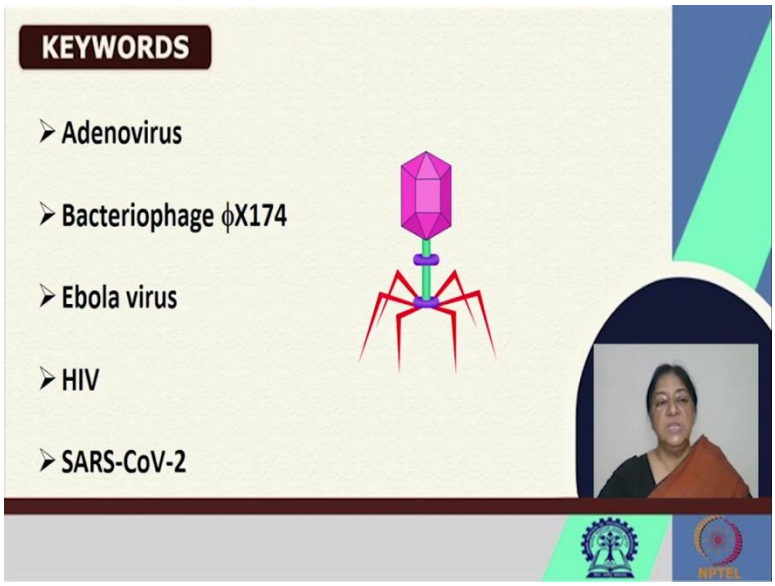
In the last of the lectures on special topics in protein chemistry, we will be looking at viral proteins.

(Refer Slide Time: 00:32)



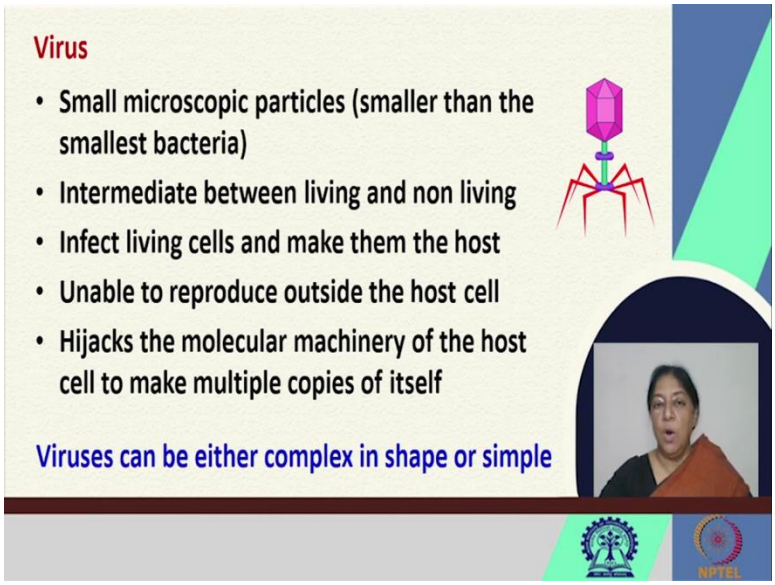
In the last lecture we will be looking at an overview of the course, the topics that have been covered and the interesting aspects of protein chemistry. In this lecture we will be studying virus basics, the importance of studying viral proteins, what are the different types of viral proteins, their compositions, their structures and the specific functions associated with them.

(Refer Slide Time: 00:48)



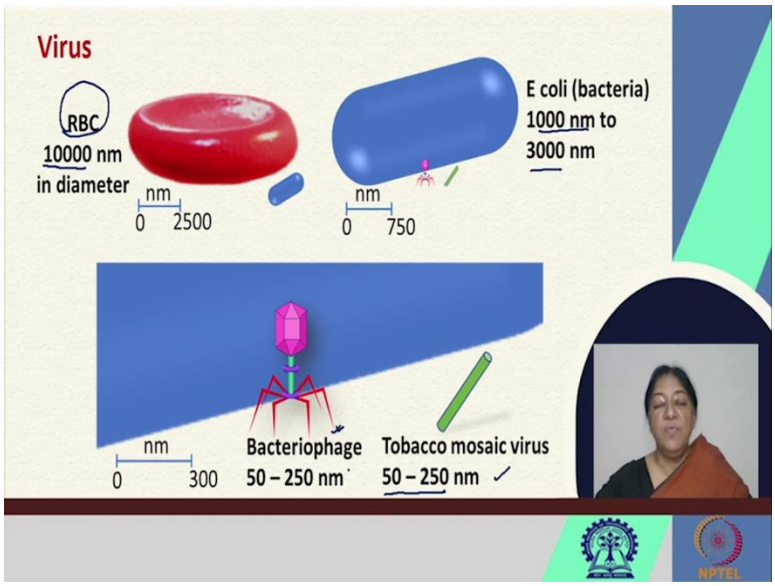
In the process, we will be looking at the adenovirus, the bacteriophage, the figure of which is given here [refer to slide], the Ebola virus, the HIV and SARS-CoV-2, given that it has been the most talked about virus for the past one and a half years and created this pandemic that has been associated with all of us and all of us have been affected in some way or the other.

(Refer Slide Time: 01:22)



The virus is a very small microscopic particle and it is smaller than the smallest bacteria known. It is intermediate between living and non-living organisms and it infects living cells and makes them the host. It does not have the capability to reproduce outside the host cell. So what it does, it hijacks the molecular machinery of the host and makes multiple copies of itself, thus creating a viral disease. These viruses can be either very complex in shape or simple in shape.

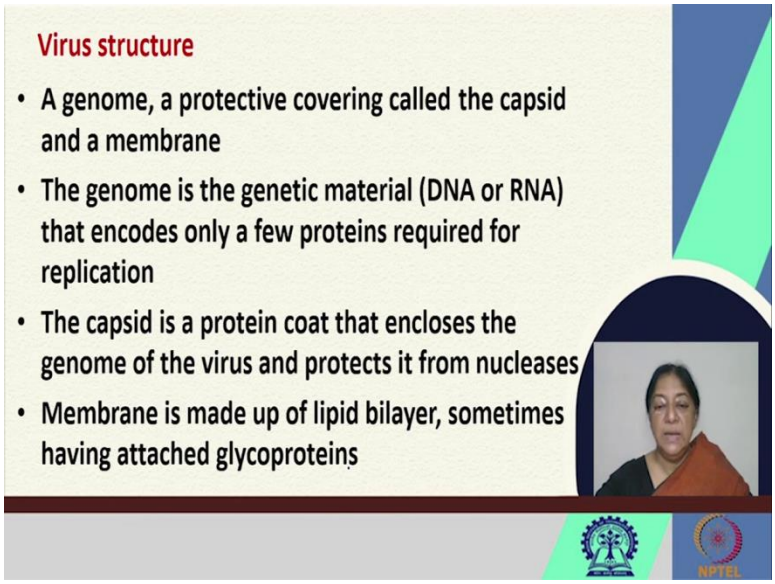
(Refer Slide Time: 02:05)



However, they have a common motive. If we look at the various sizes or the size comparison of the virus this [refer to slide] is the RBC, the red blood cell as we can see in the diagram here and associated with it, this blue rod is an E coli bacteria. The size ranges from 10000 nm to 1000 to 3000 nm.

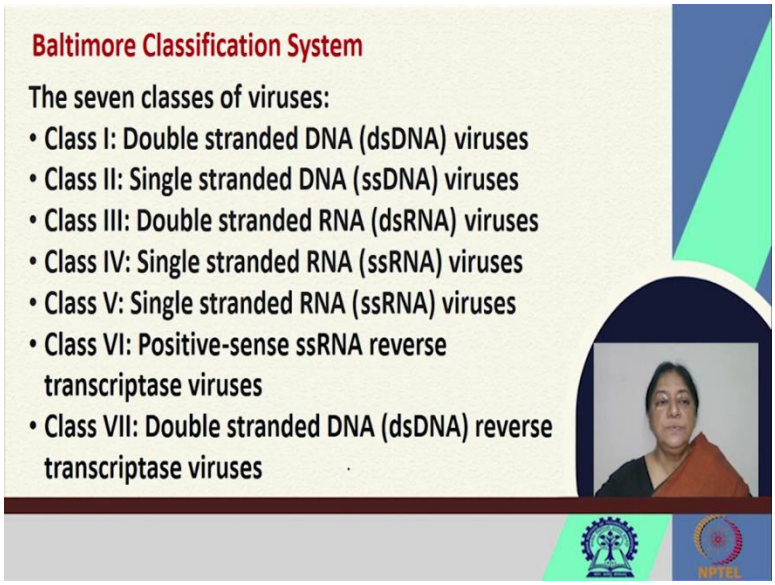
Following this we see another rod shaped structure and the tobacco mosaic virus that is 50 to 250 nm and the bacteriophage which is also 50 to 250 nm. So, we look at these size ranges and we realize that these small particles, the viruses can create a lot of havoc, based on the way they take over the machinery of the host cells.

(Refer Slide Time: 03:04)



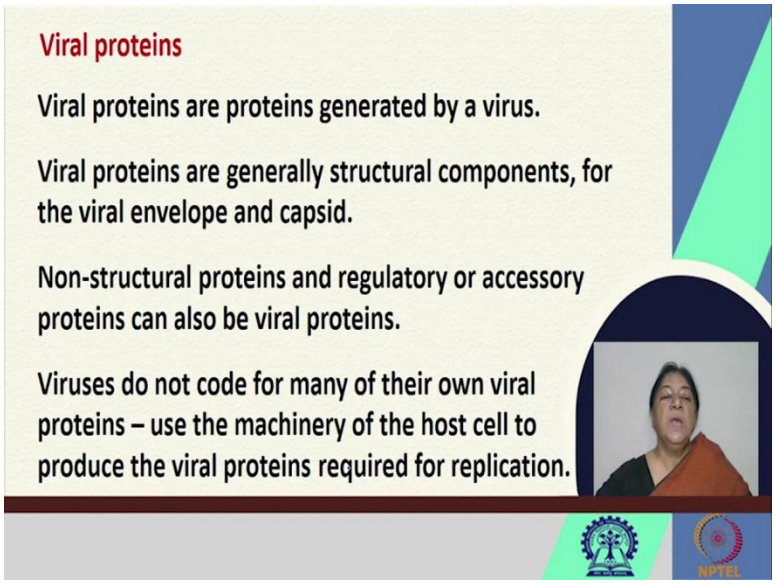
The virus structure is common to all. There is a genome, a protective covering called the capsid and a membrane. The genome is the genetic material, the DNA or the RNA that encodes only a few proteins that are required for the replication of the virus and the capsid is a protein coat that encloses the genome of the virus and protects it from nucleases, that may be present in the host cell. The membrane in this case is made of a lipid bilayer and has attached glycoproteins.

(Refer Slide Time: 03:45)



There is a classification system based on the types of viruses, the genome content of the viruses, whether it is double-stranded DNA or single-stranded DNA, double-stranded RNA or single stranded RNA so on and so forth.

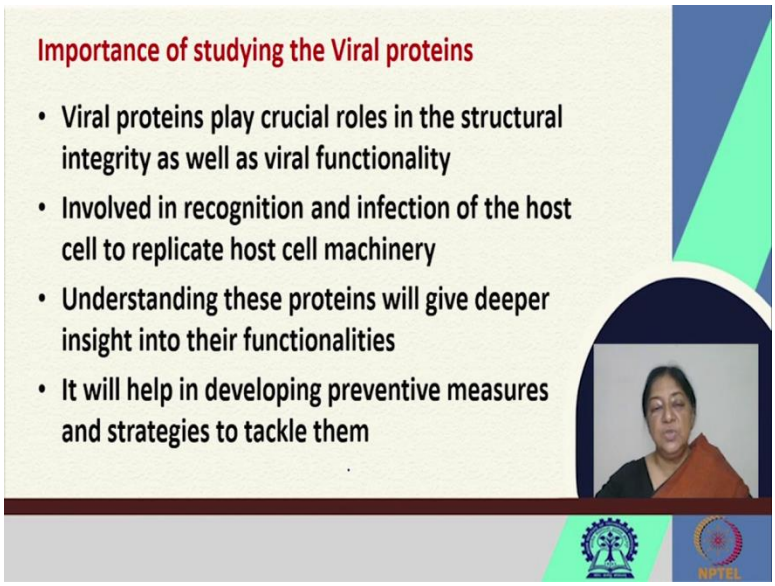
(Refer Slide Time: 04:04)



Viral proteins are the proteins that are generated by a virus. They are structural components that are required for the viral envelope and the capsid. The non-structural proteins and regulatory or accessory proteins, can also be these viral proteins.

Now viruses do not code for many of their own viral proteins. As a result of which they have to use the machinery of the host cell to produce the viral proteins that are required for its own replication.

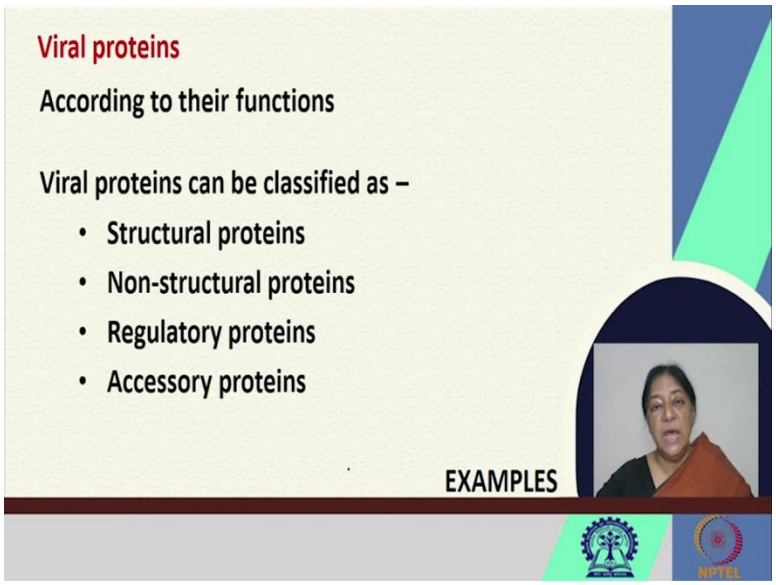
(Refer Slide Time: 04:44)



To understand or to know about the viral proteins, we realize that these play crucial roles in the structural integrity as well as their functionality. They are involved in the recognition and infection of the host cell, so that it can replicate the host cell machinery and produce copies of itself.

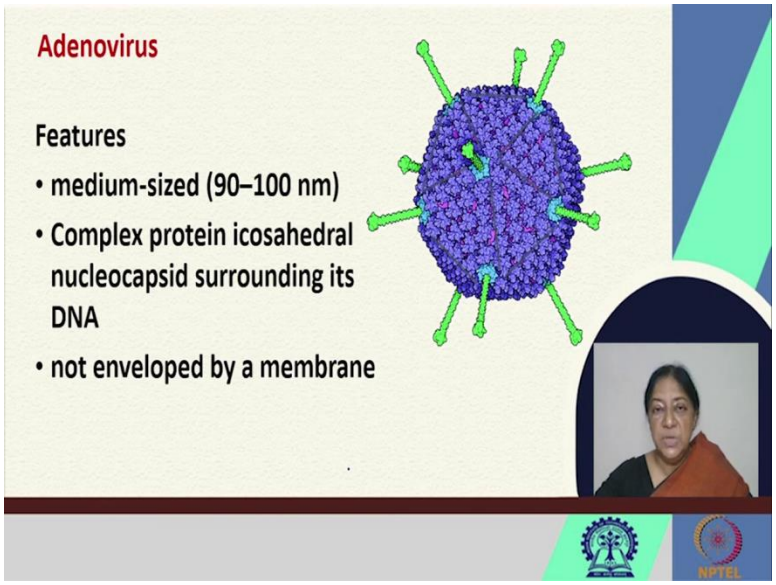
So understanding the proteins involved, is going to give a deeper insight knowledge about their functionalities and it can help us develop preventive measures and strategies to tackle them.

(Refer Slide Time: 05:28)



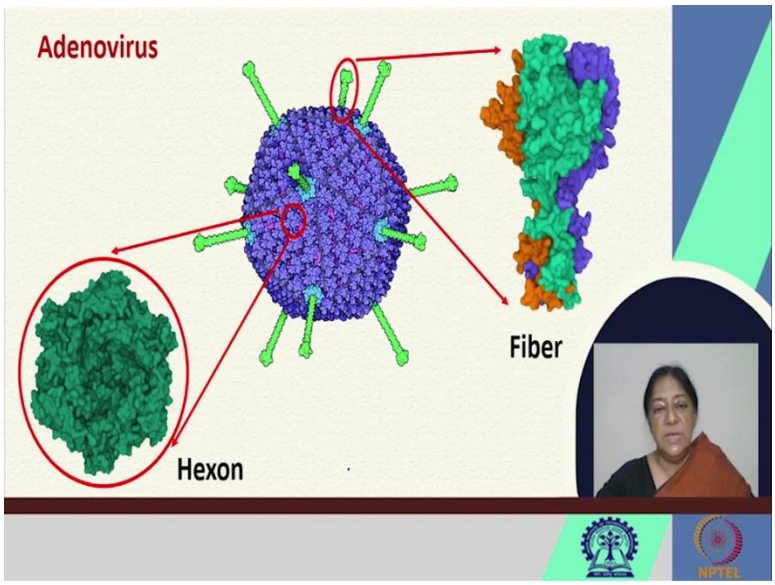
The viral proteins are divided according to their functions. They can be classified as structural proteins, non-structural proteins or what are called the NSPs, the regulatory proteins or the accessory proteins. Examples of each of these proteins will now be discussed and we will see how versatile they are in their activity, in their action and in their function.

(Refer Slide Time: 05:59)



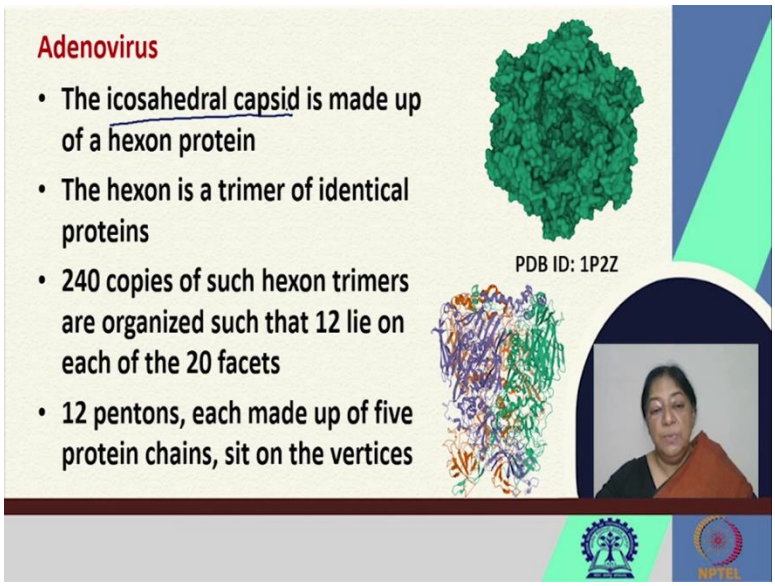
The adenovirus is a medium-sized virus of the order of 90 to 100 nm. The specific features are that the complex protein icosahedral nucleocapsid surrounds the DNA, that is the genome of the adenovirus. It is not enveloped by a membrane, so we have the genome material covered by this capsid protein.

(Refer Slide Time: 06:29)



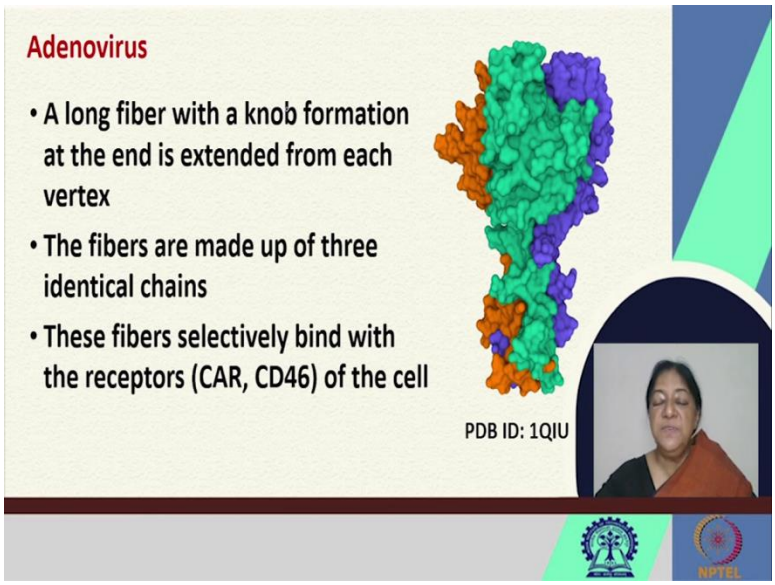
This capsid protein has a six group of structure that has six units to it called the hexon and there is also a fiber, that is also a protein on the surface of the adenovirus.

(Refer Slide Time: 06:46)



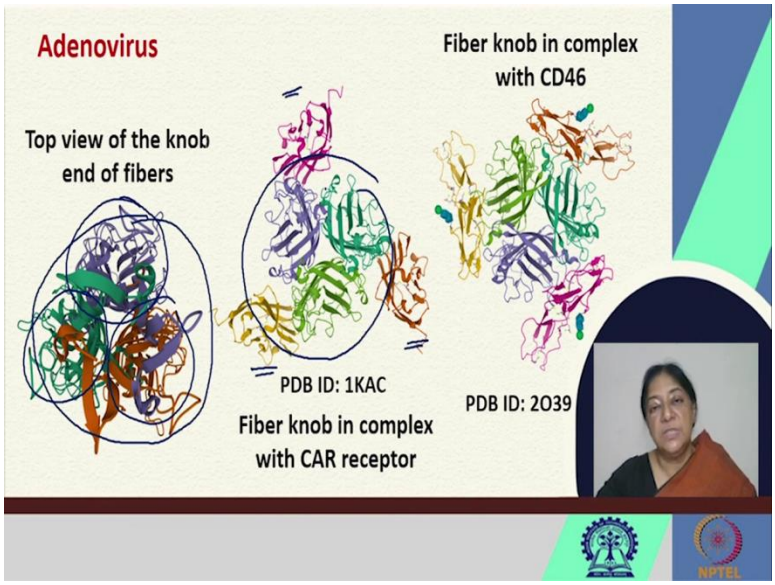
The icosahedral capsid is made up of this hexon protein, which is a trimer of identical proteins. There are 240 copies of such hexon trimers that are organized such that 12 lie on each of the 20 facets of the icosahedral capsid. Now there are 12 pentons, each made up of five protein chains that sit on the vertices.

(Refer Slide Time: 07:20)



Then a long fiber with a knob formation at the end, is extended from each vertex. This is what it looks [refer to slide] like and these fibers are made up of three identical chains and these bind selectively with specific receptors of the cell.

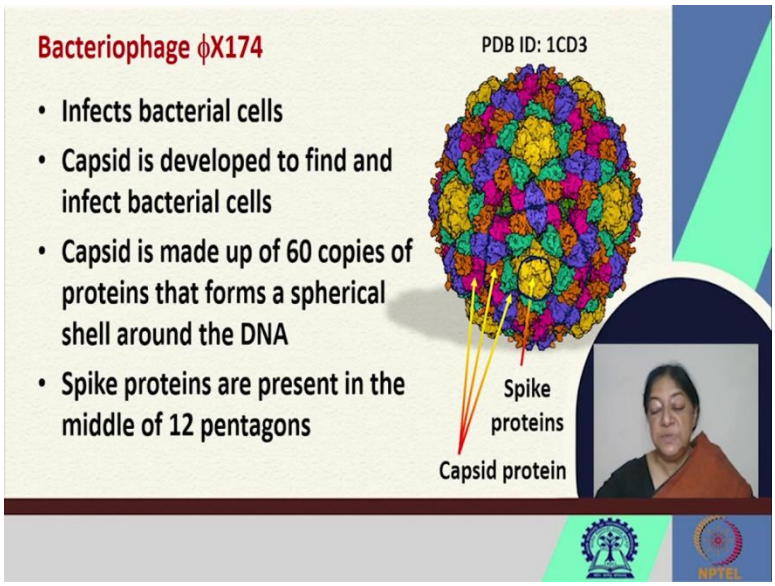
(Refer Slide Time: 07:43)



So we have a top view [refer to slide] of the knob end of the fibers, where we see three of the units here. There is one, two and three of these units. If we look at their interaction now with a specific receptor, we see how they can interact with the receptors.

This is the top view of the knob end of the fiber. Similarly, we can look at its complex with CD46. So when we are looking at these receptor interactions, these are specific types of protein-protein interactions, specific recognition that occurs between the proteins.

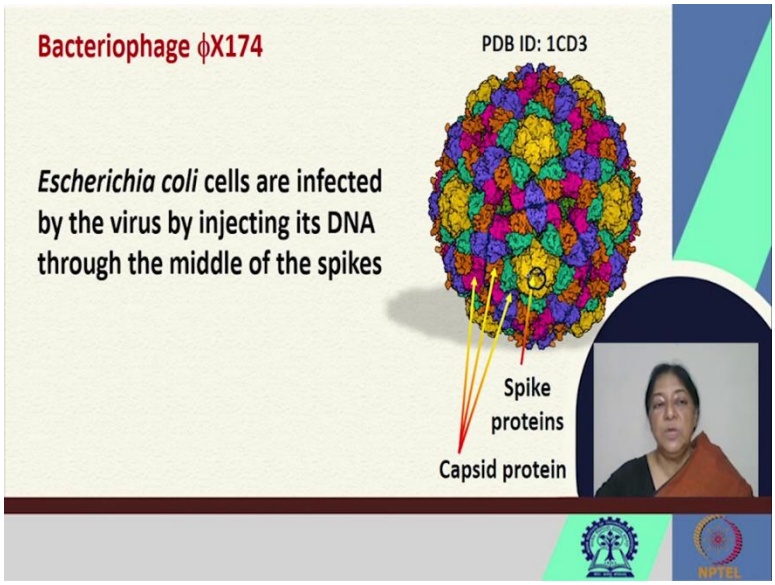
(Refer Slide Time: 08:31)



The bacteriophage is another of these virus, that infects bacterial cells. This capsid is developed to find and infect bacterial cells only. The capsid is made up of 60 copies of proteins, that form a spherical shell around the genome, in this case the DNA. So, these [refer to slide] are the capsid proteins that encompass the DNA present in the cell..

The spike proteins are present in the middle of the 12 pentagons that are present in this capsid. These are the spike proteins marked here and these are the specific capsid proteins.

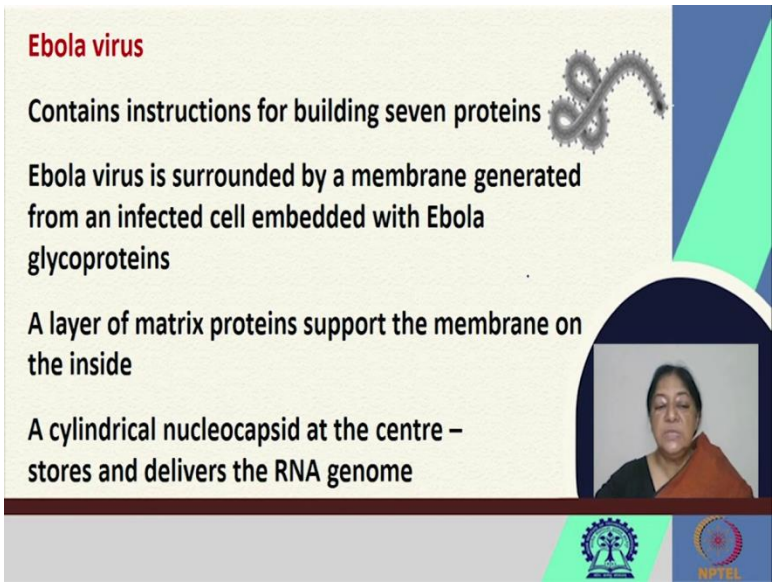
(Refer Slide Time: 09:21)



When we look at the E coli cells, these are infected by the virus, by injecting the DNA through the middle of these spikes. So, the ejection occurs through the middle of the spikes and the DNA

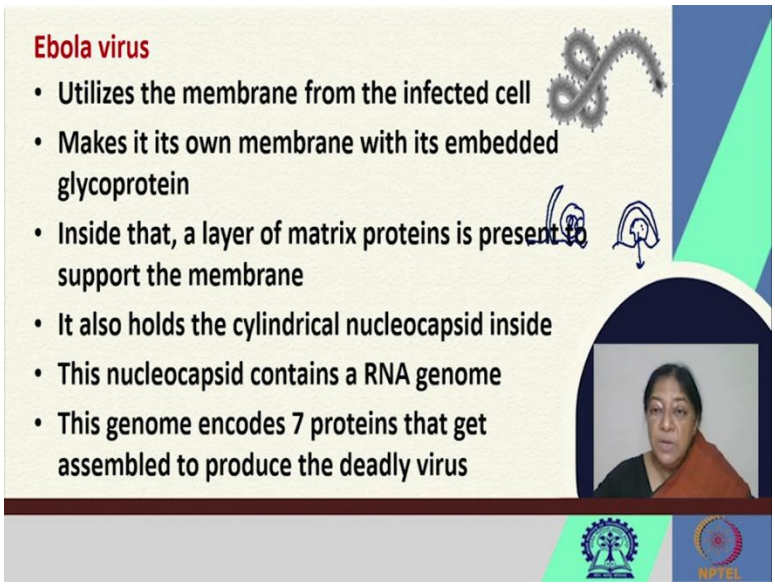
is injected into the host cell, in this case the E coli cell. The machinery of the cell is hijacked so that it can make multiple copies of itself.

(Refer Slide Time: 09:47)



In the Ebola virus, this contains instructions for building seven proteins that are required for its replication. So, the Ebola virus is surrounded by a membrane that is generated from an infected cell embedded with Ebola glycoproteins. In this case there is a layer of matrix proteins, that support the membrane on the inside and there is a cylindrical nucleocapsid at the centre that stores and delivers the RNA genome.

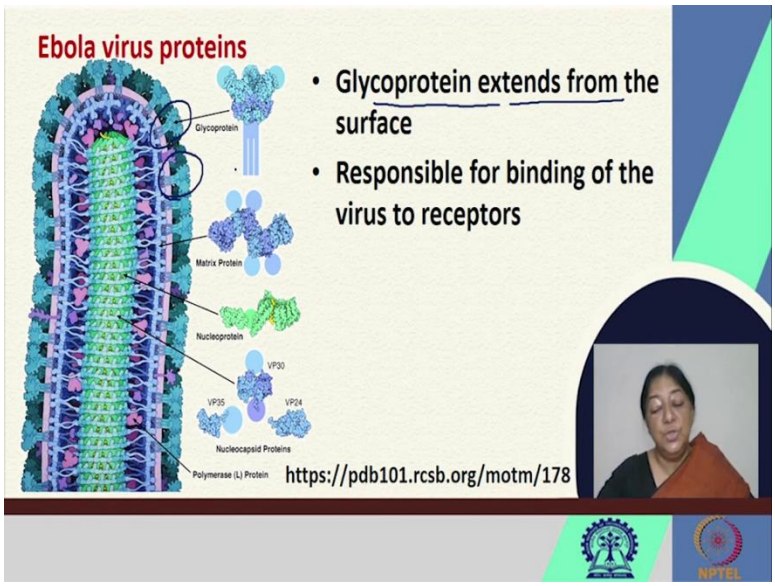
(Refer Slide Time: 10:25)



This utilizes the membrane from the infected cell and makes its own membrane with its embedded glycoprotein. So, in the membrane there are protrusions. There is the infection of the Ebola virus in a form here [refer to slide] and then what happens is, this takes over and there is a structural formation in this fashion and then this enters the membrane.So it makes its own membrane with the embedded glycoprotein.

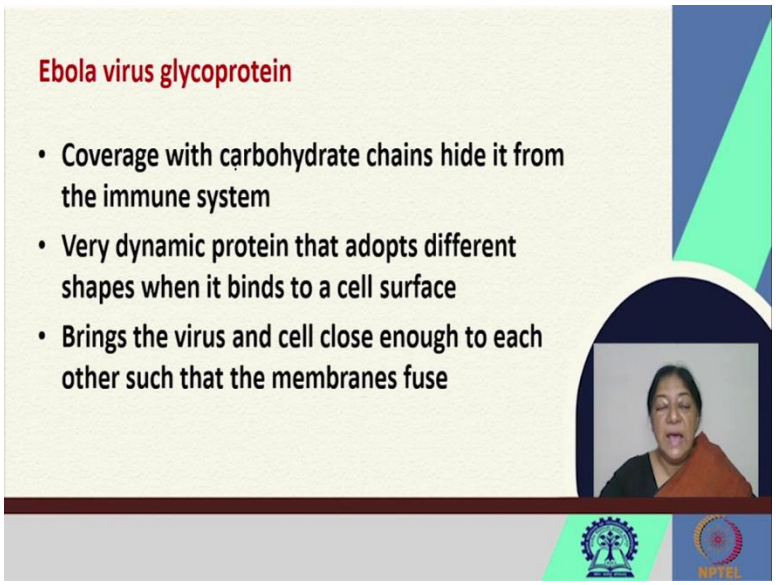
Inside that, a layer of matrix protein is present to support the membrane and it holds the cylindrical nucleocapsid inside and this nucleocapsid contains the RNA genome and the genome can encode for the 7 proteins that then assemble, to produce the deadly virus.

(Refer Slide Time: 11:26)



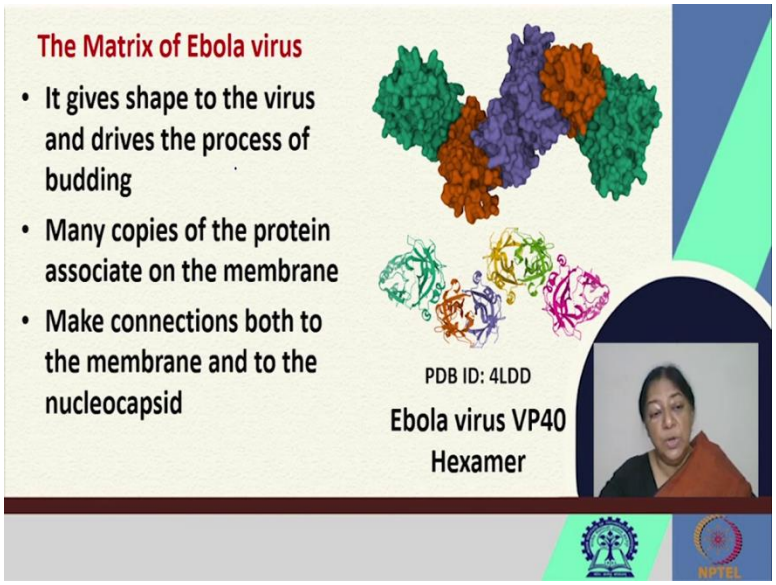
This [refer to slide] is the structure of the Ebola virus, where we can see the membrane. There are glycoproteins that extend from the surface and these are responsible for binding of the virus to the receptors and if we look at the Ebola virus glycoprotein, there is a coverage with the carbohydrate chains that hide it from the immune system.

(Refer Slide Time: 11:46)



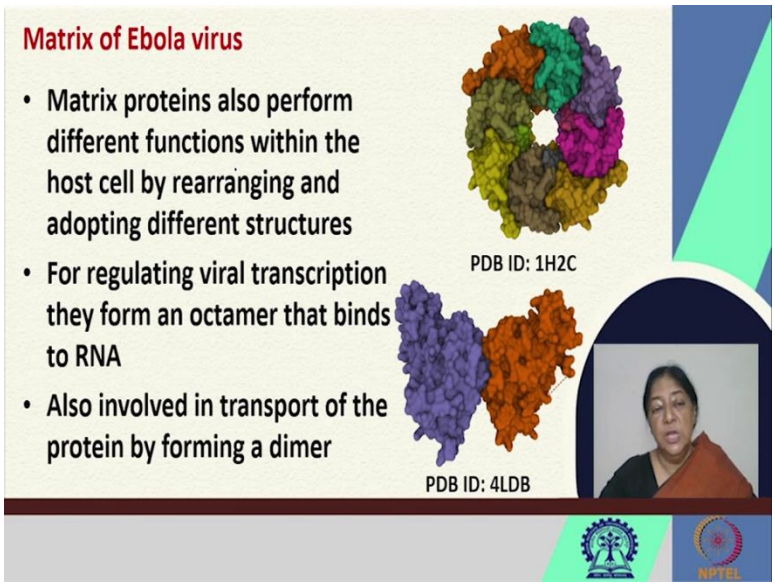
It is a very dynamic protein that can adopt many different shapes when it binds to a cell surface, so that it is not recognized easily. In this case it brings the virus and the cell close enough to each other, such that the membranes fuse together and the virus enters the cell as shown.

(Refer Slide Time: 12:19)



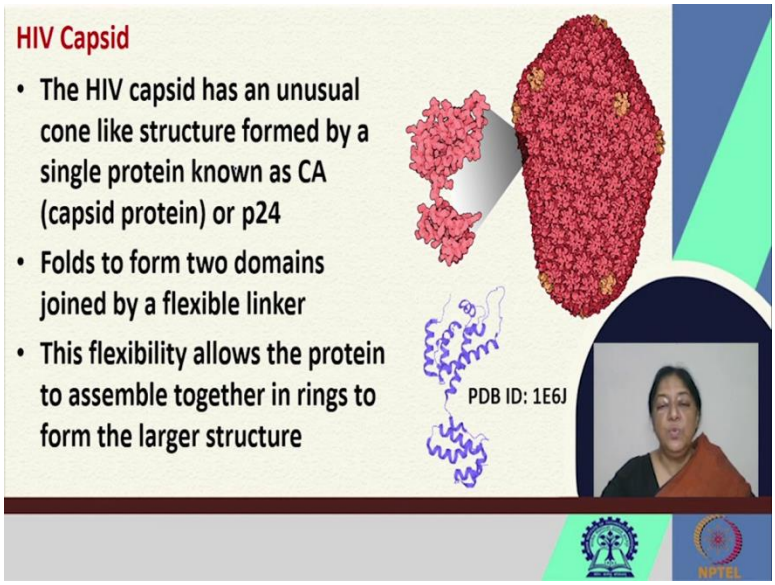
The matrix of the Ebola virus is such, that it can adopt different shapes and this drives the process of what is called budding. Many copies of the protein associate on the membrane and they make connections to the membrane and to the nucleocapsid and the matrix proteins therefore have to perform different functions within the host cell, by rearranging and adopting different structures. For example it needs to regulate viral transcription; for that case it forms an octamer that binds to RNA.

(Refer Slide Time: 12:42)



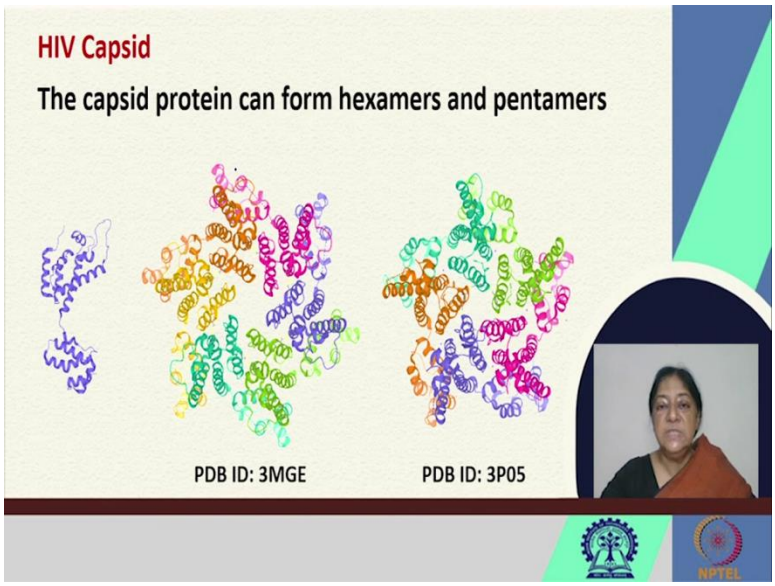
It is also involved in the transport of the protein and in this case it forms a dimer. So the versatility of the viral proteins can be seen here, in the way that they adapt and take up different structures for the different functions that they have to perform.

(Refer Slide Time: 13:34)



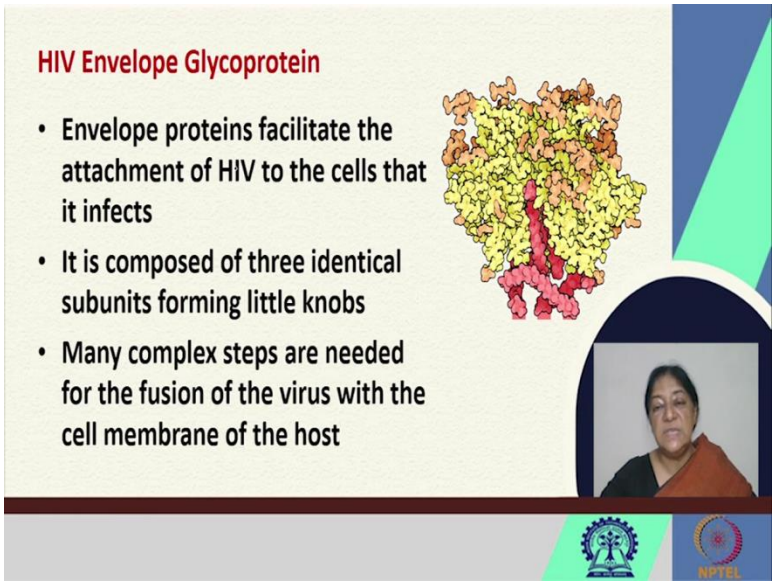
If we look [refer to slide] at the HIV capsid, this HIV capsid has a cone like structure formed by a single protein called the CA or the capsid protein and in this case, this folds to form two domains that are joined by a flexible linker, to accommodate for the packing in the capsid. This flexibility allows the protein to assemble together in rings, that would result in this large structure that forms the HIV capsid.

(Refer Slide Time: 14:01)



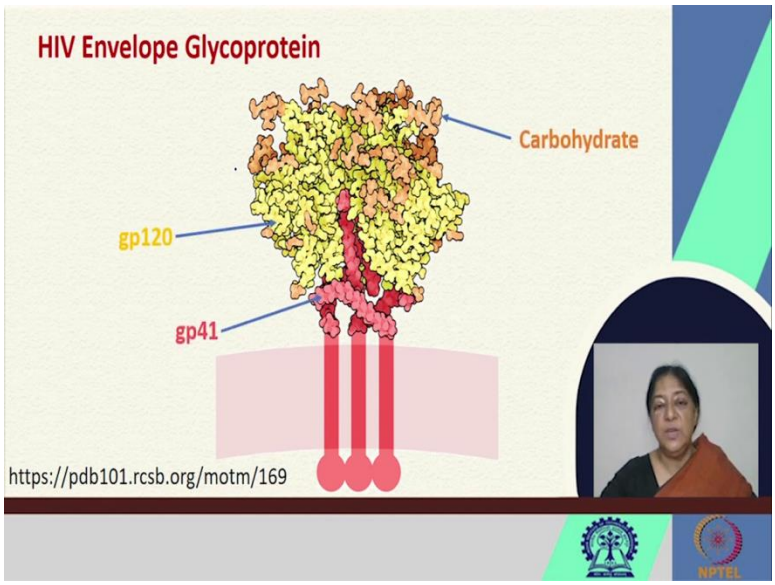
The capsid protein can again adopt different structures. It can form hexamers, it can form pentamers; the assembly is such, these then stack up to form the cone structure.

(Refer Slide Time: 14:13)



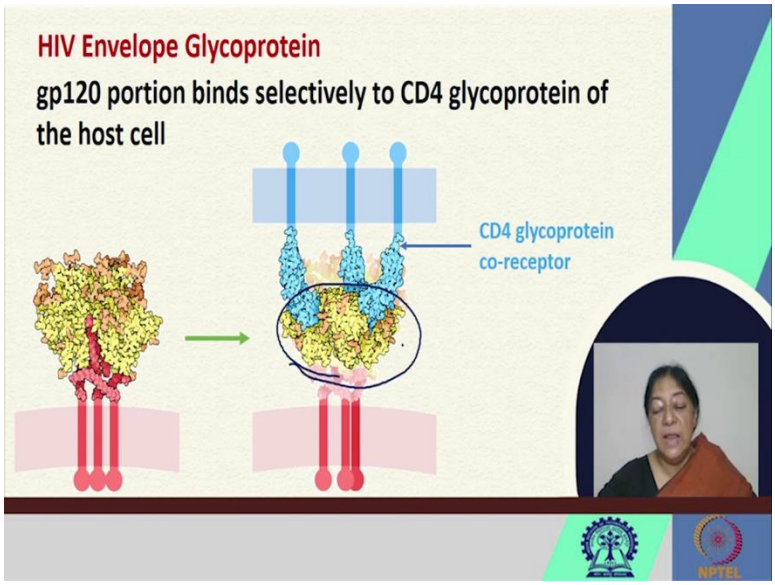
The HIV envelope glycoprotein, facilitates the attachment of HIV to the cell that it infects. This is composed of three identical subunits forming little knobs on the surface and there are several steps that are required for the fusion of the virus with the cell membrane of the host. The fusion of this is required for the cell to be infected with the virus.

(Refer Slide Time: 14:48)



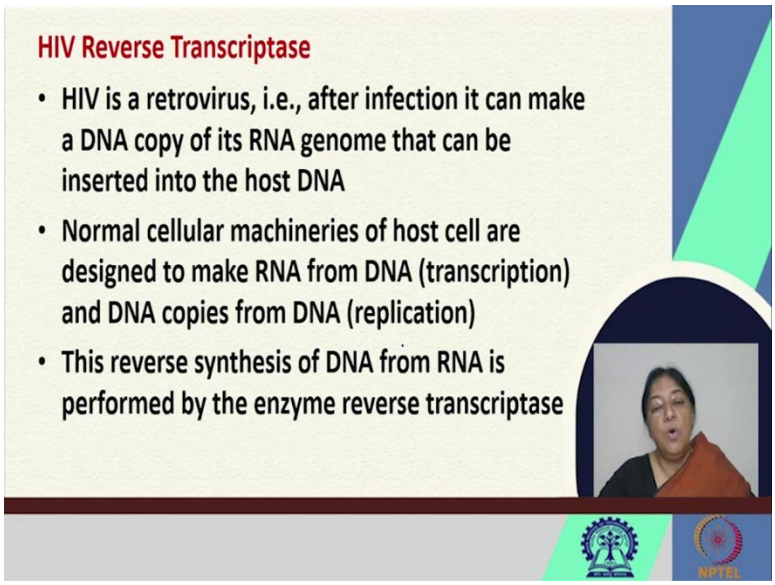
The HIV envelope glycoprotein therefore has these different glycoproteins, the carbohydrates associated with them and the gp120 portion, binds selectively to a CD4 glycoprotein of the host cell and the CD4 glycoprotein that is the core receptor, binds to this specific glycoprotein.

(Refer Slide Time: 14:58)



So, it is this interaction that allows the fusion and the infection to occur.

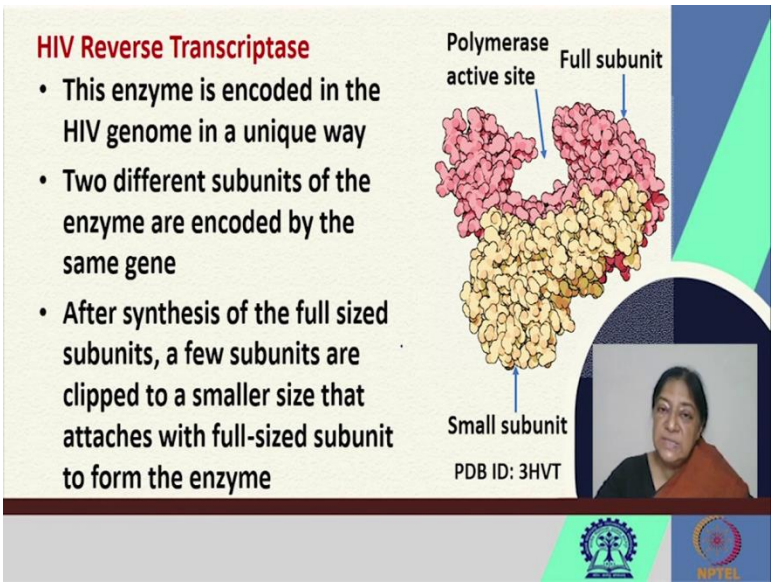
(Refer Slide Time: 15:18)



In HIV reverse transcriptase is a retrovirus. That is, after infection what it can do is, it can make a DNA copy of its RNA genome that can be inserted into the host DNA. This is an RNA genome that is present, but being a retrovirus it makes a DNA copy of itself. So the normal cellular machineries of host cells, are designed to make DNA from RNA as we know from the central dogma of biology; DNA to RNA to protein.

It is designed to make RNA from the DNA which is the transcription process, but here it makes DNA copies from the DNA in a replication process and the reverse synthesis of the DNA from the RNA, is performed by a specific enzyme called the reverse transcriptase.

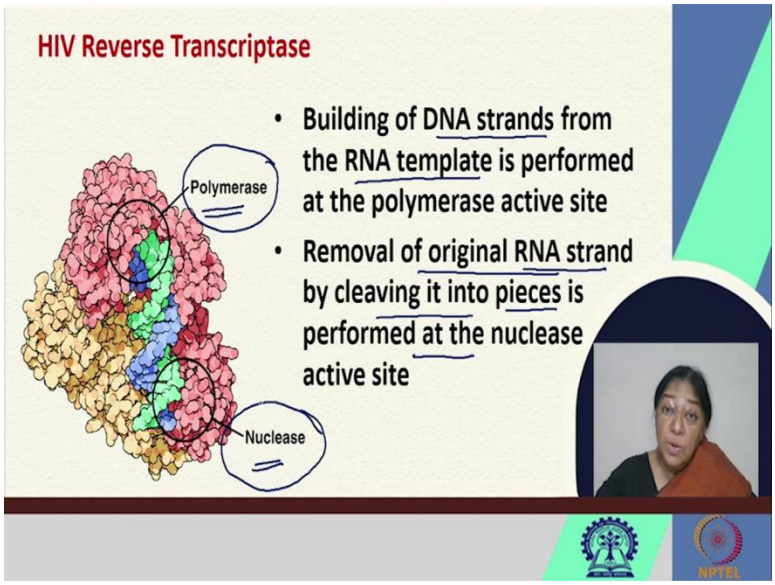
(Refer Slide Time: 16:21)



In this case there is a specific polymerase active site. This enzyme is encoded in a very unique fashion. There are two different subunits of the enzyme that are encoded by the same gene. So

there is a full subunit that has a nucleus site and there is one that has the polymerase active site. Now after synthesis of the full size subunits, a few subunits are clipped to a smaller size that then attaches with the full-size subunit, to form the enzyme.

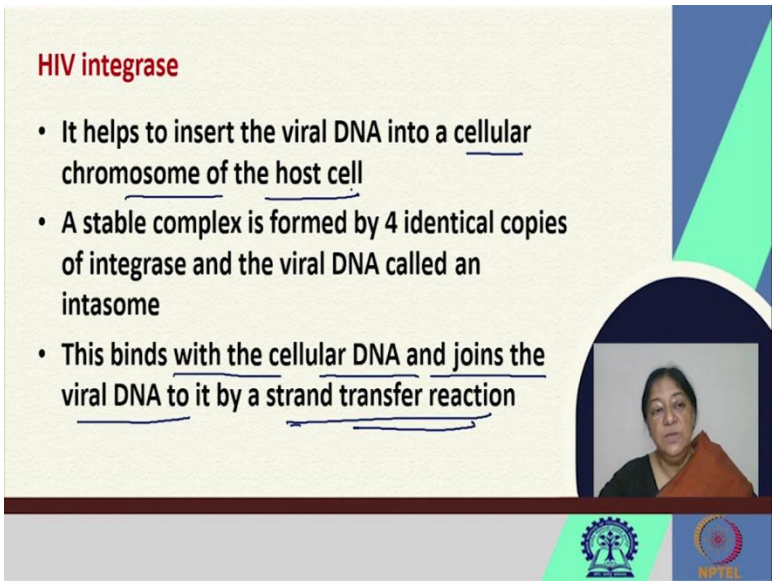
(Refer Slide Time: 17:09)



So, there is a polymerase site and there is a nuclease site. The building of the RNA strands from the RNA template, is formed at the polymerase active site. So this is where we have the building of the DNA strands from the RNA template, that is performed at the polymerase site.

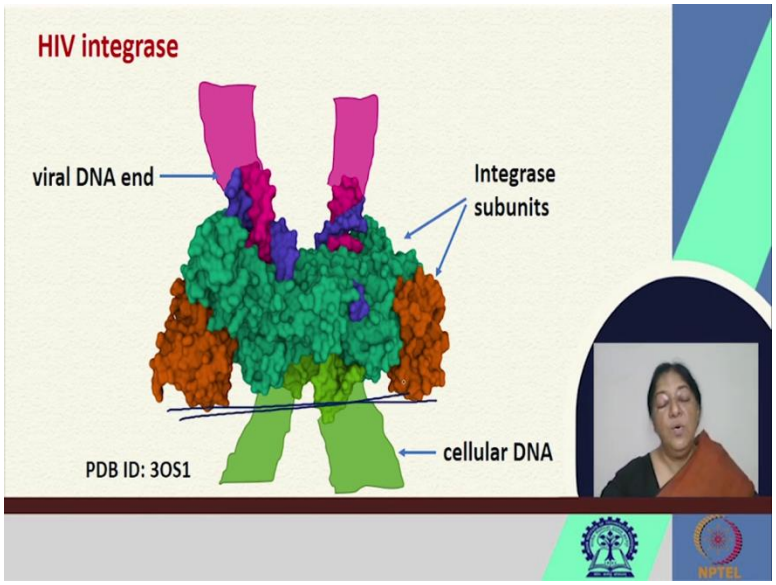
However the removal of the origin RNA strand, implies that there has to be a nuclease presence. So, this [refer to slide] is then cleaved into small pieces and this is performed at the nuclease active site. So, there are two functions occurring here. One is the building up the DNA strands from the RNA template and the other is the destruction of the original RNA strand, by cleavage at the nucleus active site.

(Refer Slide Time: 18:07)



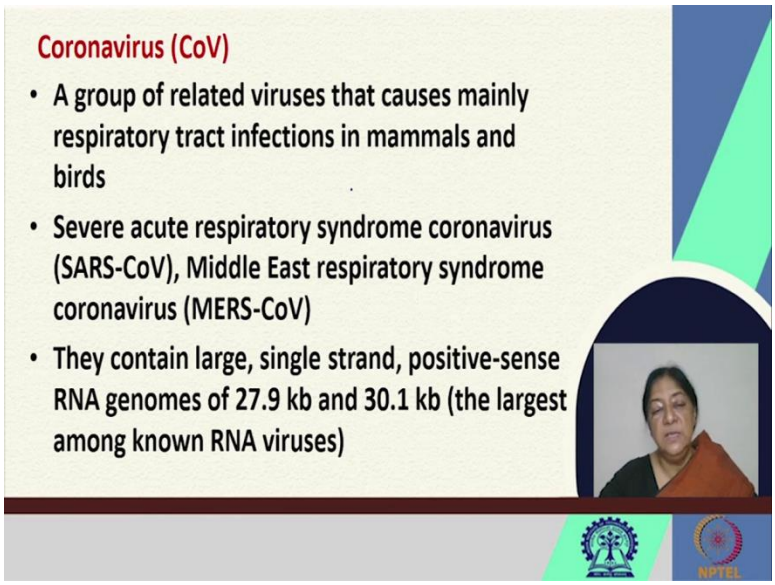
In HIV integrase, this helps to insert the viral DNA into a cellular chromosome of the host cell. Then a stable complex is formed by 4 identical copies of the integrase and the viral DNA. This complex is called an intasome. This binds with the cellular DNA and joins the viral DNA to it by a strand transfer reaction, taking over the host cell machinery.

(Refer Slide Time: 18:39)



So this [refer to slide] is how the structure of the HIV integrase looks like. There is a viral DNA end and a cellular DNA and these integrase subunits will allow the binding of the cellular DNA, that joins the viral DNA by a specific strand transfer reaction, that will be put into the host cell and then this insertion into the cellular chromosome of the host cell, the machinery take over in that cell to produce more copies of it.

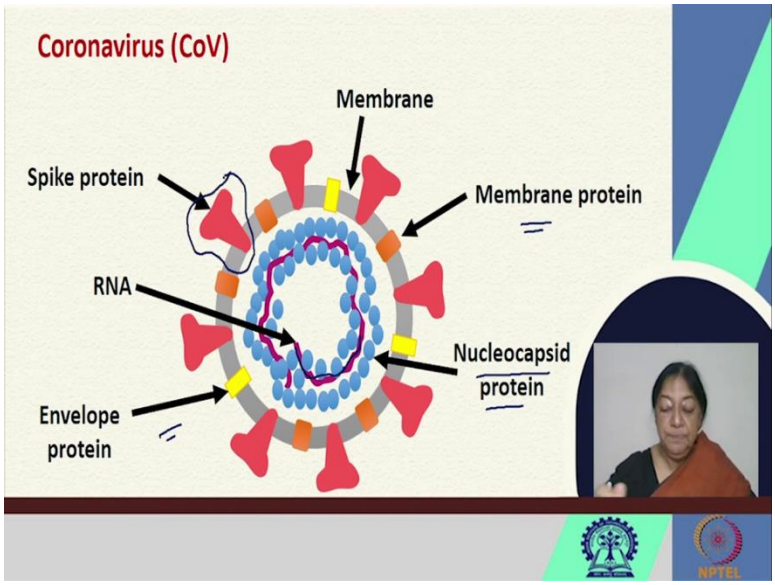
(Refer Slide Time: 19:15)



The coronavirus, is one that has affected all of us in some way or the other. This is a group of related viruses that cause many respiratory tract infections in mammals and in birds. The diseases known are the severe acute respiratory syndrome coronavirus the SARS-CoV, the SARS-CoV-2 that we have now as well.

The Middle East respiratory syndrome coronavirus, the MERS-CoV; they contain very large single strand positive sense RNA genomes of the order of 27.9 and 30.1 kb, which are the largest among the known RNA viruses.

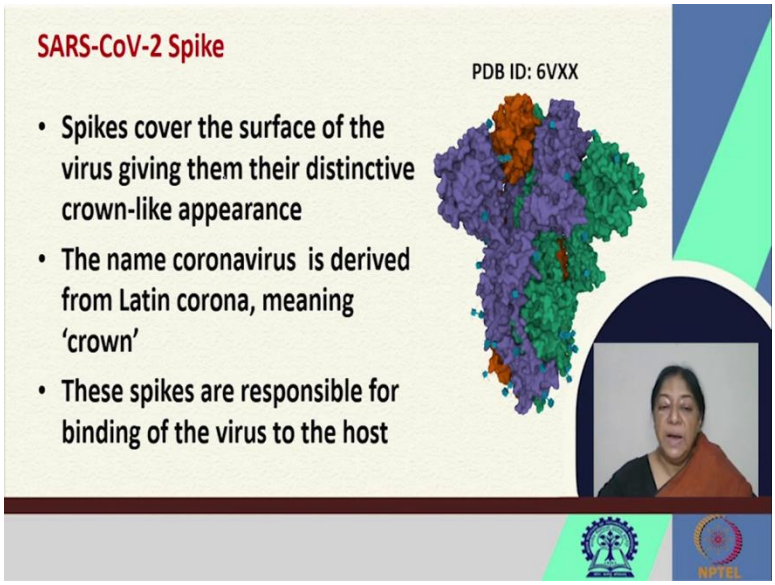
(Refer Slide Time: 20:07)



This [refer to slide] coronavirus, a diagram or a figure that we have been seeing very commonly now, has in here its RNA. There is a nucleocapsid protein that surrounds the RNA. It has a

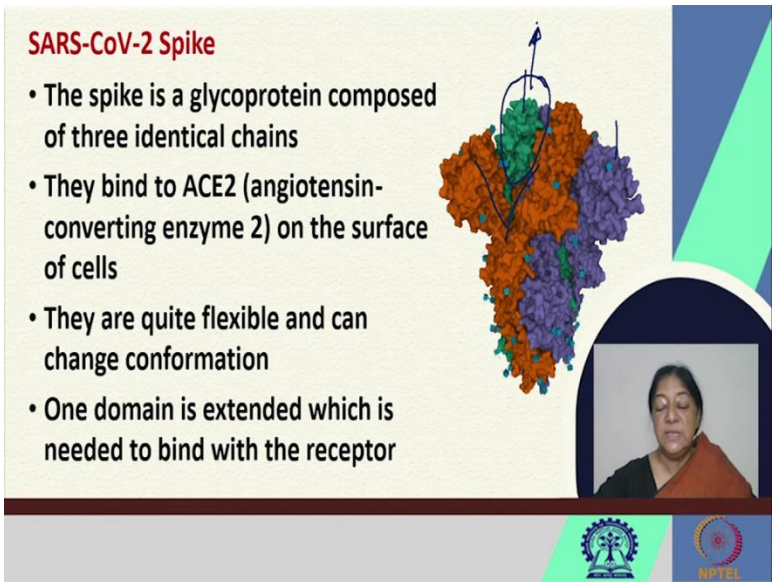
membrane. In this membrane there are envelope proteins, membrane proteins and embedded in them are the spike proteins. This gives it the specific structure of the coronavirus.

(Refer Slide Time: 20:39)



The spikes cover the surface of the virus, that give them their distinctive crown like appearance. Hence the name corona which comes from the Latin corona, which means crown and these spikes are responsible for binding of the virus to the host.

(Refer Slide Time: 21:04)

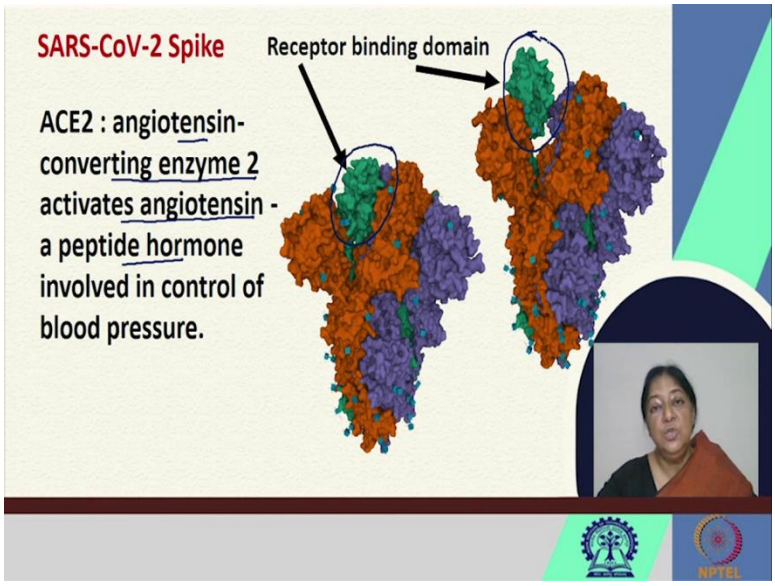


What happens is, the spike is a glycoprotein that is composed of three identical chains. We see [refer to slide] the three identical chains, one chain second chain and the third chain. These three

identical chains bind to ACE2, the angiotensin-converting enzyme 2 which we had discussed earlier, that is present on the surface of the cells.

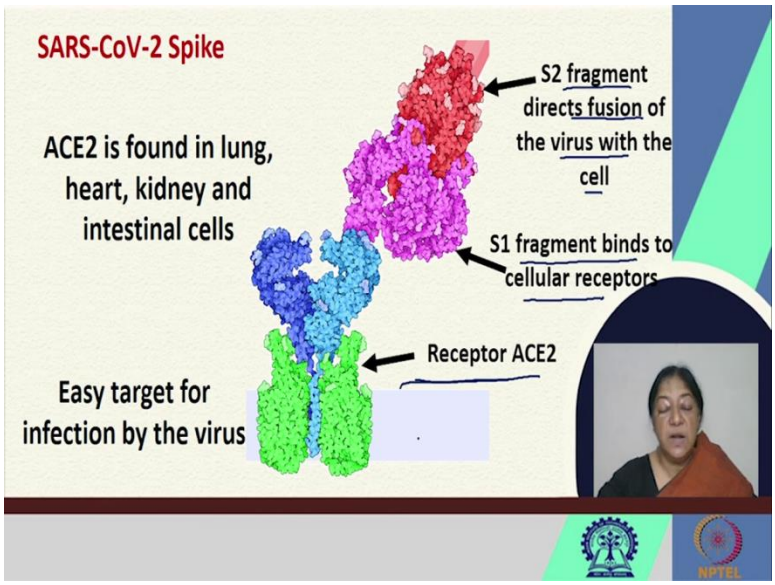
These are quite flexible and can change conformation and what is required is one domain is extended, which is needed to bind with the receptor. So, say we have a specific domain that is then ejected or extended to bind with the receptor.

(Refer Slide Time: 21:53)



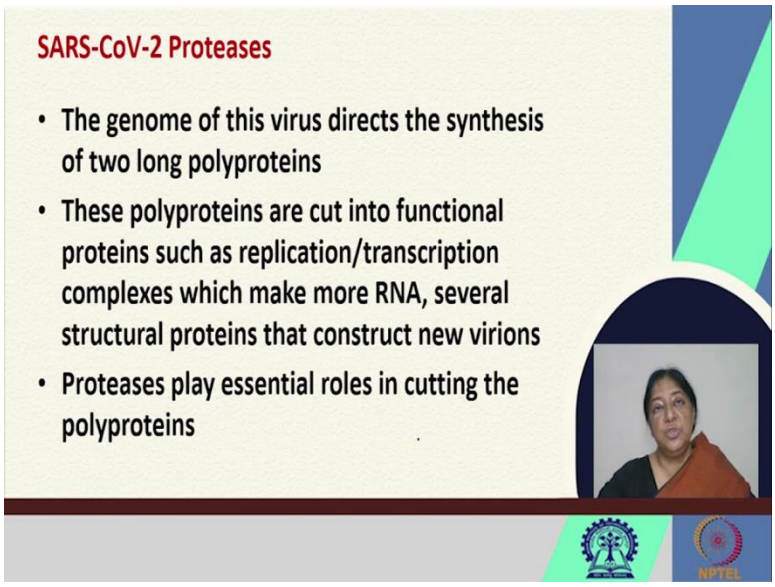
We have this [refer to slide] specific receptor domain that extends itself for a possible binding with the receptor. So the ACE2, the angiotensin-converting enzyme 2, activates angiotensin as we know, a peptide hormone that is involved in the control of blood pressure. So the receptor binding domain present here, is extended in a manner to facilitate the binding.

(Refer Slide Time: 22:30)



ACE2 is found in lung, heart, kidney and intestinal cells. As a result of which it becomes an easy target for infection by the virus. So we [refer to slide] have the receptor, we have the fragment that binds to the cellular receptor. The S2 fragment directs fusion of the virus with the cell, the S1 fragment binds to the cellular receptor and ACE2 is the receptor present on the cell membrane, providing an easy target for infection by the virus.

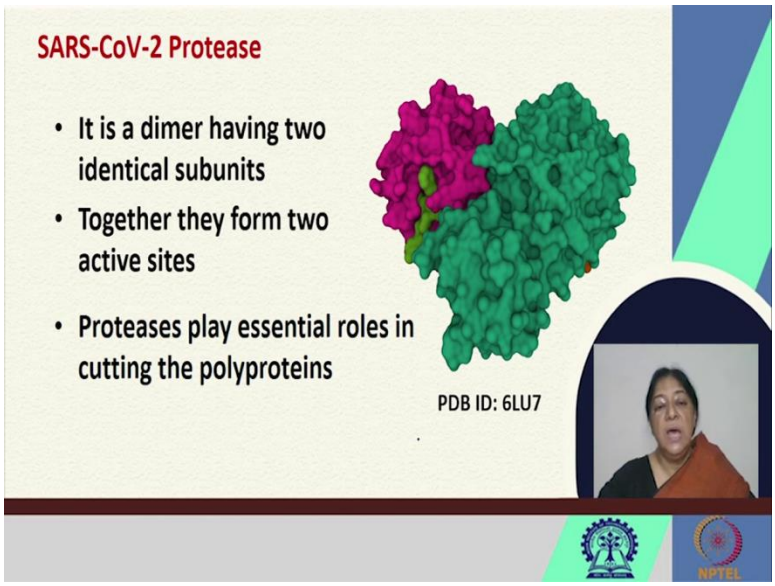
(Refer Slide Time: 23:04)



The SARS-CoV-2 proteases, this genome of the virus directs the synthesis of two long polyproteins. These polyproteins are cut into functional proteins such as a replication transcription complex, which then make more RNA and other several structural proteins that can construct new virions.

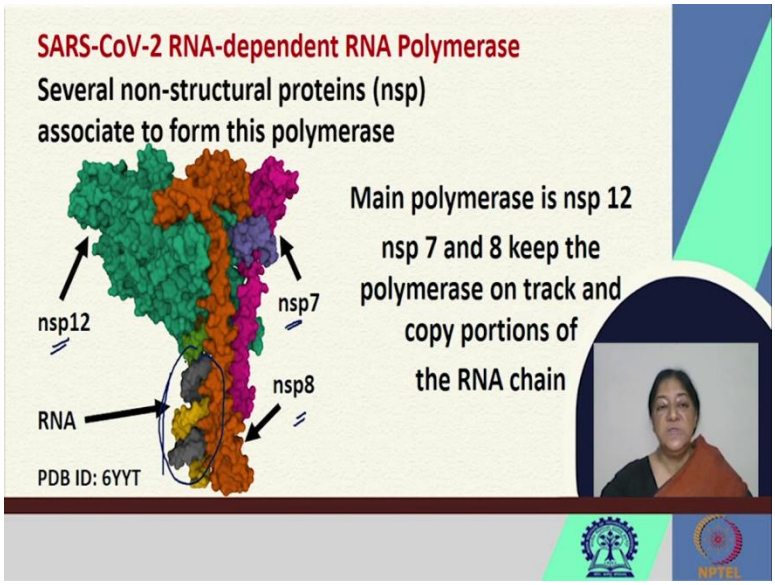
These proteases play very essential roles in cutting the polyproteins, so that their functional proteins can be created. The complexes required for the generation of more RNA can be created.

(Refer Slide Time: 23:50)



The structure of this SARS-CoV-2 protease is a dimer that has two identical subunits and together they form two active sites and these have essential roles in cutting the polyproteins.

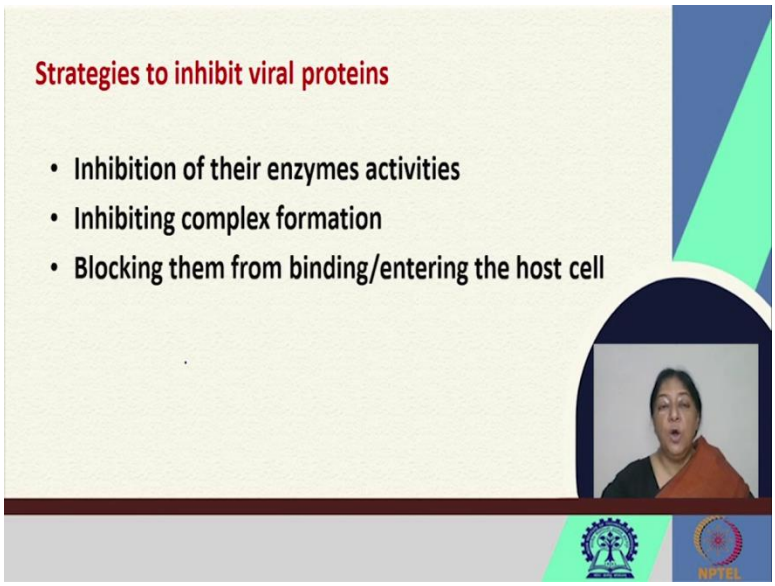
(Refer Slide Time: 24:09)



The SARS-CoV-2 RNA-dependent RNA polymerase, has [refer to slide] what are called non-structural proteins. So we have the RNA associated with this. We have what are called the nsp; the non-structural proteins 7, 8 and 12 in this case, where they associate to form this polymerase.

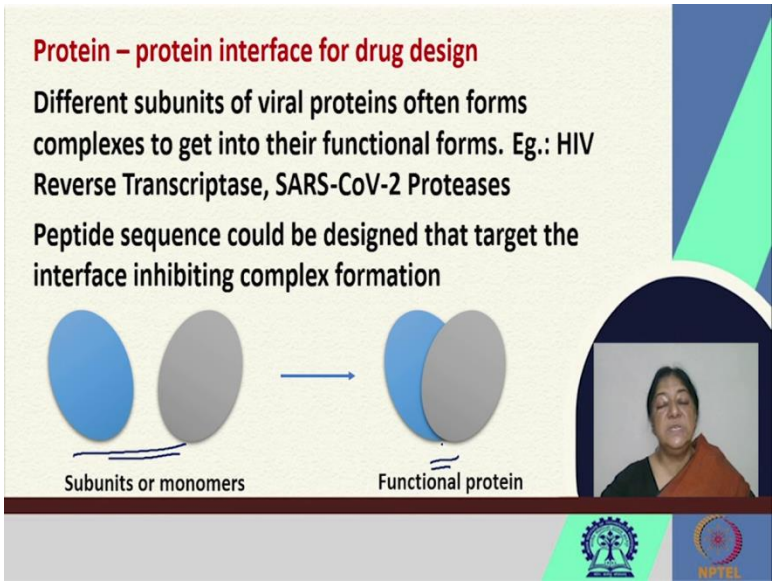
It is an RNA-dependent RNA polymerase and the main polymerase is nsp12 and nsp7 and 8 keep the polymerase on track and copy portions of the RNA chain to facilitate this.

(Refer Slide Time: 24:49)



This means that, we now need to design or think of strategies to inhibit viral proteins. We have seen the way they act and the way they function. So there may be a possibility to inhibit their enzyme activities, inhibit the complex formation that is necessary for the action to occur of the virus or block them entirely from binding or entering the host cell.

(Refer Slide Time: 25:21)

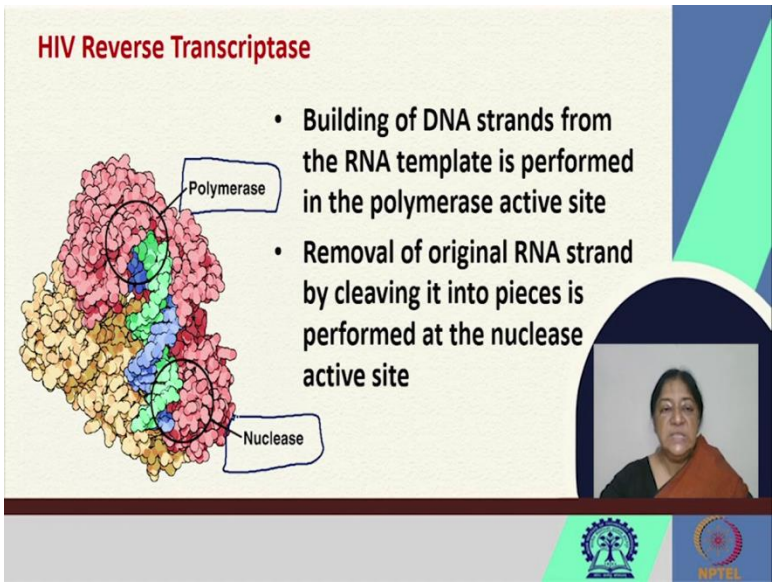


This is where we can have a protein-protein interface for drug design, where we could have different subunits of the viral proteins that form the complexes to get to their functional forms.

So we have the different subunits or the monomeric units, that are required to form the functional protein.

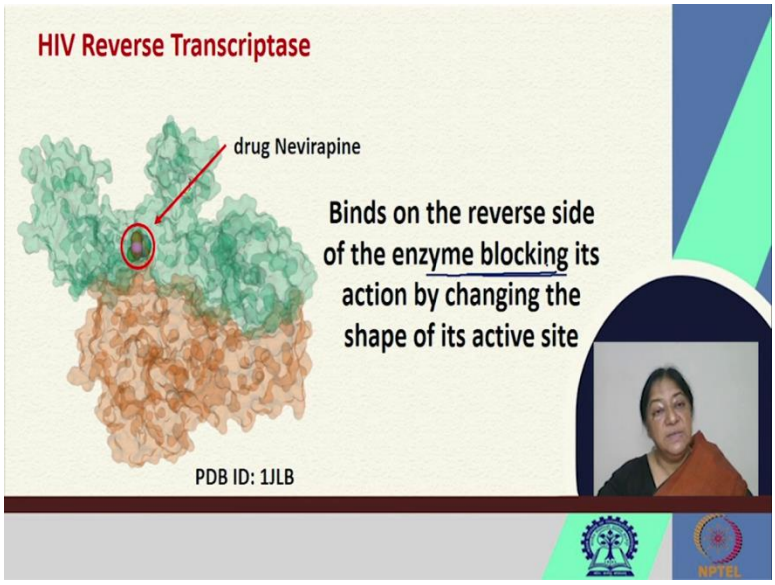
There is an interface, a specific interface and it could so happen that peptide sequences or specific drugs could be designed that target the interface, thus inhibiting the complex formation.

(Refer Slide Time: 25:57)



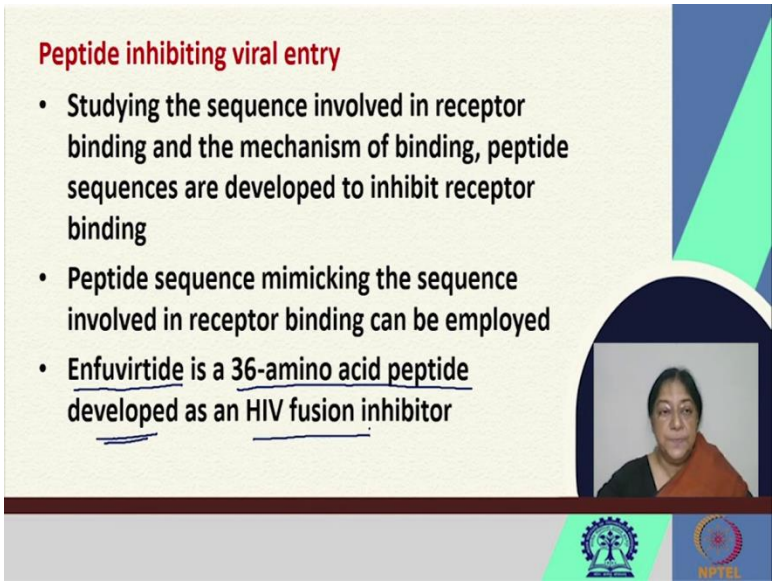
For example, we saw in HIV reverse transcriptase there were two functional units. One was the polymerase and one was the nuclease. The building of the DNA strands from the RNA template was performed at the polymerase active site and the cleaving of the original RNA strand was performed at the nuclease site.

(Refer Slide Time: 26:22)



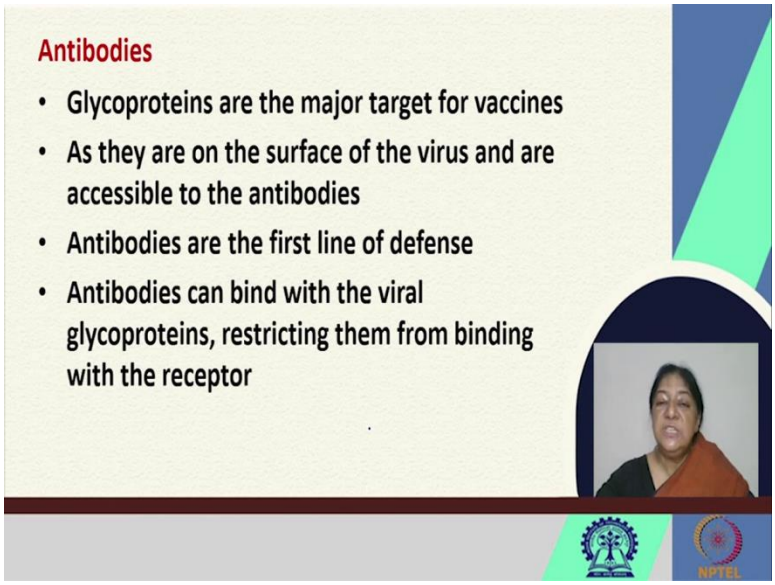
This means that we can develop drugs that can block the enzyme from working either the polymerase activity or the nucleus activity. This particular drug binds on the reverse side of the enzyme and it blocks its action, by changing the shape of the active site so that it can no longer bind the substrate.

(Refer Slide Time: 26:48)



So, we can have peptides that can inhibit viral entry. The sequence involved in receptor binding and the mechanism of binding, peptide sequences can be developed to inhibit the receptor binding. The peptide sequence would be required to mimic the sequence that is actually involved in receptor binding and then, this can be used to stop viral entry. For example, enfuvirtide is a 36-amino acid peptide, that has been developed as an HIV fusion inhibitor.

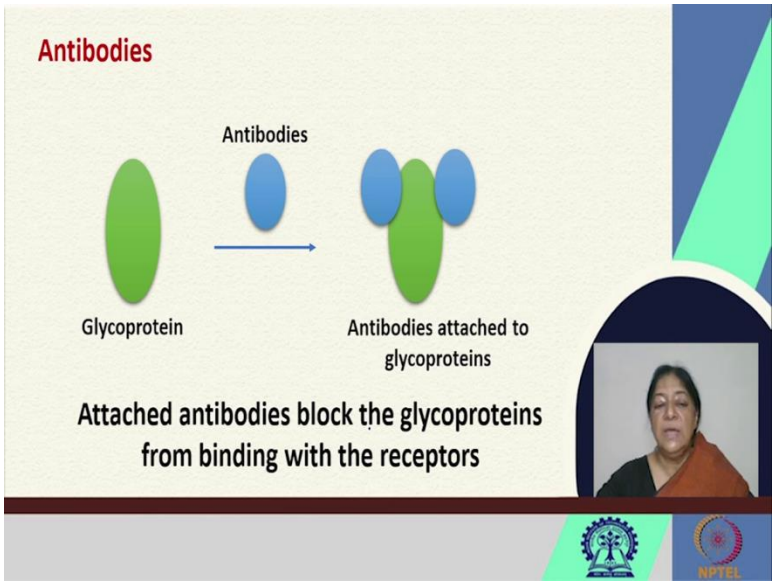
(Refer Slide Time: 27:28)



Antibodies can also work in this fashion because glycoproteins are the major target for vaccines. We saw how the glycoproteins interacted with each other to form the complex or to facilitate the complex formation and then be attached or then attack the host cell. So because they are on the surface of the virus, they are accessible to the antibodies.

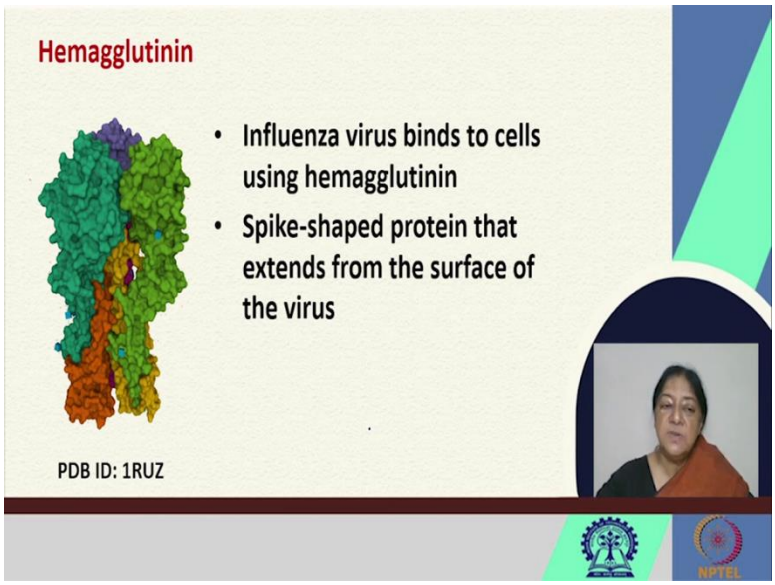
The antibodies become the first line of defense and they can bind with the viral glycoproteins, restricting them from binding with the receptor glycoproteins or in the membrane fusion that occurs.

(Refer Slide Time: 28:08)



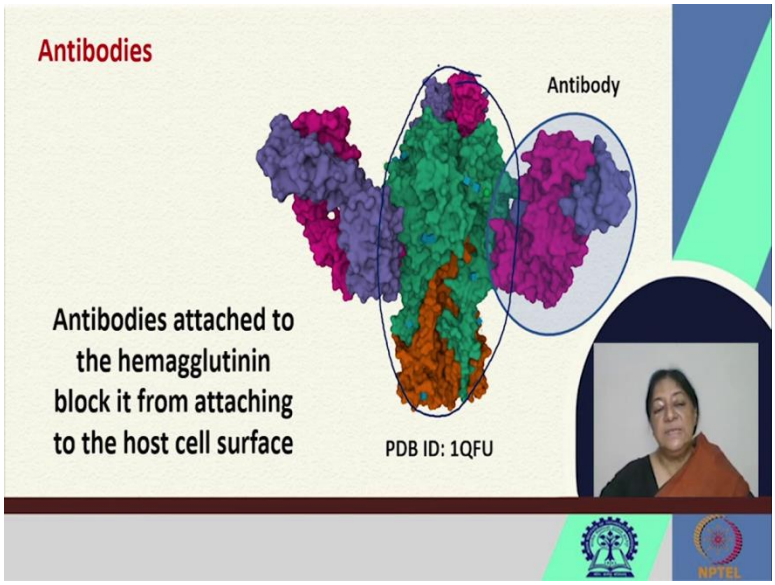
So if we have a glycoprotein and the antibodies associated with them, they can be specific antibodies that are attached to the glycoproteins that can block the glycoproteins from binding with the receptors.

(Refer Slide Time: 28:22)



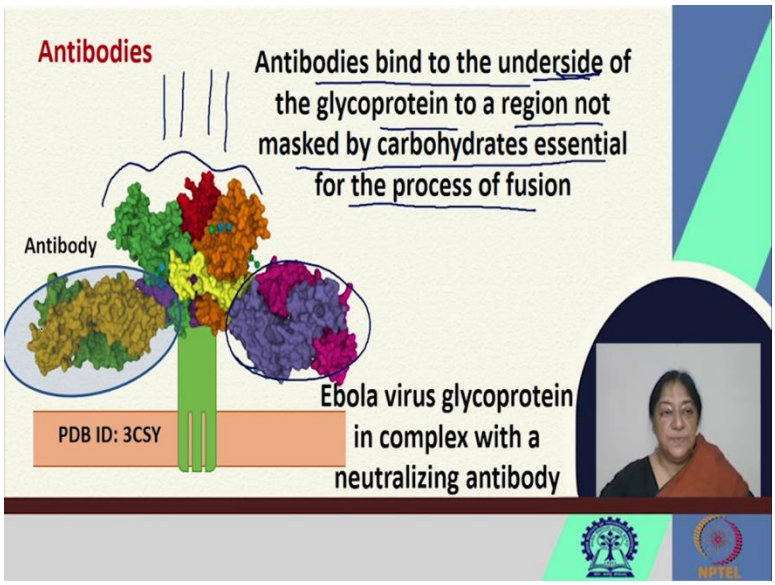
Hemagglutinin is one such possibility where the influenza virus binds to cells, using hemagglutinin. The spike shape protein that extends from the surface of the virus, can have antibodies attached to it.

(Refer Slide Time: 28:35)



So, this [refer to slide] is where we have the hemagglutinin and we have the antibodies that are attached to it. The antibodies attached to the hemagglutinin, block it from attaching to the host cell surface.

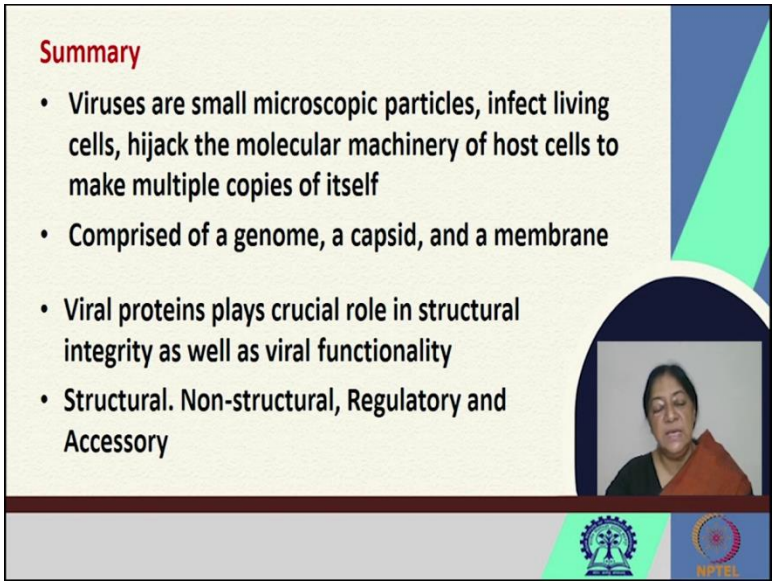
(Refer Slide Time: 28:53)



If we look [refer to slide] at the general design of these, the Ebola virus glycoprotein is shown here in complex with a neutralizing antibody. The antibodies bind to the underside of the glycoprotein, to a region that is not masked by the carbohydrates, that are essential for the process of fusion.

Here we have the glycoproteins or the carbohydrates attached to the specific membrane protein, that are going to be involved in the attachment with the virus. But the antibodies present here are there to the underside of the glycoprotein, so they do not mask the carbohydrate.

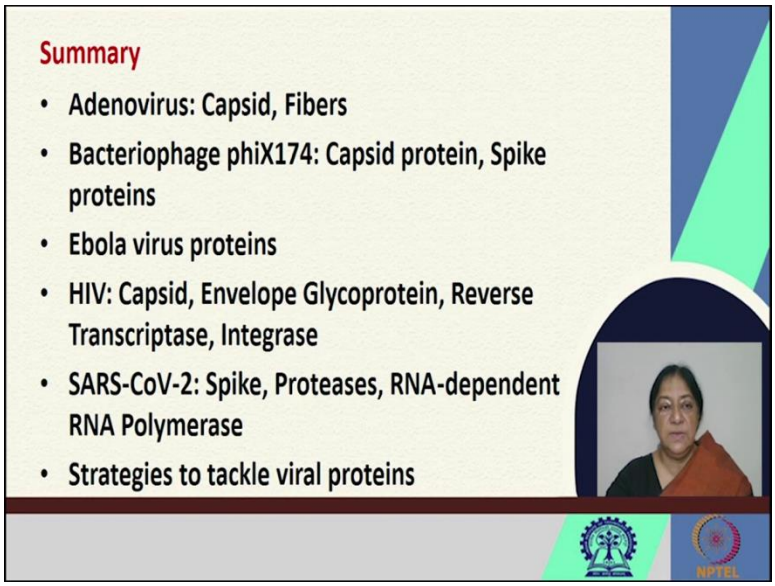
(Refer Slide Time: 29:47)



So in summary, we looked at viruses that are small microscopic particles, that infect living cells, hijack the molecular machinery of host cells to make multiple copies of itself. It is comprised of a genome, a capsid and a membrane. The viral proteins that form these capsids play crucial roles

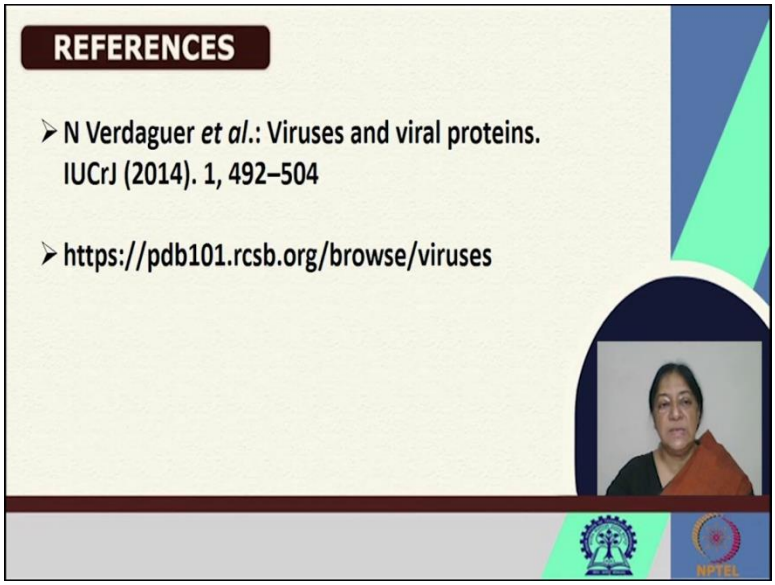
in structural integrity, as well as viral functionality and these proteins can be structural, non-structural, regulatory and accessory.

(Refer Slide Time: 30:20)



We looked at different types of viruses; the adenovirus, the bacteriophage, the Ebola type, HIV and the SARS-CoV, where we saw the role these proteins can play in the assembly, to protect the genome that is then transferred to the host cell to build up more copies of itself. We looked at more strategies to tackle viral proteins and how viral diseases may be combated.

(Refer Slide Time: 30:55)



These [refer to slide] are the references.

Thank you.