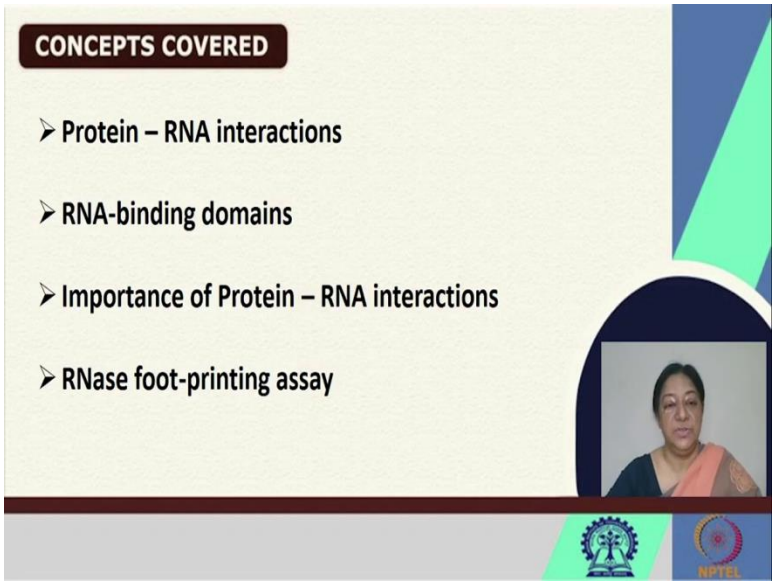
Fundamentals of Protein Chemistry
Prof. Swagata Dasgupta
Department of Chemistry
Indian Institute of Technology, Kharagpur

Module - 10
Protein Macromolecule Interactions I
Lecture - 50
Protein Nucleic Acid Interactions - III

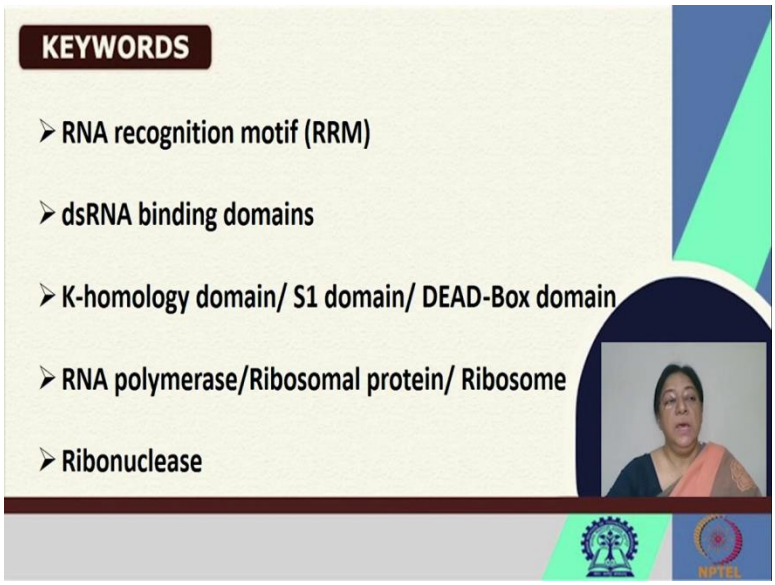
(Refer Slide Time: 00:41)



In the last lecture of protein nucleic acid interactions of our module on protein macromolecular interactions, we will be looking at the interactions of proteins with RNA. We have studied the DNA-RNA structure, DNA domains and DNA interactions.

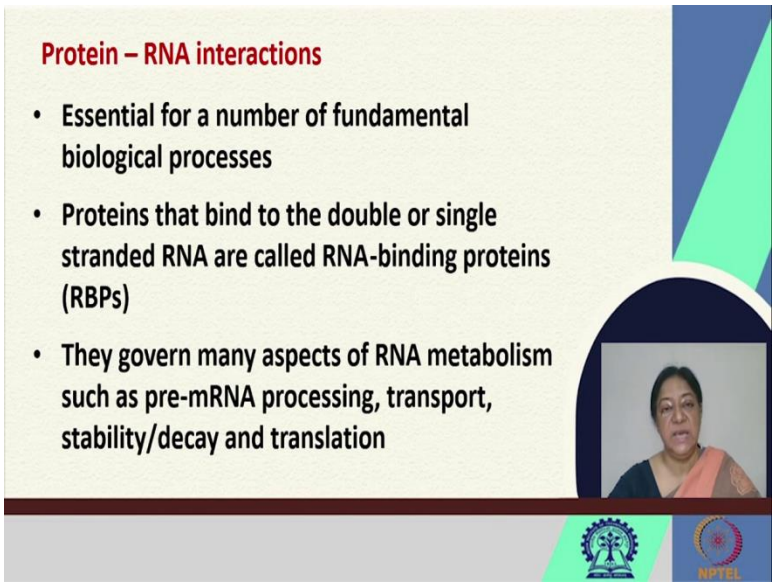
In this lecture we will be looking at protein RNA interactions, RNA binding domains, the importance of protein RNA interactions and the last of the detection methods that we mentioned, the RNase foot-printing assay.

(Refer Slide Time: 00:58)



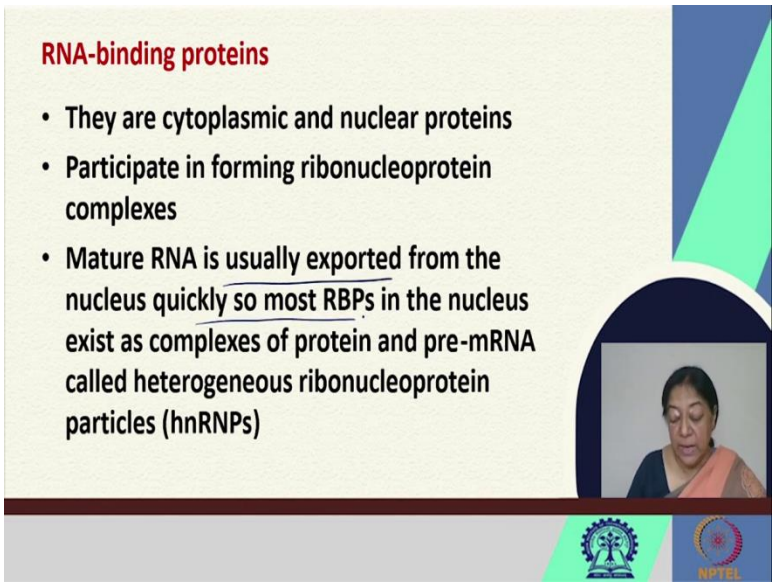
In the RNA recognition motif, which is important in the interaction process, we will be looking at the several domains that are known; the double stranded RNA domain, the K-homology domain, the S1 and the DEAD-box domains and then we will look at specific proteins; RNA polymerase, the ribosomal proteins and ribonuclease. We had looked at the polymerase proteins when we discussed our lectures on motor proteins.

(Refer Slide Time: 01:35)



When we look at protein RNA interactions, we understand that they are essential for a number of fundamental biological processes because we know that we have our central dogma of biology, where we go from DNA to RNA to protein. So RNA has an extremely important role to play in this specific process. Proteins that can bind to the double or the single stranded RNA are called RNA binding proteins or abbreviated as RBPs. What they do is, they govern many aspects of RNA metabolism such as pre-mRNA processing, transport, stability, decay and translation.

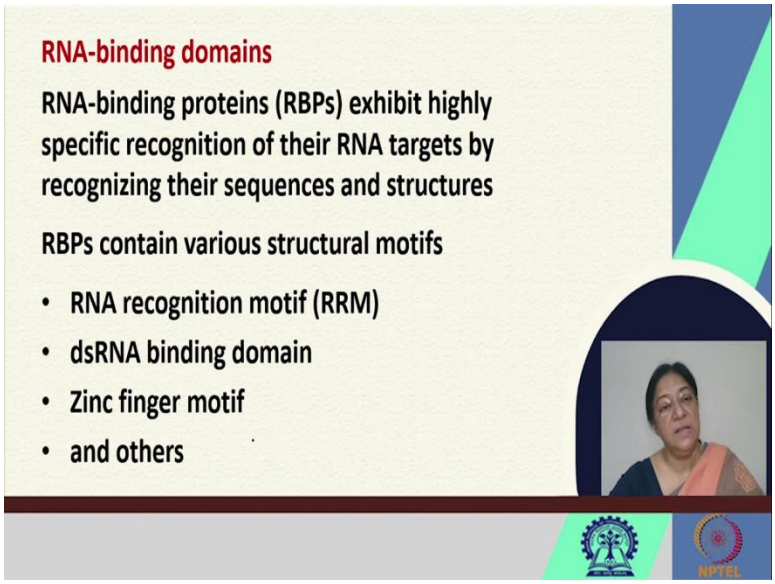
(Refer Slide Time: 02:21)



The RNA binding proteins are cytoplasmic and nuclear proteins because we realize that their importance in the specific processes that they are involved in. They have to be present at many places in the cell.

They also participate in the forming of ribonucleoprotein complexes and mature RNA is usually exported from the nucleus very quickly, so that the RNA binding proteins in the nucleus that exist as complexes of the protein and pre-mRNA. They are called heterogeneous ribonucleoprotein particles or hnRNPs. So this mature RNA that is present now, has to be exported and the process involves the binding with several types of proteins.

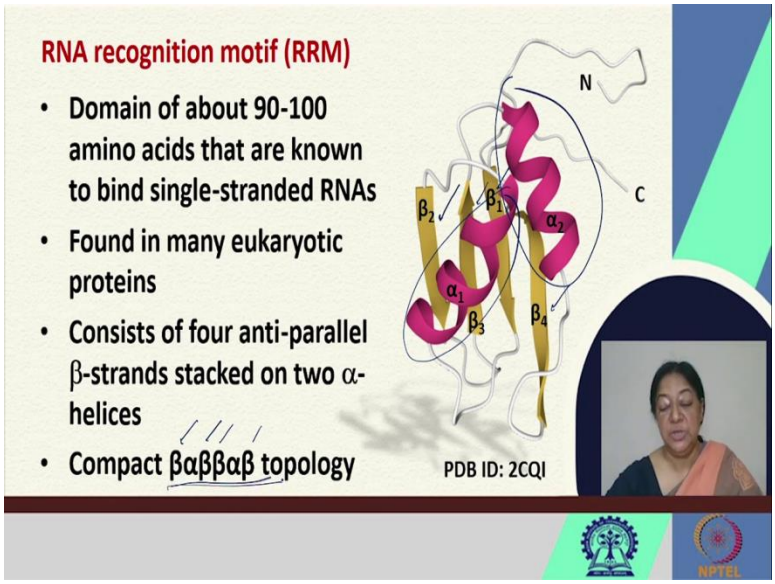
(Refer Slide Time: 03:15)



In the RNA binding domains, we have the RBPs. They exhibit highly specific recognition of the RNA targets, by recognizing the sequence and structures. When we consider the RNA structure, we looked at a pentose sugar the d-ribose sugar, to which we had the phosphate and we had the nitrogenous bases attached.

So, the recognition of a specific sequence by the RNA binding proteins, is important for any activity to occur in the specific biochemical processes that are important. The RBPs therefore, contain several structural motifs. There is an RNA recognition motif, there is the double stranded RNA binding domain and a zinc finger motif and several others. We looked at the zinc finger motif in our previous lecture.

(Refer Slide Time: 04:18)

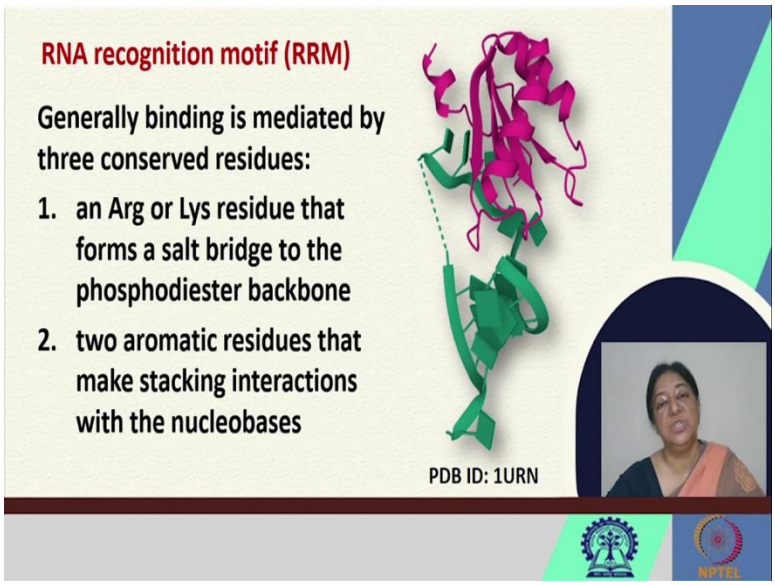


When we look at the RNA recognition motif of the RRM, there are specific characteristics of the types of residues involved and more so, structural importance based on domains in the protein that have specific secondary structures. The domain in this case would be about of 90 to 100 amino acids that are known to bind single stranded RNAs. This is found in many eukaryotic proteins and it consists as we can see [refer to slide], of four of these anti-parallel β-strands that are stacked along two of these α-helices.

So this is the specific recognition motif and specific domain. It is referred to as βαββαβ. Now, this recognition domain or motif is important to realize that the RNA structure the backbone or the specific residues involved, will be interacting with the recognition motifs that we have. These are the N and the C terminals and these are marked with the specific strands that we can see.

So there are four strands of the β-sheet antiparallel in nature, that have β_1 followed by an α followed by 2 β, an α and then another β, in the specific topology that has been mentioned here.

(Refer Slide Time: 05:58)



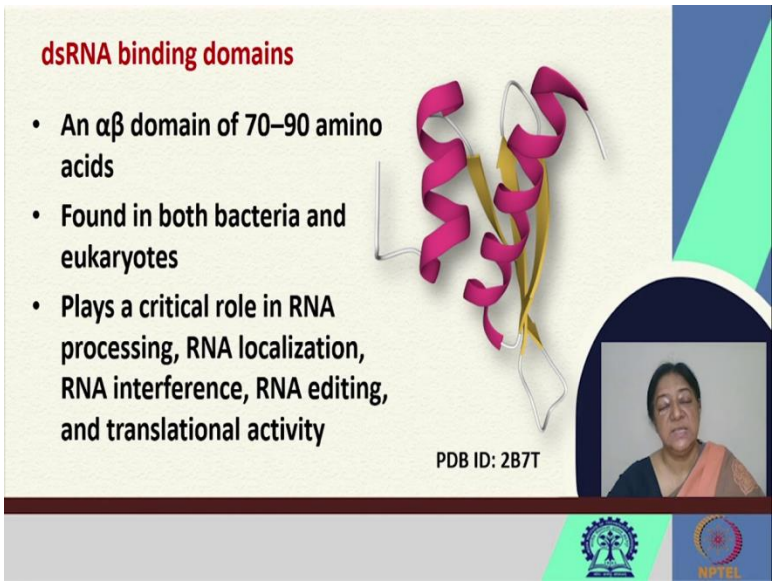
In the recognition motif we have the single stranded bases, that are specifically recognized by the β-sheet and through the two loops that connect the secondary structure elements. So these [refer to slide] are the four strands that we have here and we have the α-helices. Now we have specific single stranded bases that are recognized.

For any protein ligand interaction, the recognition is the most important event that occurs. So whether we are talking about a small molecule ligand or whether we are talking about a macromolecule, we realize that the recognition is extremely important. The motif or the specific domain involved in the process, has specific structural aspects to it, has specific chemical aspects associated with it, that brings about this recognition for the specific function that it is involved in.

The binding is usually mediated by several conserved residues. It could be an arginine or a lysine residue, that forms a salt bridge to the phosphodiester backbone. We know that we have the sugar phosphate backbone and the bases attached to it. Also there are two aromatic residues that can make stacking interactions with the nucleobases.

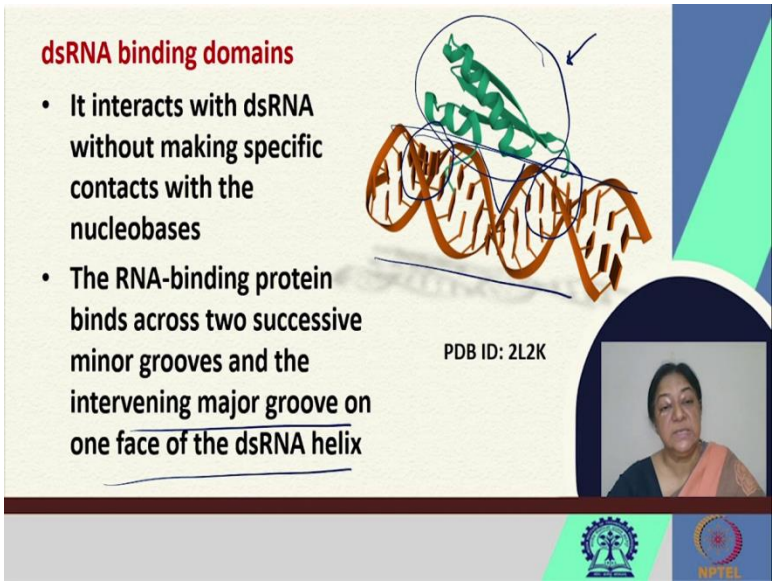
So we studied the specific types of interactions that are possible. There are hydrogen bonding interactions, Van der Waals interactions, electrostatic interactions, as well as stacking interactions in this case, with the aromatic residues that can be involved in static interactions with the bases.

(Refer Slide Time: 07:55)



The double stranded RNA binding domains have a specific motif. This [refer to slide] is an αβ domain of around 70 to 90 amino acids and this is found in both bacteria and eukaryotes. This plays a critical role in RNA processing, RNA localization, RNA interference, RNA editing and translation activity. We realize the importance of the specific binding domain and these are large domains because they have specificity, structural complementarity in the way they are going to bind with their specific ligand of interest; in this case double stranded RNA.

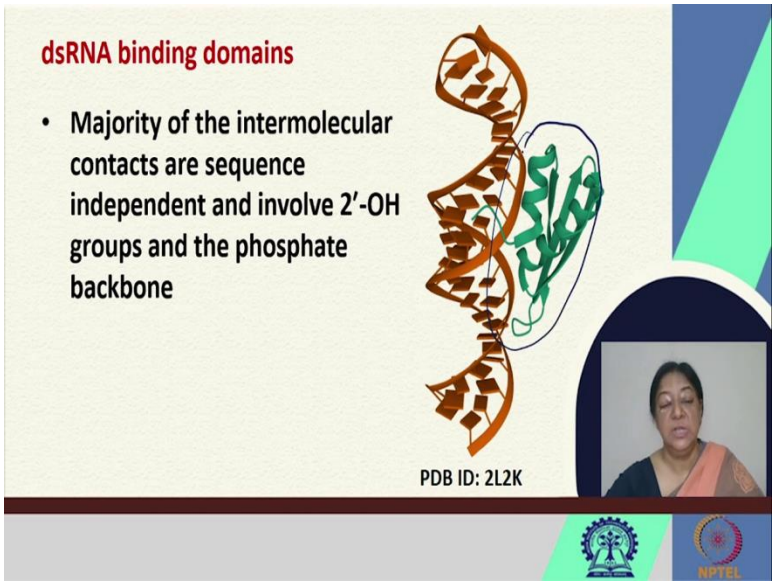
(Refer Slide Time: 08:41)



This binding domain [refer to slide] where we have our double stranded RNA and our protein that is going to bind to the RNA. So, it interacts with double stranded RNA without making specific contacts with the nucleobases. As we can see it is more near the sugar phosphate backbone. The RNA binding protein in this case binds across two successive minor grooves and the intervening major groove.

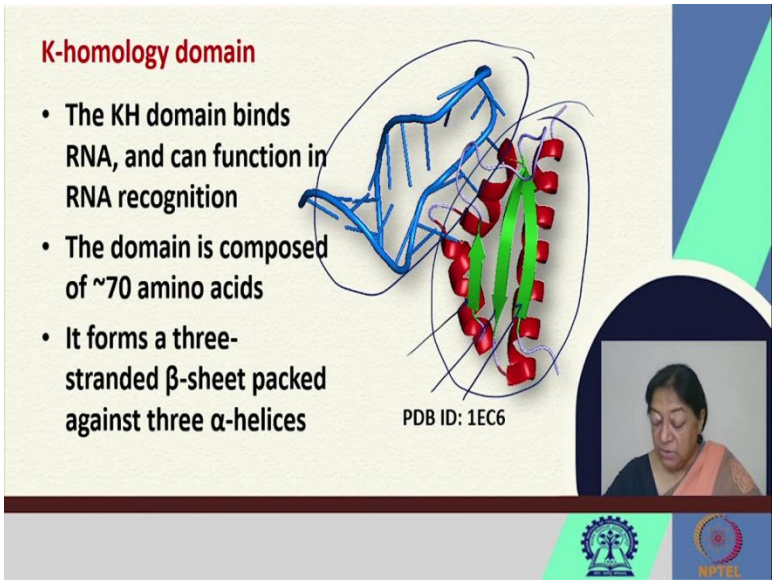
We have the minor groove here, another minor groove here and this is the intervening major groove that is on one face of the dsRNA helix. So, this is our dsRNA binding domain that recognizes the double stranded RNA. It does not make any specific contact with the nucleobases; but nevertheless interacts with the backbone, in a sense that it crosses over two minor grooves and a major groove, in the face of the DNA helix that we are looking at.

(Refer Slide Time: 09:50)



The double stranded RNA binding domains are a majority of the intermolecular contacts, that are sequence independent and they involve the 2'-OH groups and the phosphate backbone, as we can see [refer to slide] from the nature of the interaction in this case.

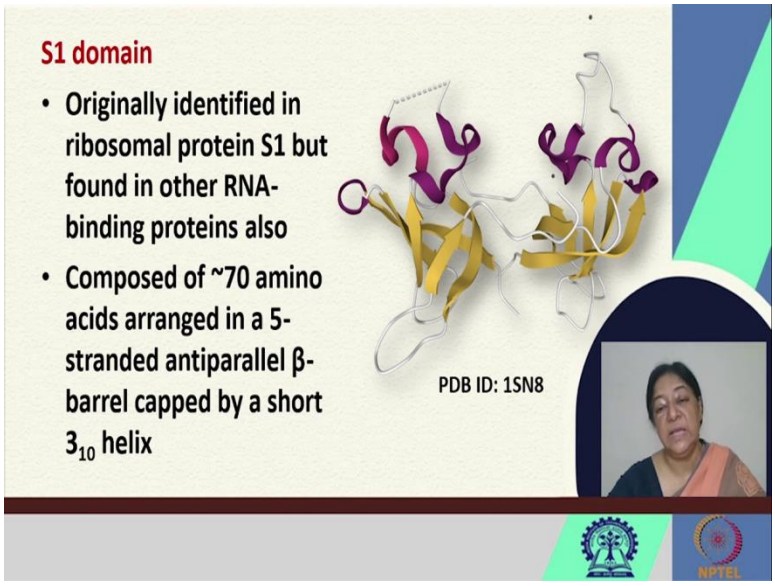
(Refer Slide Time: 10:06)



Another domain is the K-homology domain. Each of these domains have their own characteristics as we can see. So, this [refer to slide] is our RNA molecule and this is our protein that has the specific domain that is going to interact with the RNA. This KH domain binds to RNA and it functions in RNA recognition. The domain again we can see is composed of a large number of amino acids, approximately 70 and it forms a three stranded β-sheet packed against three helices, which can be identified.

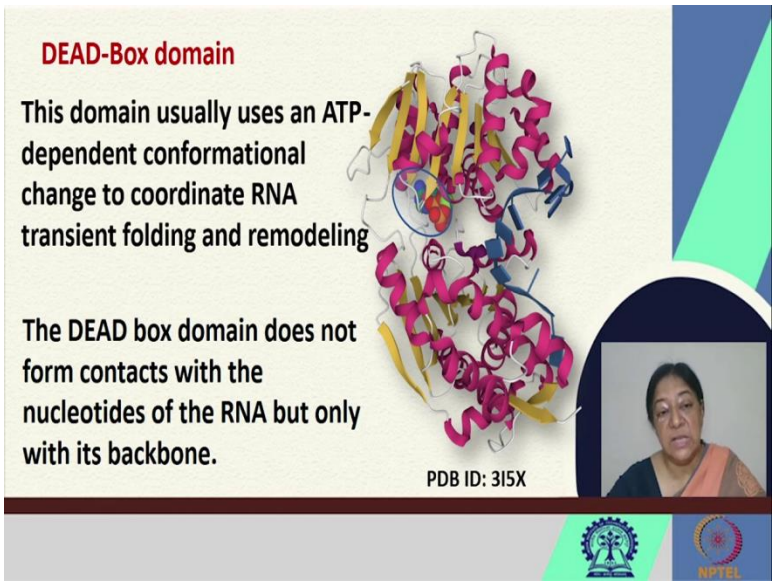
The β-strands are marked in green and the three helices that we see here are marked with red. So, the specific structural motif and the way the geometry of the strands in the β-sheet and the α-helix are located, is important in the recognition motif that we are looking at in the formation of this specific domain.

(Refer Slide Time: 11:09)



The S1 domain was originally identified in a ribosomal protein, but was found in other RNA binding proteins as well. Again this is composed of approximately 70 amino acids that are arranged in a 5-stranded antiparallel β-barrel, that is capped by a short 3_{10} helix. So the necessity of having these structural motifs we realize, is in the recognition.

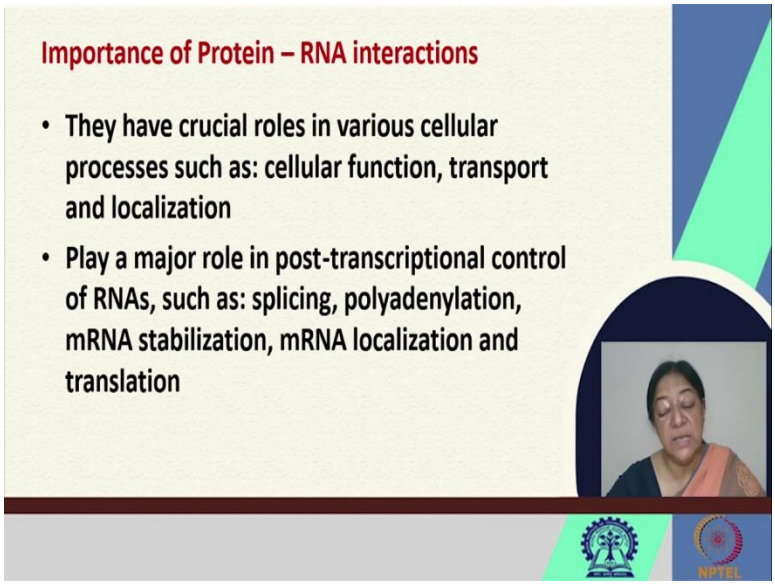
(Refer Slide Time: 11:42)



The domains are large in recognizing the RNA structure. The DEAD box domain is another domain that does not form contacts with the nucleotides but again, interacts with only the backbone of the RNA. This domain usually uses an ATP dependent conformational change, to coordinate RNA transient folding and remodeling.

This is an extremely important aspect of several functionalities associated with RNA, the transient folding and the remodeling of the RNA that is brought about by an ATP dependent conformational change, to the DEAD box domain.

(Refer Slide Time: 12:26)

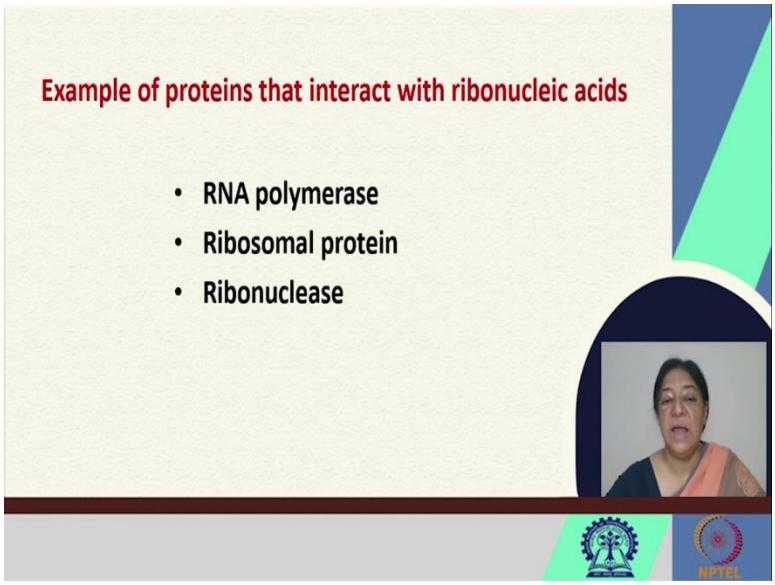


So if we look at the importance of protein RNA interactions, we realize that they have very crucial roles in various cellular processes, bringing from cellular function, transport and

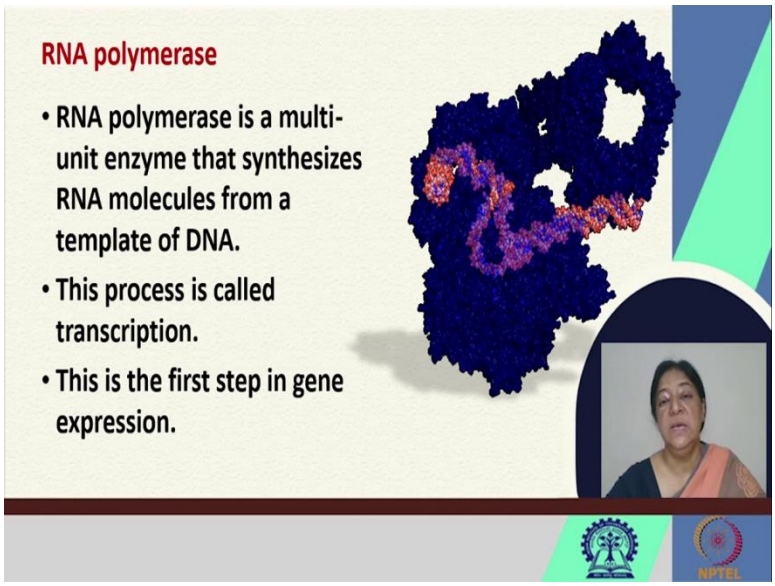
localization. It plays a major role in the post transcriptional control of RNAs such as splicing, polyadenylation, mRNA stabilization, mRNA localization and translation.

In all the translation and post transcriptional processes, we have RNA and RNA processing is very important. Which means that we have the protein RNA interactions that need to be understood to modulate these processes.

(Refer Slide Time: 13:09)



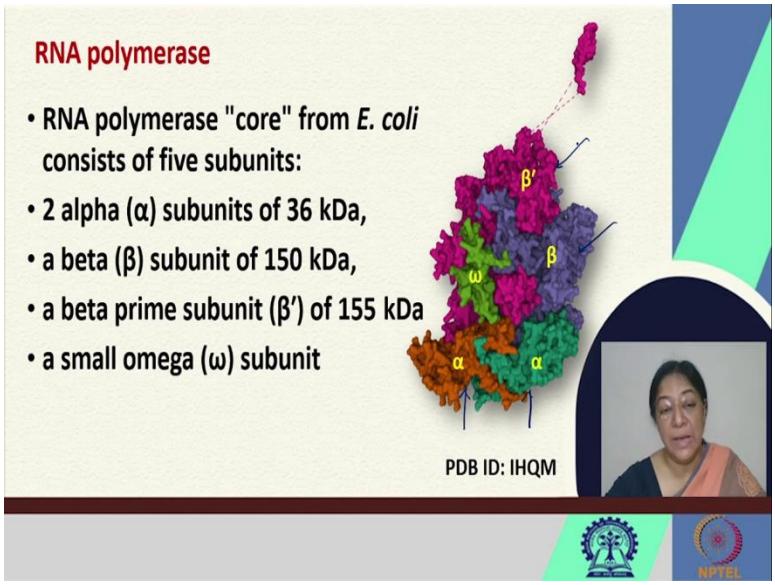
(Refer Slide Time: 13:23)



The examples of other proteins that interact with RNA, are RNA polymerase as we had seen; ribosomal proteins and ribonucleases. The RNA polymerase as we have looked at before, is a multi-unit enzyme that synthesizes RNA molecules from a template of DNA. So, we have the

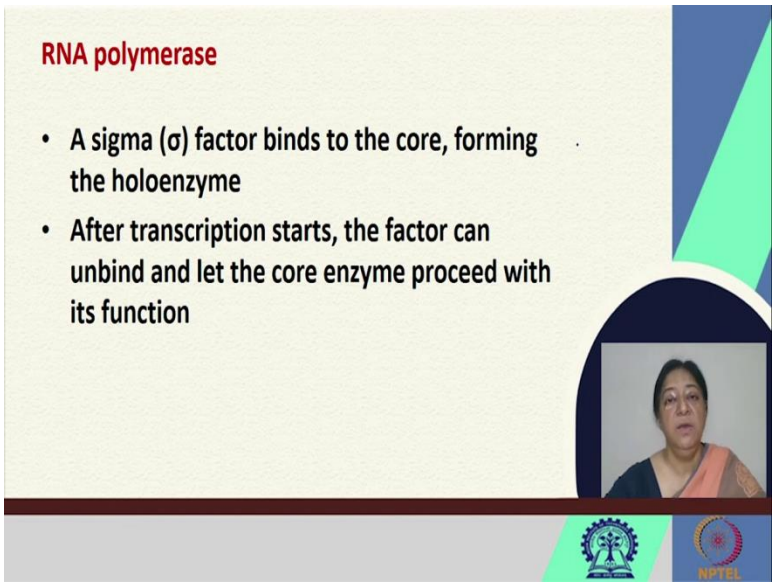
DNA to RNA, the process of transcription which takes us then to the protein in the process of translation. This is the first step in the gene expression that involves an RNA polymerase, interacting with our RNA.

(Refer Slide Time: 13:51)



When we look at the RNA polymerase, it has a "core" from the E. coli that consists of five subunits. There are two α subunits of 36 kDa each, a β subunit of 150 kDa and a β' subunit.

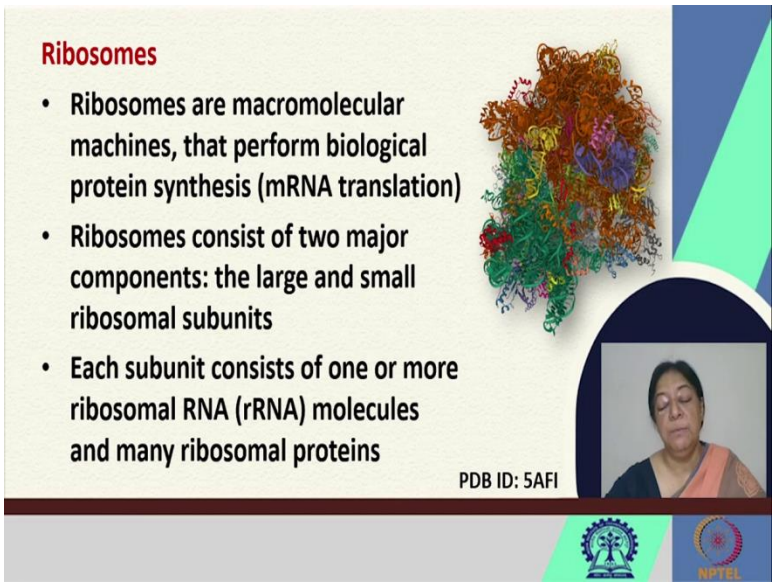
(Refer Slide Time: 14:19)



Then, a small ω subunit and a σ factor that binds to the core, forms the holoenzyme. After transcription begins, the factor then unbinds and lets the core enzyme proceed with its function. So, we realize that the there is a very tight regulation of the process that goes on. We have this σ

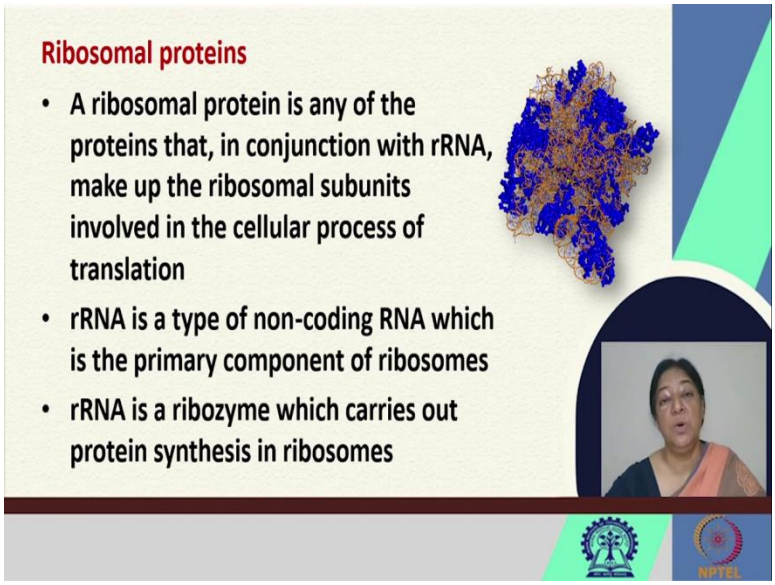
factor that forms a holoenzyme, then the factor will unbind and let the core enzyme proceed with its function; that is in the preparation of the synthesis of RNA from the DNA temporary.

(Refer Slide Time: 14:57)



When we look at the ribosomes, we have these very large proteins. They are macromolecular machines that perform biological protein synthesis. We had mentioned this specific macromolecular assembly, when we spoke about other types of proteins in the motor proteins that we mentioned. The ribosomes themselves consist of two major components. There is a large and a small ribosomal subunit and each subunit consists of one or more ribosomal RNA molecules and many ribosomal proteins because it is the complete machinery that prepares the protein.

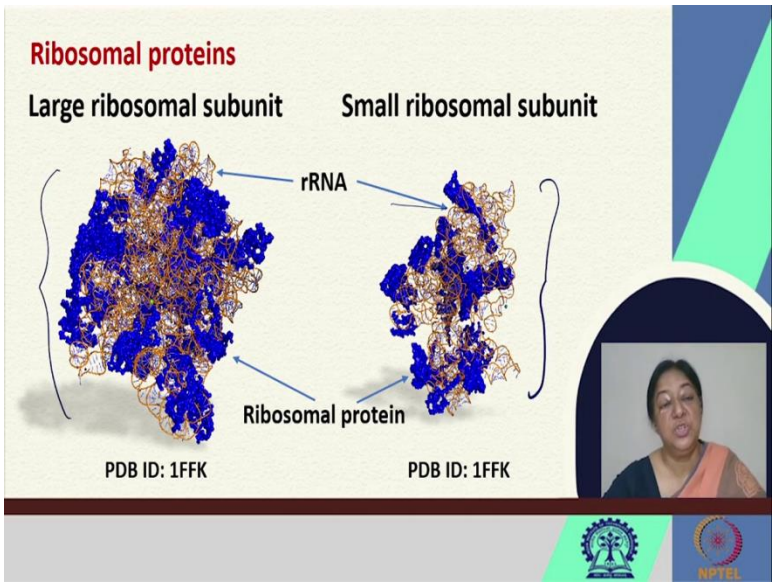
(Refer Slide Time: 15:43)



The ribosomal proteins therefore, is any one of the proteins that in conjunction with ribosomal RNA, make up the ribosomal subunits that are involved in the cellular process of translation. So, we have the DNA to RNA in the transcription process. Now we have the RNA going to the protein in the cellular process of translation and there is the involvement of RNA in both these processes.

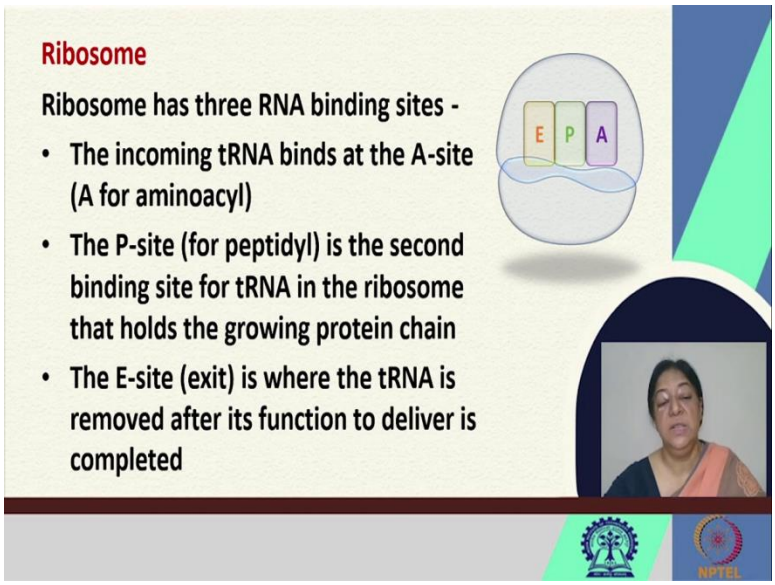
The ribosomal RNA is a type of non-coding RNA, which is the primary component of these ribozymes and it is ribosomes and RNA itself is a ribozyme, which we had talked about in our enzyme lectures, which carries out protein synthesis in ribosomes.

(Refer Slide Time: 16:31)



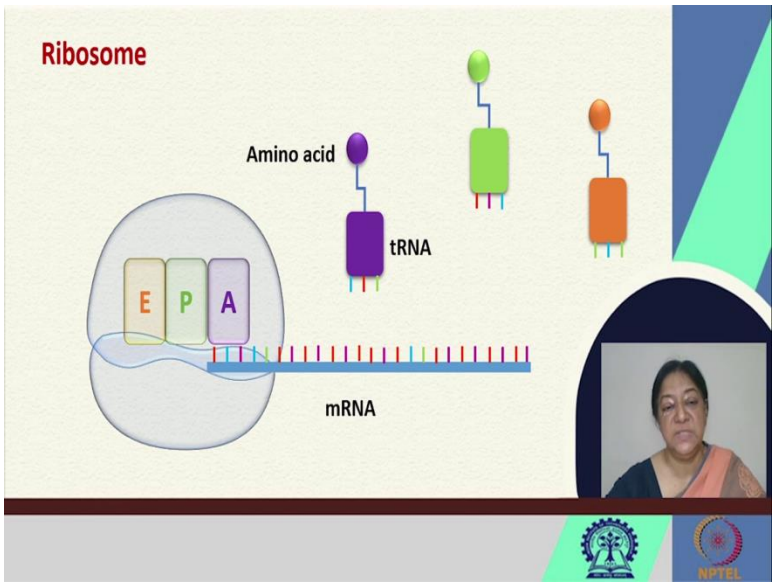
So we have the large ribosomal subunit and a small ribosomal subunit. These two subunits are important. We can see [refer to slide] the ribosomal RNA structure marked here and the ribosomal proteins structure marked in blue and we can see how complex the formation of the large ribosomal subunit and the small ribosomal subunit, together make up these ribosomal proteins.

(Refer Slide Time: 17:05)



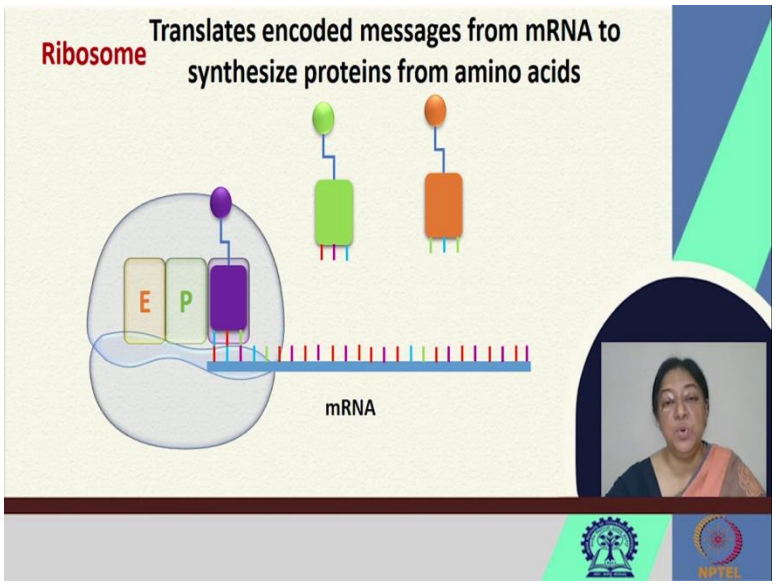
In the ribosome, there are three RNA binding sites; they are the E, P, A. The incoming tRNA binds at the A-site, where A is for aminoacyl. We have then the P-site, P for peptidyl is the second binding site for the tRNA, that holds the growing protein chain because we know that this is involved in protein synthesis. Then the E-site is the exit site, where the tRNA is removed after its function to deliver the amino acid, is completed.

(Refer Slide Time: 17:39)



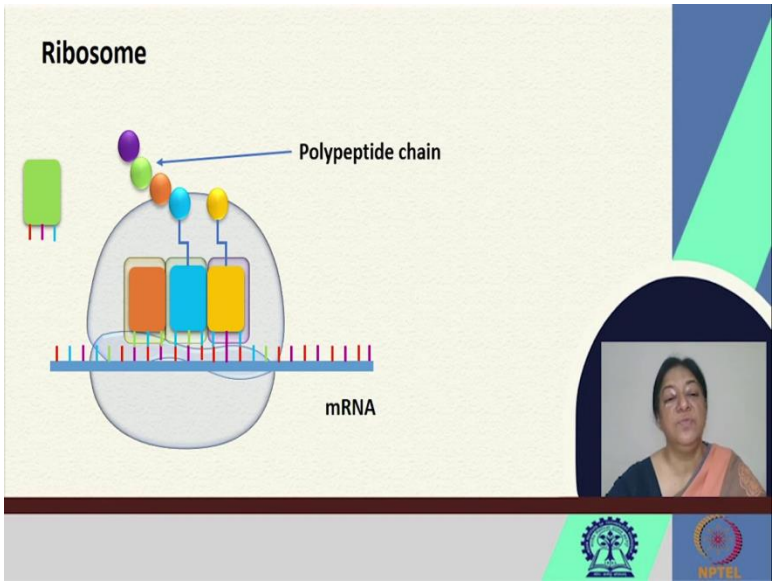
So if we just look [refer to slide] at a schematic of how this works, we have our m RNA, we have our unit with the E, P, A-sites attached to it and then, we will have the amino acids tagged along with the specific tRNA that are then going to bind to the specific site here, with the anti-codon that is present.

(Refer Slide Time: 18:01)



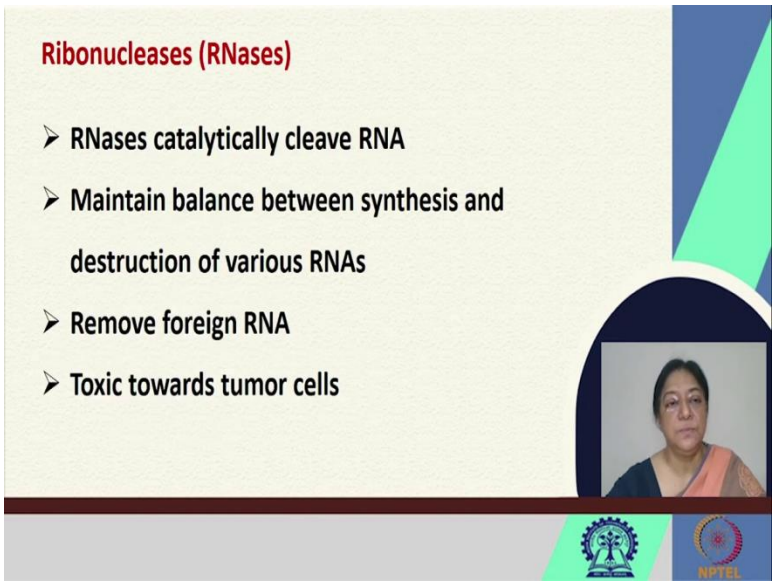
So without going into details of the process, we realize that there is a translation occurring now, the translation of the encoded messages from mRNA, that synthesizes the proteins from the amino acids in the formation of the growing polypeptide chain.

(Refer Slide Time: 18:20)



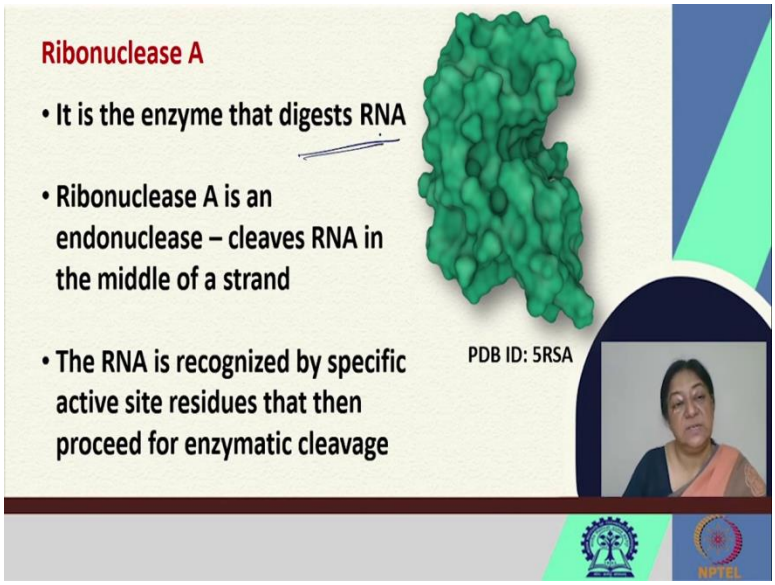
So, we would have the shift of the mRNA the formation of the peptide bond and gradually, the tRNA exit, by delivering its specific amino acid cargo to the growing polypeptide chain. As this continues, we have the growth of the polypeptide chain in the synthesis of proteins, in the translation process that involves this very large macromolecular assembly. Probably one of the most important ones in proteins that we will be looking at.

(Refer Slide Time: 18:57)



The ribonucleases now, are a set of proteins that catalytically cleave ribonucleic acid. We have the RNA polymerase that prepares or synthesizes the RNA and we have the ribonucleases which are rather RNA depolymerases, that cleave the RNA. Now the necessity of ribonucleases is to maintain a balance between the synthesis and the destruction of various RNases. It can also remove any foreign RNA and is always toxic towards tumor cells. We visited ribonuclease A when we studied enzyme mechanisms, into looking at how it cleaves the RNA.

(Refer Slide Time: 19:48)

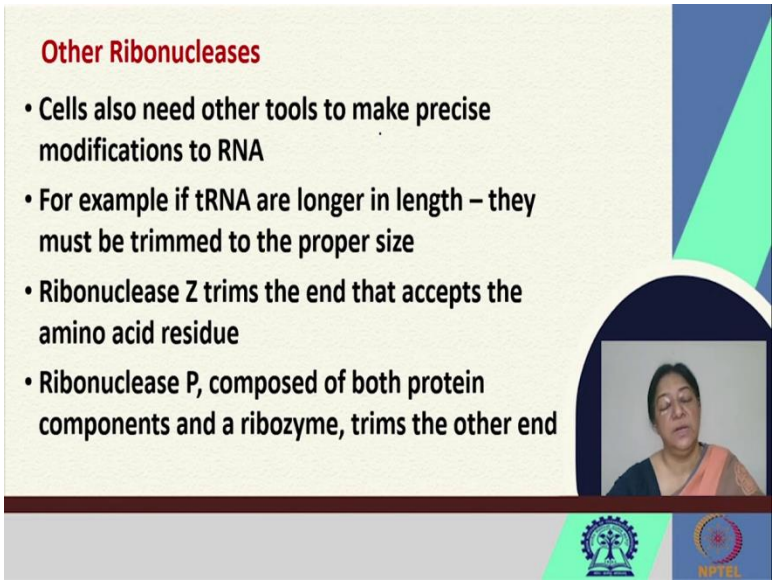


The ribonuclease A protein is a very important protein, that is one of the most well studied proteins. It is one of the enzymes that helps or can digest RNA. It is an endonuclease that cleaves RNA in the middle of a strand and it is recognized by specific active site residues, that then proceed for enzymatic cleavage, which means it will be able to digest RNA.

There again are specific recognition sites, as we have seen before in the enzyme classes; we have a base recognition site, a sugar recognition site and a phosphate recognition site. Knowing that the structure of RNA has these three components in the nucleotide; the sugar, the base and the phosphate.

So we have these specific recognition sites on the ribonuclease A protein, that will then proceed with its enzymatic activity in the cleavage of RNA, a process that is necessary to maintain a very crucial balance between the synthesis of RNA and the destruction of RNA. This is going to be brought about by the RNA polymerase and the destruction by the RNA depolymerase, known as ribonucleases.

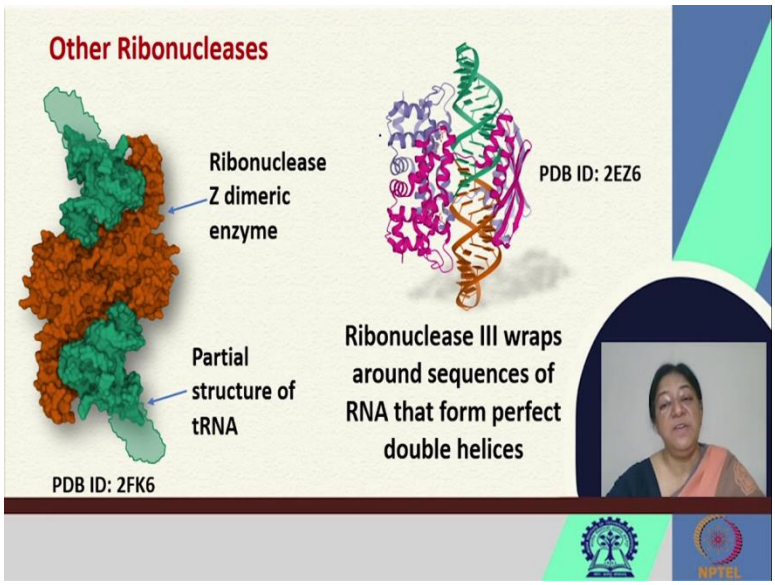
(Refer Slide Time: 21:07)



There are other ribonuclease and each of these have a regulatory role to play and these cells also need other tools to make precise modifications to RNA that are possible with these ribonucleases. For example if the tRNA is too long in length, then they must be trimmed to the proper size to bring about their activity.

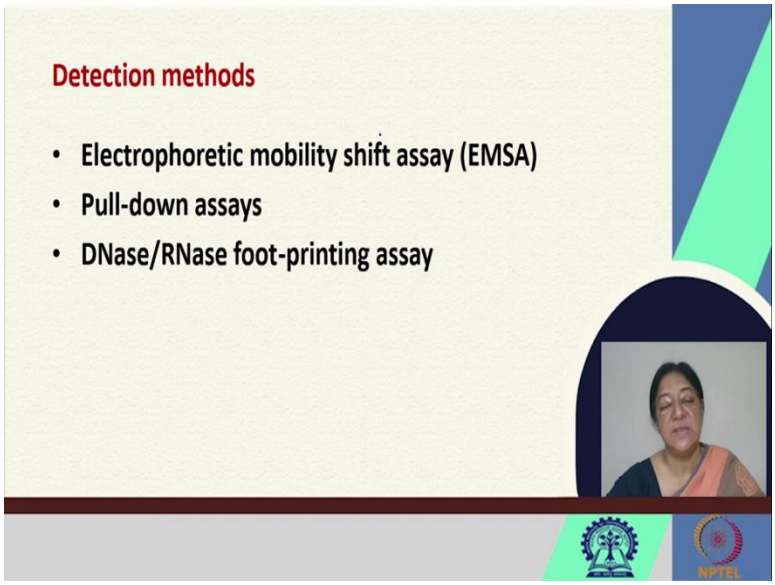
Ribonuclease Z trims the end that accepts the amino acid residues and we have ribonuclease P, which will compose of both protein components and a ribozyme that will trim the other end. So, we realize the importance of having these ribonucleic acid depolymerases or ribonucleases.

(Refer Slide Time: 21:45)



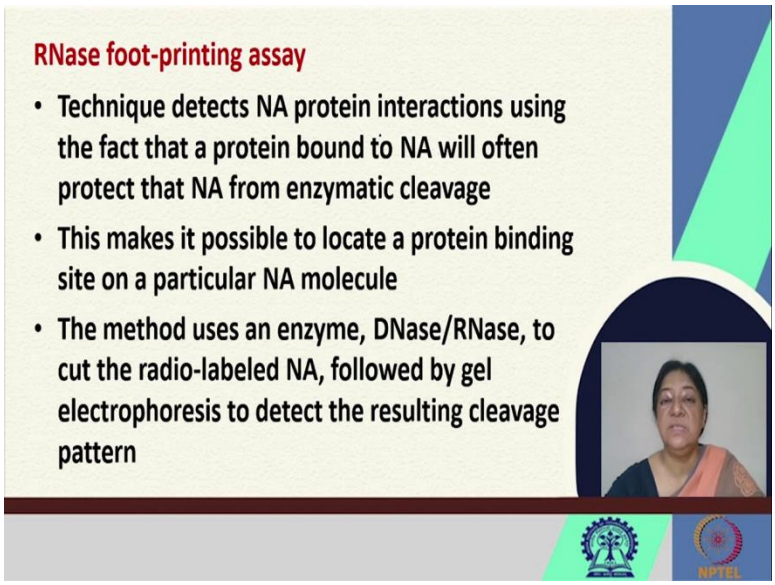
There are some others, where we have the Z dimeric enzyme and this [refer to slide] is the partial structure of tRNA, where it is used to trim the tRNA, to make it of the exact length required for its activity. Then we have ribonuclease III, that wraps around sequences of RNA that form perfect double helices in a very compact structure.

(Refer Slide Time: 22:12)



In the detection methods that we studied in the previous two lectures, we looked at electrophoretic mobility shift assay, we looked at the pull down assay and in this lecture we will study the foot printing assay that is used for DNase and RNase.

(Refer Slide Time: 22:31)

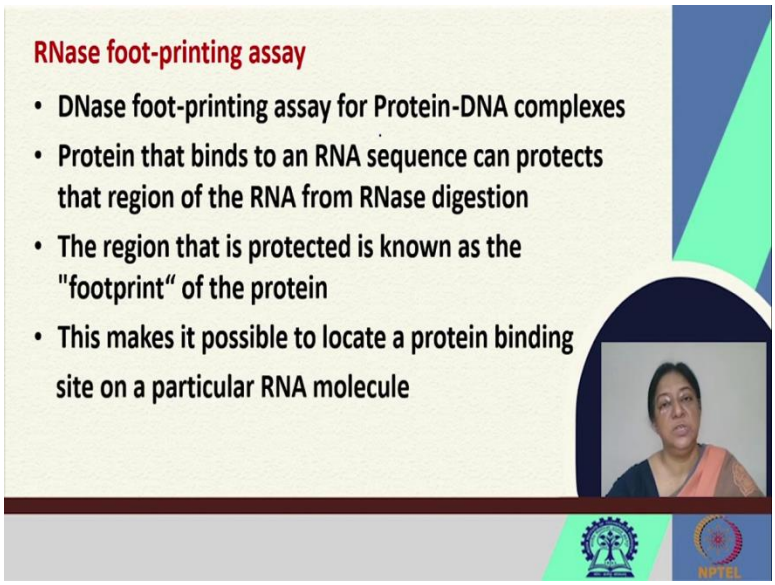


The RNase foot printing assay is a technique that can detect nucleic acid protein interactions, using the fact that a protein that is bound to the nucleic acid, will protect the nucleic acid from any enzymatic cleavage. This makes it possible to locate a protein binding site on a particular nucleic acid molecule.

So, the process involves the interaction of the nucleic acid with a protein. This complex is then cleaved with DNase or RNase, knowing that if there is a protein nucleic acid complex, this will be protected from the cleavage that is going to be brought about by a DNase or a ribonuclease.

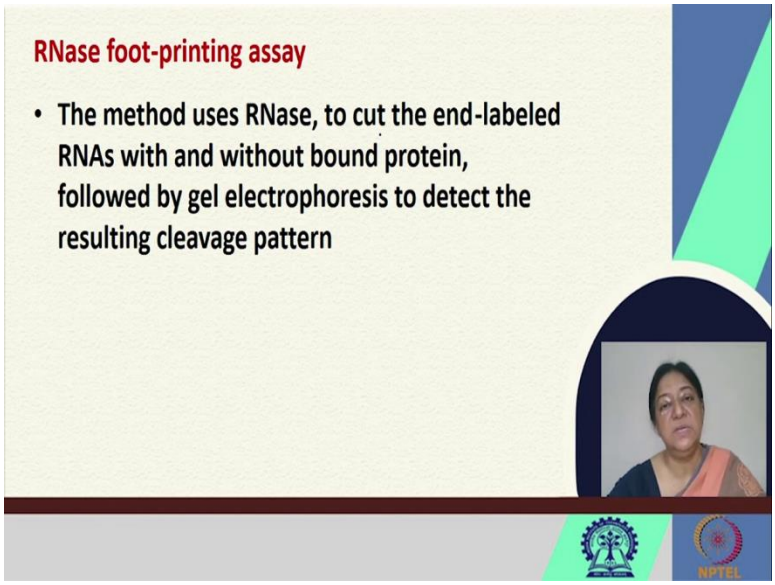
We just saw that a ribonuclease cleaves RNA. However, if this RNA is protected by the interaction with the protein, then this cleavage will not be possible and the RNA will be protected. So this method uses the enzyme, the DNase or the RNase to cut a radio labeled NA, this is followed by gel electrophoresis, that is going to detect the resulting cleavage pattern and then there will be a pattern that would decide whether there is an interaction.

(Refer Slide Time: 23:58)



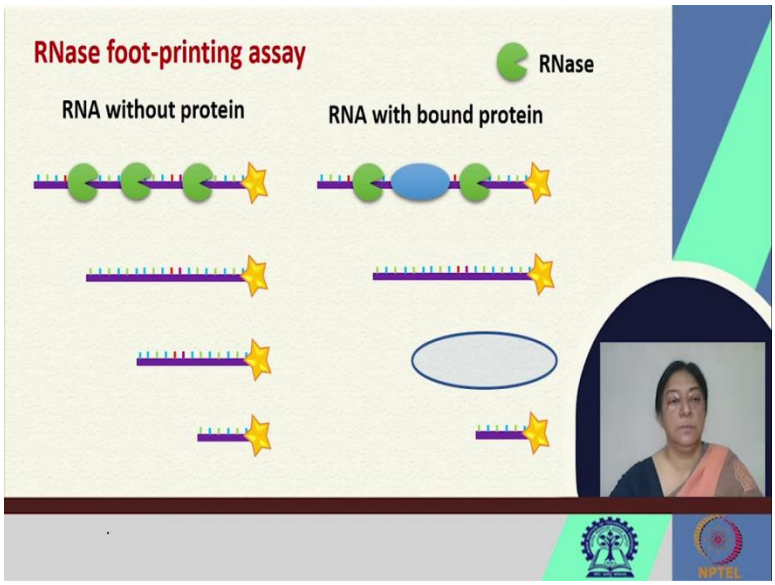
The DNase or the RNase foot printing assay; the DNase is used for protein DNA complexes and the RNase is used for protein RNA complexes. So, the protein that binds to an RNA sequence can protect the region of the RNA from any RNase digestion and the region that is protected is known as the footprint of the protein. This makes it possible to locate the specific protein binding site on a particular nucleic acid molecule.

(Refer Slide Time: 24:31)



The method uses RNase to cut the end labeled RNase with and without bound protein. This will be followed by the gel electrophoresis, that would help in the detection of the cleavage pattern.

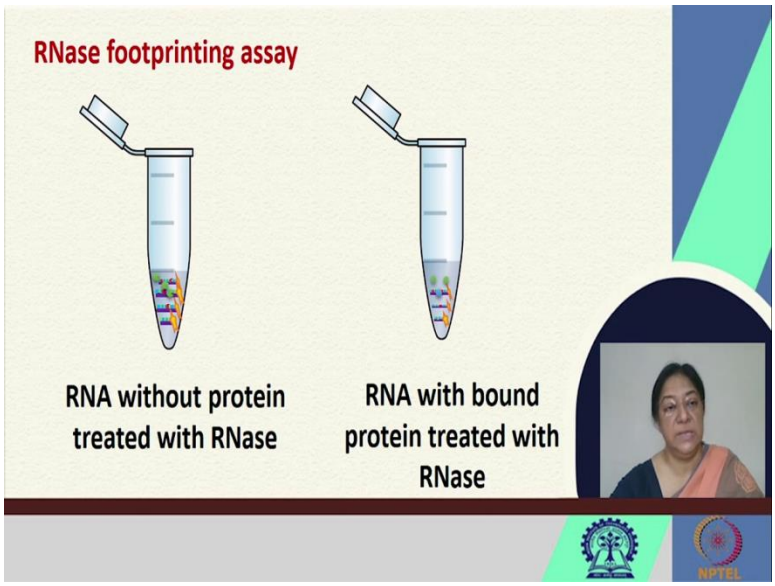
(Refer Slide Time: 24:45)



If we look [refer to slide] at RNA without protein and then we have the RNA with the bound protein, so this is where we have our protein RNA complex form. Now RNase will chop up the RNA at specific locations depending upon its specificity. So, we will have fragments of the RNA without the protein, but once the RNA is bound with the protein, the specific site will not be able to be cleaved.

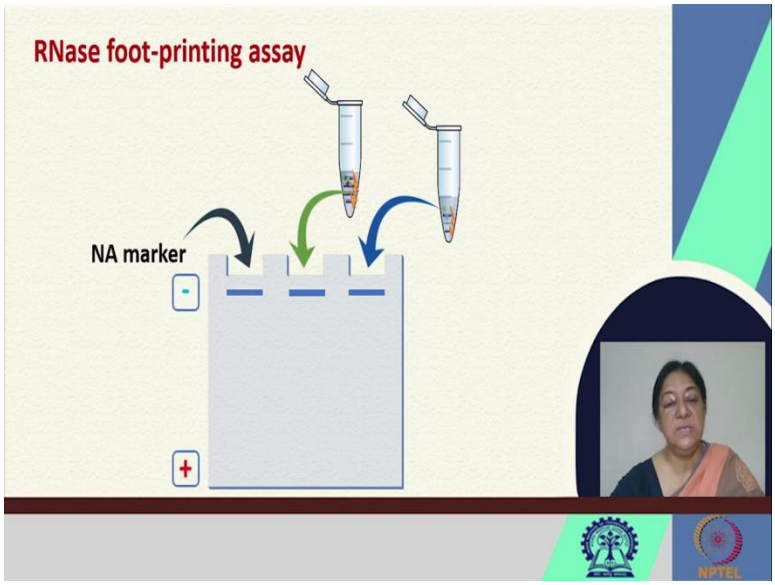
So, we will miss a specific fragment. Then this cannot bind. What we see is we have the cleavage at these points, but no cleavage possible at this point because the RNA has the bound protein to it.

(Refer Slide Time: 25:44)

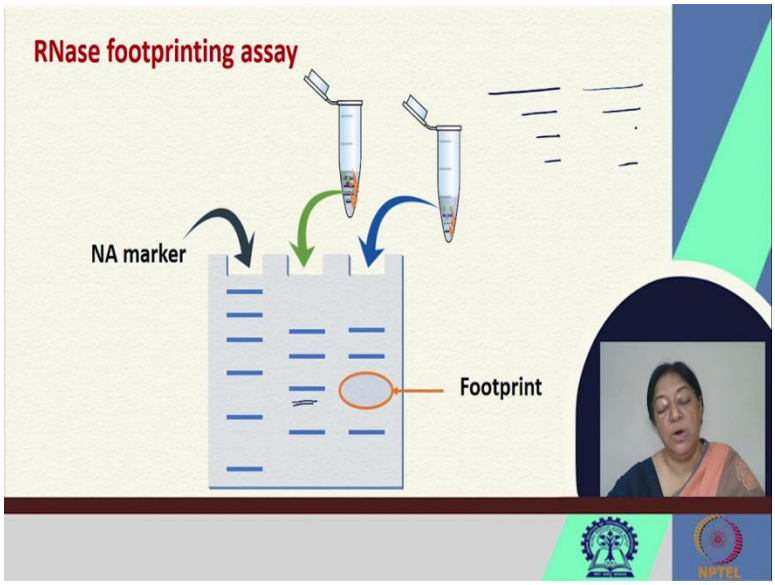


What we have now is two sets of proteins, two sets of samples; one where we have RNA without protein that has been treated with RNase and one we have RNA with the bound protein, also treated with RNase.

(Refer Slide Time: 26:01)



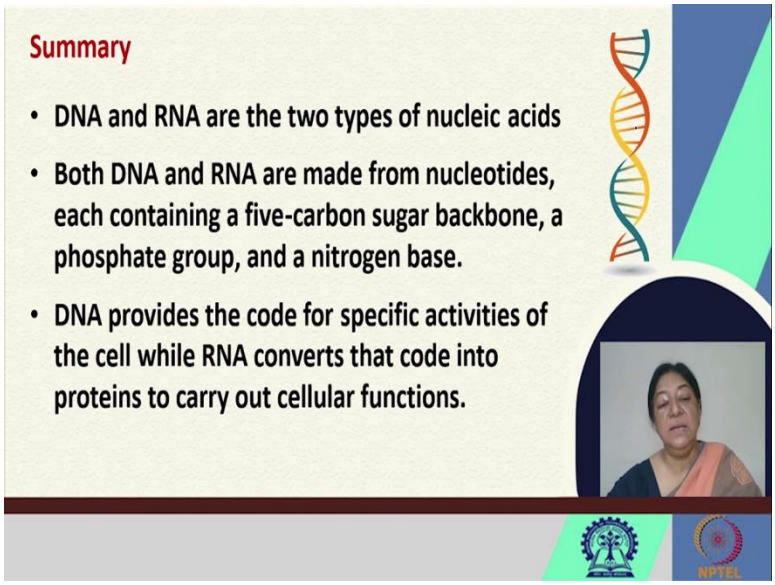
(Refer Slide Time: 26:18)



This is then subjected to electrophoresis. What we will see in an electrophoretic assay is that, when we run the gel in our experiment in the presence of some nucleic acid markers, we will see a distribution, a migration based on the specific size. However at the region where we do not have this specific fragment, we will see an absence which is known as the footprint.

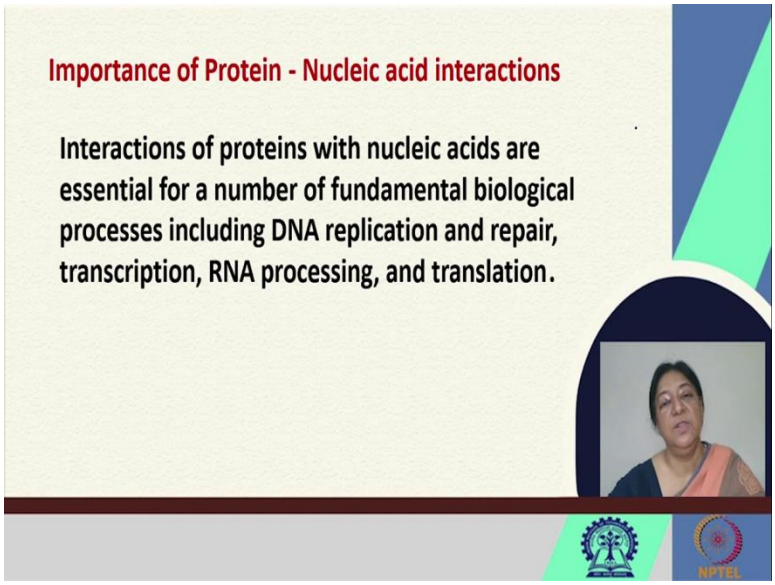
This is indicative of the fact that the protein has bound to the RNA at a point, where the RNase in this case, could not cleave the specific protein. So we do not have the fragment. In one case, we have the different fragments and in the other case, we are missing one fragment because the RNA binding protein had bound to that specific sequence of the RNA.

(Refer Slide Time: 27:06)



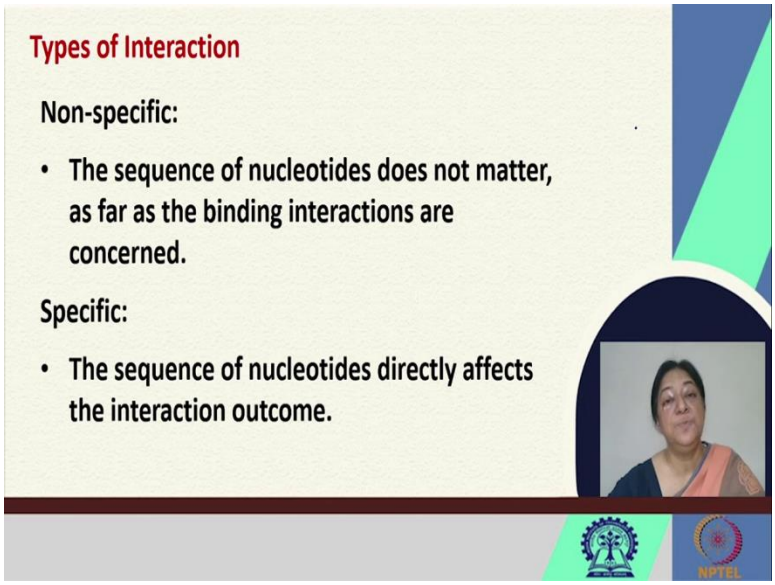
We can say that we looked at the DNA and the RNA that the two types of nucleic acids in their basic structure, they were made of nucleotides and each containing a five-carbon sugar backbone, phosphate groove and a nitrogen base. The DNA provided the code for the specific activities of the cell, while RNA converts that code into proteins, in our central dogma from DNA to RNA to protein.

(Refer Slide Time: 27:38)



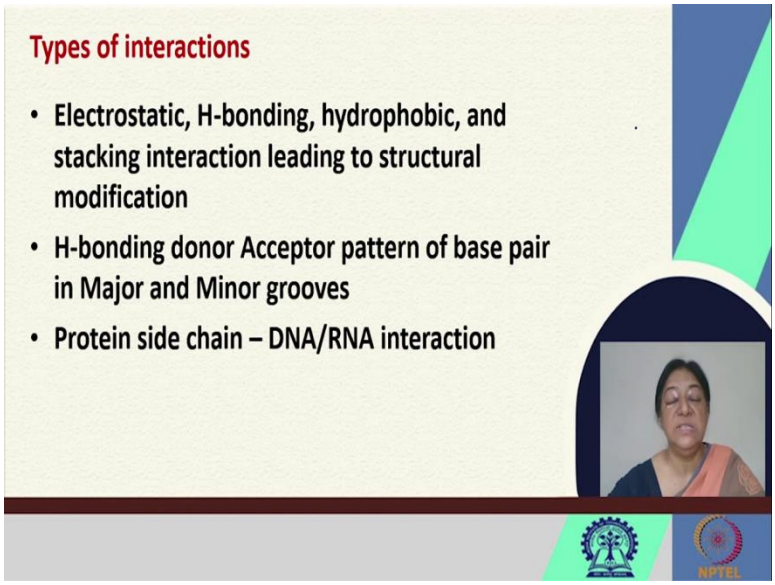
The importance of protein nucleic acid interactions we saw, lies in a variety of biological processes starting from DNA replication, repair, transcription, RNA processing and translation. So, the importance of understanding these interactions of the nucleic acids with proteins, is an extremely important process that is going to be involved in all biochemical processes, in the formation of proteins.

(Refer Slide Time: 28:14)



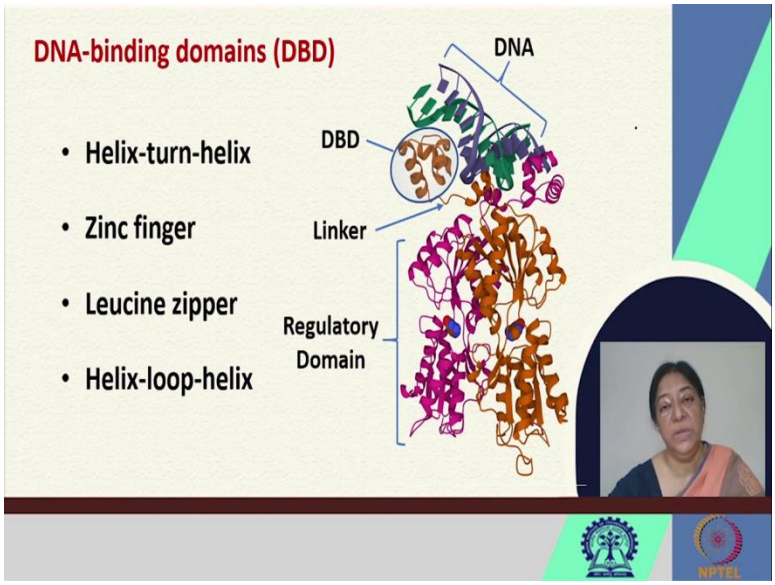
We looked at the types of interactions that could have been non-specific in nature or specific in nature, where the sequence of the nucleotides did not matter in the non-specific types of interactions. Whereas in the specific types of interactions of the proteins with the nucleic acids, the sequence of the nucleotides did matter in the way they interacted with the nucleic acids.

(Refer Slide Time: 28:37)



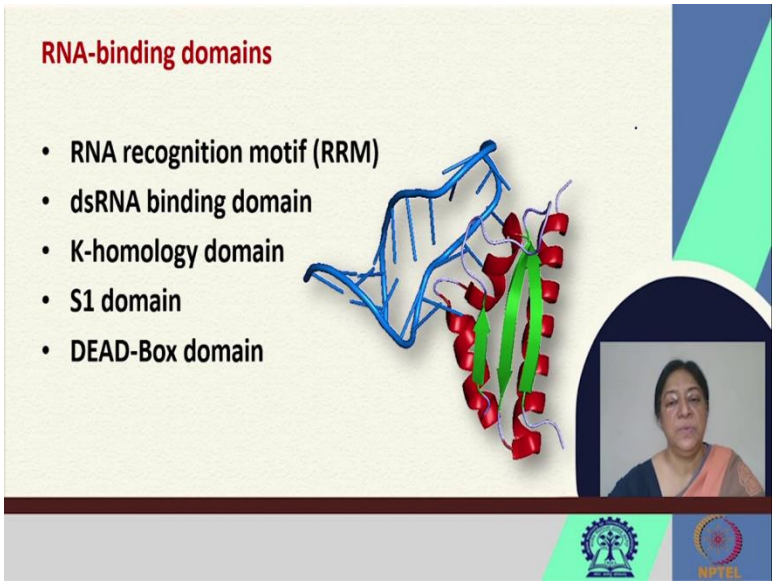
These interactions could be of several types; electrostatic, hydrogen bonding, hydrophobic and stacking interactions that could lead to structural modifications.

(Refer Slide Time: 28:57)



There would be the major and the minor groove interactions and we also looked at protein side chain, DNA-RNA interactions. The specific DNA binding domains that are important in how the protein recognizes the DNA in the helix-turn-helix, the zinc finger type, the leucine zipper type and the helix-loop-helix.

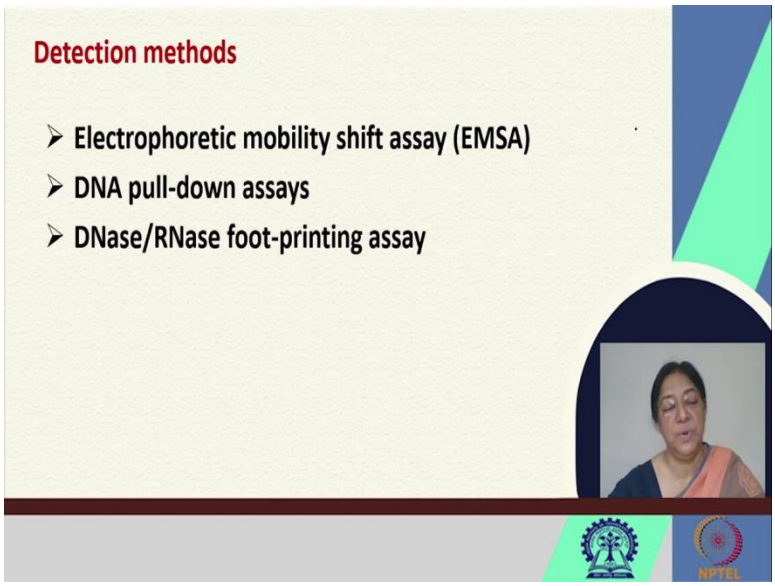
(Refer Slide Time: 29:15)



Similarly, when we had RNA binding proteins; there were specific RNA recognition motifs that involve the recognition of double stranded RNA, a K-homology domain, S1 domain, a DEAD-

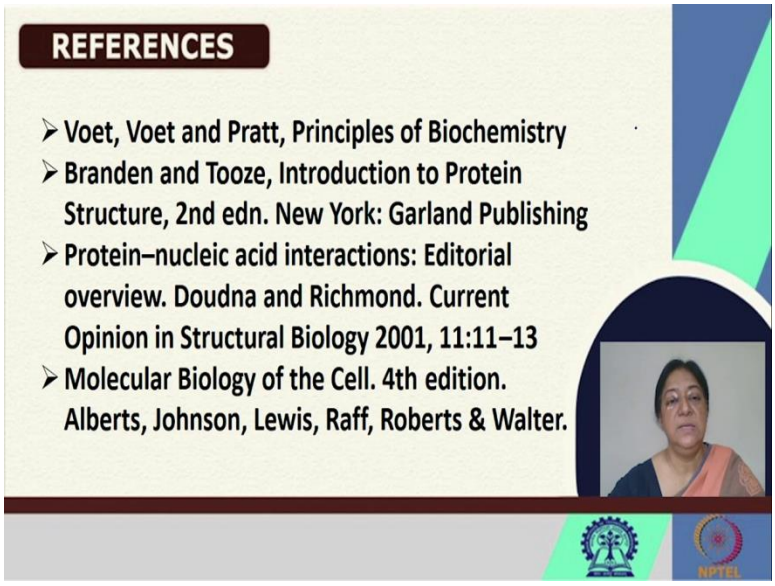
box domain that involve a large number of amino acid residues, a specific structural motif to them, that formed the domain that was required for the specific recognition.

(Refer Slide Time: 29:42)



Then, we looked at the detection methods of the electrophoretic mobility shift assay, the DNA pull-down assay and in this case, the foot-printing assay.

(Refer Slide Time: 29:51)



These [refer to slide] are the references.

Thank you.