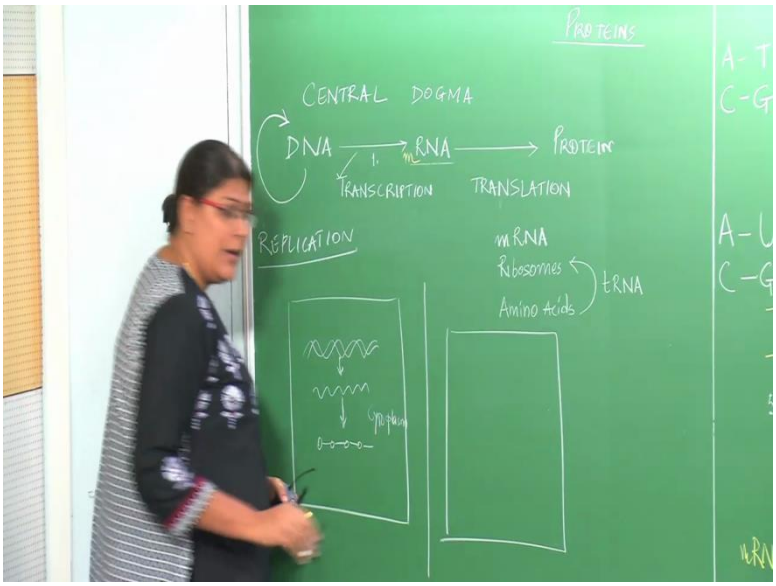
Biology for Engineers and Other Non-Biologists
Prof. Madhulika Dixit
Department of Biotechnology
Indian Institute of Technology, Madras
Week- 04
Lecture - 21
Transcription:
"The Decoding Mechanism"

So, hi and welcome again to these series of videos. Today we are going to talk about one of the most interesting problems in biology and that is, how is it that in our day-to-day lives and in each and every cell of our body whatever information that is stored in the DNA is read and interpreted. Now this decoding mechanism is what I am going to cover today and in my next video. And for the sake of conceptual understanding of the topic I am going to skip a lot of details.

I do not want to bother the audience with too much of the mechanistic nitty-gritty and (meta) mechanistic details. So we are trying to understand it at a very superficial level but it is important to understand the concept as to how the DNA information is read, decoded and interpreted and then what does it lead to. It obviously leads to formation of the working horses of our cell which nothing but the proteins. Now protein is one of them, there are quite a few others but we will stick to proteins in this particular video.

(Refer Slide Time: 24:16)



So I want you to start imagining that DNA which is packaged in our nucleus is like our huge instruction manual, and it is in this instruction manual that all the information is kind of catalogued and kept. And then there has to be a mechanism which needs to, as and when the

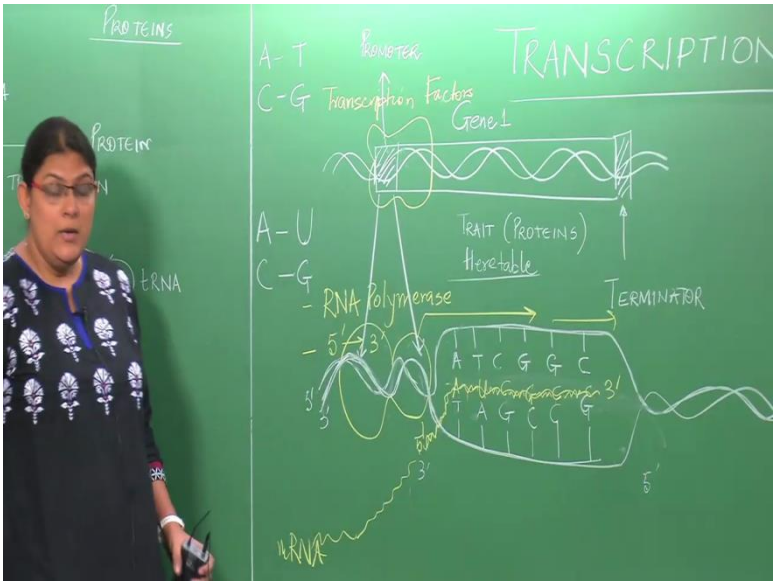
need is, to pick up the relevant instructions, read it and then convert it into a product. So we will start with what is called as the Central dogma of biology, all right? Now Central dogma, what it says is that most of our information is stored in DNA, our genetic material. And then, this is like what I said is your instruction manual.

Now these instructions which are stored in DNA are then copied into RNA. You know, imagine like this, you have a instruction manual which has got a series of recipes and you need to pick out one specific recipe to prepare, let us say, a protein. Now this process of copying the recipe from the instruction manual, let us say into cue cards, right? So this RNA is like your cue card in which you are going to note down the recipe for the synthesis of a particular protein. So that is DNA to RNA.

And then the RNA has all the information it carries this information from the nucleus to the cytoplasm which happens in case of eukaryotes. We will also see how it happens in prokaryotes in a bit. So this cue card is then read by the chef and then it will bring in all the necessary ingredients which are needed to make this protein. Now this protein is the final product, all right? Now, this process, the step 1 in which the instructions are read and are communicated from the nucleus into the cytoplasm by one specific kind of RNA which we call as the mRNA.

There are other species of RNA as well, but this conversion from DNA to RNA is process one which is called as transcription. And then, the (tran) the all the information which is present in the RNA is then decoded, is understood and accordingly the ingredients are brought in for the formation of a complete protein. That (for) process of decoding and the actual formation of proteins is what you call as translation. Now in today's video we are going to talk about transcription, I will come to details of translation in the next video.

(Refer Slide Time: 16:18)



So what are the ingredients for a protein formation to happen? Well you definitely need a messenger RNA, all right? You need the chef is going to put in all these ingredients together and the chef is nothing but the ribosomes which are present in the cytoplasm. There are also ribosomes which are present on the endoplasmic reticulum in eukaryotic cells that is also a site for synthesis of proteins.

So you need mRNAs, you need ribosomes which are the chefs, and these chefs will read the information in the mRNA and bring in the other ingredients which are the amino acids.The amino acids are the building blocks of the proteins, ads you would have read in your biomolecules class. And these amino acids are brought to the chef; they are ferried to the chef by another class of RNA which is called as the tRNA. Now we look at all this in translational video. For this video we will stick to transcription.

So in transcription, what is happening is that the information which is present in the DNA is read and copied onto an RNA molecule, and then the RNA molecule moves out of the nucleus into the cytoplasm. So how is it that, where is it in a DNA that the information is stored. Now, (a) let us assume that this is your long, linear, uncoiled DNA molecule. Now the DNA molecule in a stretch will have specific segments which will carry this information or in simple terms these are like specific files or folders in which the information is kept.

Each unit which codes for a particular trait, let us say this (s) entire set from here to here has a stretch of nucleotides which are coding information, let us say, for synthesis of a protein, protein 1, right? So that information is coded in a stretch of DNA, so each stretch of DNA

which codes for a particular trait; in this case we are talking about proteins, you can also have traits like, which are coding for the tRNA itself. So these sections of DNA which are coding for specific traits and the traits have to be heritable, right, they have to be passed on from (heritable right), they have to passed on from (genen) one generation to other.

This basic unit which carries certain information is what you call as a gene, okay? And each gene has a section which determines from where you need to start reading the code, from where you need to start copying the code. So if I give you the instruction manual and I tell you, start copying from the start page, so you need to know from where to start copying it so that portion from where it gives in the instruction that the gene can be copied from here on is called as a promoter, okay?

Similarly the gene will end, the code message for a particular recipe will end at a certain point and that end point where it ends is called as a terminator, okay? So at the time of transcription the DNA is lose and it is lose chromatin, their promoter is present and let us say this gene has to pass on and it has to be copied into an RNA, read by the RNA, so the script which is written here is going to be rewritten into an RNA molecule.

Now remember, unlike DNA, RNA is single-stranded but just like DNA, RNA can undergo complimentary, it shows complementarity with DNA. So let us see, if you were to have this DNA, okay, this DNA has to first uncoil, this segment of the gene has to uncoil, right, the hydrogen bonds are to be broken to set apart and this is where the gene has to read, okay?

Now this uncoiling has to happen, the reading has to take place and all this; remember in DNA (d) replication the uncoiling was happening by helicases and then the nucleotides were being added by an enzyme called as DNA polymerase based on complementarity. Now what is that complementarity?

As I mentioned, in DNA, an adenine will always base pair with a thymine and a cytosine will always form of base pair with a guanine. Now in case of RNA, wherever there is an adenine, there is no thymine so instead of thymine in case of RNA you will see a uracil. And then again as seen in DNA for cytosine you will always have a guanine, okay?

So this is the rule which has to be followed, but rest of it is the same, it is still complementarity. The hydrogen bonds have to be broken to kind of separate the 2 strands and then whatever is the information, let us say there is a set of sequence here, it has a A, it has a T, a C, a G, a G again and a C, okay? Now the other sister strand on the DNA will obviously

be having a T, a A, a G here, here it will have a C because G pairs with a C, here again G pairs with a C and here you have a G, okay?

And let us look at, let us give it some directionality (so) let us say this strand is 5 prime, this one which is going all the way from here, this strand is a 5 prime to 3 prime, right? Then obviously the second strand is going to be antiparallel, so this will be 5 prime and it will go 3 prime, okay? Now whatever is the enzyme which is going to read it; so enzyme is like the pen which is going to copy it and write it down, now that enzyme in case of transcription is similar but not exactly same, similar, it is called as RNA polymerase.

Now just like DNA polymerase, it needs a template to read, the first thing, second just like DNA polymerase it will keep on adding the ribonucleotides; now these are not deoxyribonucleotides, here you are bringing in the ribonucleotides; it will read the (ribonucleotides) or it will polymerise the chain of RNA in the 5 prime to 3 prime direction. And I had explained why it happens in 5 prime to 3 prime because of the phosphodiester bond formation.

And the only difference, or rather one of the major differences the RNA polymerase has from a DNA polymerase is that DNA polymerase had to have a primer, it had to have a foundation on which it had to then go on adding the deoxynucleotides. That is not required in case of RNA polymerases, in that sense RNA polymerase is a smarter enzyme. So now if the RNA polymerase which will do this conversion from DNA to mRNA, okay?

Now, where does a ribosome, RNA polymerase binds? RNA polymerase actually goes and binds to this region which is called as the promoter and there are a whole bunch of proteins which allow these RNA polymerase to actually go and bind to the promoter, they are called as transcription factors. Now to avoid confusion I am not getting into details of different kinds of transcription factors, they are far more complex in eukaryotes than in prokaryotes, but (f) just for the time being remember that it is these transcription factors which assist the RNA polymerase to go and (b) hop on a to the promoter and then sit on the promoter.

Once it sits on the promoter, so let us say this was the promoter region, right, the RNA polymerase sits on it and then it starts moving in that direction, and as it moves along it starts breaking these intermediary hydrogen bonds, causes the separation of these 2 strands and then this strand becomes the template strand. Why, because this strand is running from, I

would say this particular strand, (sorry rub this off the process, right) so this strand which is going from here to here and is then opening, is (th) running at the 3 prime to 5 prime end.

But the RNA polymerase, the (v) strand has to be antiparallel, right, and it has to be complementary. So the RNA polymerase will then start reading this. So this strand will start reading and (let me just rub this up a bit); so this has separated and wherever it encounters a T, here we will (will) put a A, ribonucleotide, here there is an A, but it does not have a thymine so it will form a uracil, then it will form cytosine, a guanine, cytosine and here since there was a guanine it will form a cytosine again.

And then this synthesis is happening in the 5 prime to the 3 prime direction. And this RNA is then coming out and it (s) pulling out and as RNA polymerase is moving forward it further unwinds it, reads this strand, keeps on adding the ribonucleotides in its 5 prime (end) to 3 prime direction and the mRNA molecule is coming out at the other end, all right?

So what has happened by the end of transcription, let me just redraw this; so you had this DNA, all right? And then you had the promoter, you, to which the RNA polymerase is bound started moving in that direction and then it starts copying and starts giving you an mRNA molecule which is the exact complimentary to the template strand, okay? And that happens from 5 prime end to 3 prime end.

So now you have the mRNA molecule which is ready. And in case of eukaryotes the mRNA molecule then undergoes further processing because remember in case of eukaryotes, the mRNA has to travel a long journey, it has to come out of the nucleus and then dock on to a ribosome and then approach the chef, right? So, it has to travel. So that protection is provided by some additional modifications in case of eukaryotes wherein again the mRNA at its 5 prime end will have some additional nucleotides added, and this, and a few phosphates, and this is called as a 5 prime cap.

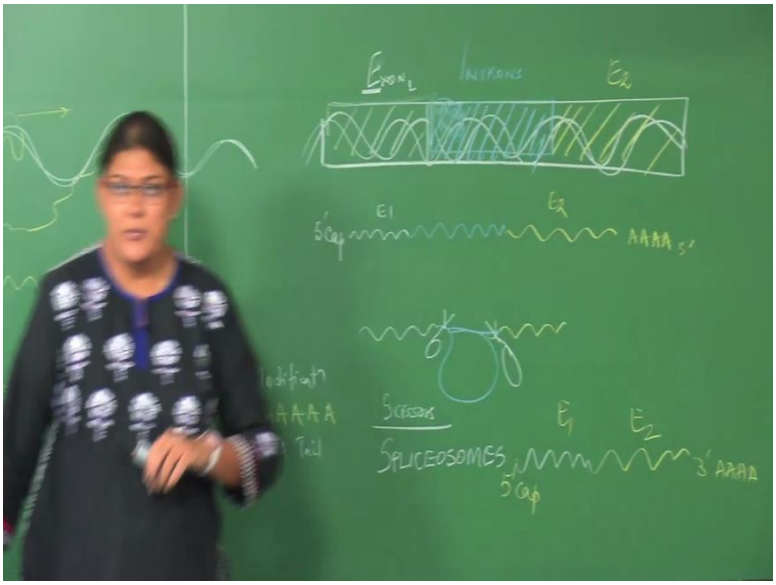
(Refer Slide Time: 19:06)

Similarly the 3 prime end of the mRNA in case of eukaryotes will have a stretch of (a) few adenines, this is like an adenine tail and this is also called as a poly-A tail. So what has happened is unlike prokaryotes; in case of prokaryotes the transcription stops here, okay; but in case of eukaryotes the RNA is further processed because it has to travel outside the nucleus and reached to either the endoplasmic reticulum or the ribosomes in the cytoplasm.

So it has a 5 prime cap, and it ends up having a poly-A tail.Now this process; so you still do not call this as an mRNA in case of eukaryotes; this process is called as RNA processing. Some of them will also call this as post, post means after, post-transcriptional modifications, okay?

So that also happens in case of eukaryotes. So now you have; whatever was written, the instructions which are written in the instruction book which is your DNA; has been (tr), the transcript has been (or) the rather the script has been copied into an (a) mRNA molecule, the molecule has been appropriately processed to further take it out to the ribosome where the actual process of translation will happen, and I will cover that process of translation separately in a separate video.

(Refer Slide Time: 22:42)



But, I want to cover 2 other important things. In case of eukaryotes what we observed is that the gene which is coding for a particular character, right, is not in continuity. What happens is the gene will have certain segments; so let us say this whole thing is one gene, all right? Now this gene will have regions which actually code for a protein so they are called the exons; E for expressing parts, the parts which get expressed into proteins; so it will have exons, let us say this is exon 1 and then, in between it has a region which does not code for anything. It is a (ta), so these regions which do not code for anything are called as the introns, okay?

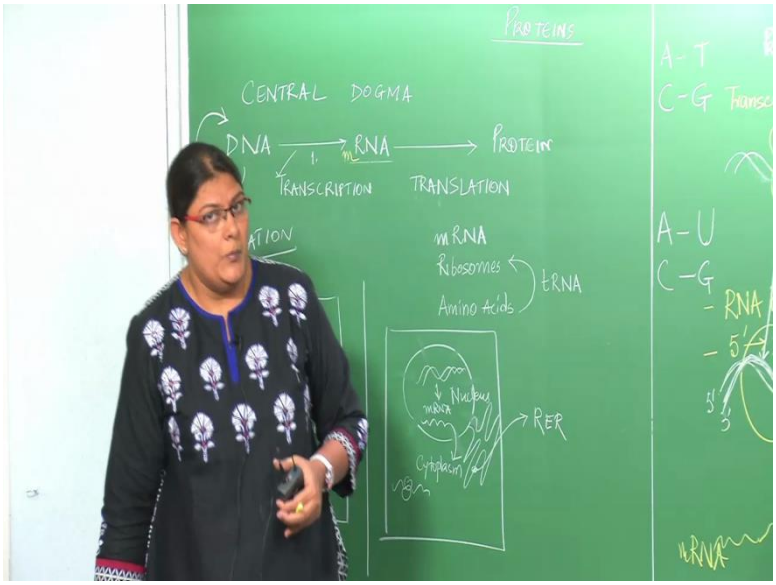
And then you will again have the next exon coming in, let us say this is exon 2. So when the RNA which is being copied is formed, you will have an RNA which will have a 5 prime cap. It will have the exon 1 and then it will have the intron in between, right? And then it will have the next exon coming in, which is exon 2 and then you end up having a poly-A tail at the 3 prime end.

Now this cannot be read, I mean complete, if this is read, the (m) recipe will get messed up. The recipe cannot afford to have this intron, the chef cannot have to have this intron. So there is a process wherein the editing takes place. It is almost like you (have) make a video film and then you kind of cut it wherever the regions you are not happy you do a crisp editing so that the final video looks absolutely crisp and continuous, right?

Similarly this RNA, it is still a pre-RNA, right, what happens is you have a machinery which is called as; and let me put it in simple terms, these are like scissors, okay? Now this scissors are called as spliceosomes. Now do not get into the details of it, what they do is they bring in these (intro s) the exons together and the kind of loop out and they cut it. So you will have a scissor cutting it here, you will have a scissor which will cut it here. You cut it and you remove the intron out, and then you get a continuous mRNA, with exon 1, exon 2, 5 prime cap and a 3 prime poly-A tail. So this happens in case of eukaryotes.

So this was a little bit of a detailing which had to be told because I told you, remember, that the DNA is a little more complex in case of eukaryotes than prokaryotes and part of it, one of the examples is this process which is necessary for the processing of the information. Now again I want to go back to Central dogma. So, Central dogma is the main thematic. There are exceptions to it but (())(23:14) for simplicity sake, the information flows from DNA to RNA which is what you call as transcription; then from RNA into protein which is what you call as translation.

(Refer Slide Time: 25:23)



When a DNA is re-copying itself this is what you call as replication. Now if one way to compare the differences in these processes, let us say in a prokaryotic cell; so this is your prokaryotic cell and this is a eukaryotic cell. So in a prokaryotic cell, there is nothing called as nucleus, the DNA is loosely sitting in the cytoplasm; the DNA gets copied into an mRNA. So the transcription happens in the cytoplasm itself and then the mRNA is read and is used for formation of proteins. The translation is also happening in the cytoplasm. So this is like in the cytoplasm.

But in case of eukaryotes you have a nucleus, right and then it has these openings which you call as the nuclear pore. The process of transcription, formation of mRNA processing of (eram) mRNA, 5 prime capping, 3 prime poly-A tail, post-transcriptional modifications, all that is happening in the nucleus. After that once the mRNA is ready, it comes out of the nucleus into the cytoplasm and then in the cytoplasm there will be either free ribosomes to which the mRNA will go and the process of translation will happen or it will go and sit on those ribosomes which are sitting on what you call as the rough endoplasmic reticulum. So you can have protein synthesis in either of these 2 areas, either in the cytoplasm or in the rough endoplasmic reticulum.

So in today's video what we have learnt is that through the process of transcription which starts at the promoter, where the RNA polymerase binds, and this RNA polymerase separates the strands of the DNA, uses one of them as a template and then from a 5 prime to 3 prime direction it starts polymerising the ribonucleotides leading to formation of an RNA molecule. So in transcription, whatever instruction which was stored in the DNA has been copied into a single-stranded mRNA molecule.

And then in case of eukaryotes this mRNA is further processed, because it needs to go out of the nucleus through the process of post-transcriptional modification. Hopefully this has helped you understand the first step in Central dogma which is transcription. In the next video we will talk about translation. Thank you.