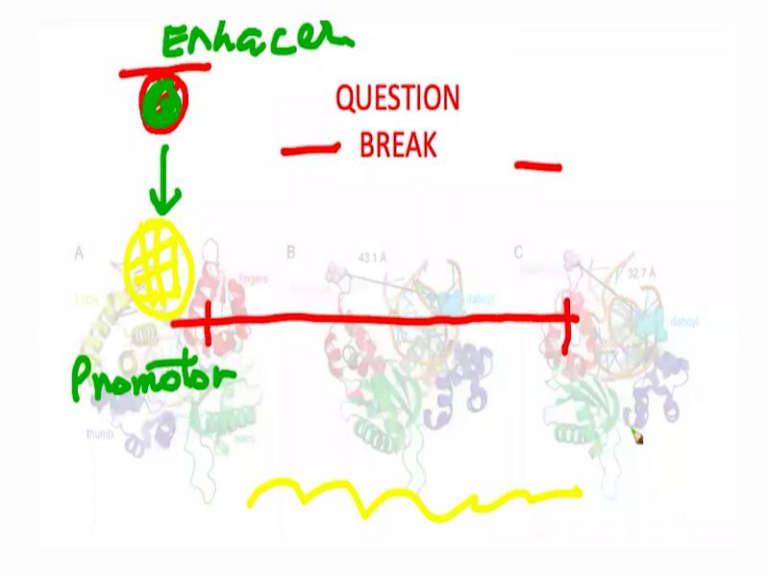
Introduction to Cell Biology
Professor Girish Ratnaparkhi and Nagaraj Balasubramanian
Department of Biology
Indian Institute of Science Education and Research Pune
Central Dogma: Transcription - Part 3

(Refer Slide Time: 00:15)



So, I will take questions over here and then continue.

Student: Sir, please repeat that the function of enhancer region.

Professor: So, the enhancer region, so what you saw over here was, let me draw it for you, is basically there is a piece of DNA which we are defining as a gene with the initiation part and a termination part. And what you saw, and I will draw this in yellow, was RNA polymerase sitting over here being called by another protein called as a transcription factor. And a transcription factor is a protein which sits on the promoter and is responsible for directing the RNA polymerase to make a RNA copy of this gene, which, with the boundaries on the left and on the right.

Now, it turns out that there are pieces of DNA, now I am going to draw them over here, which are further away from where the gene is. And it turns out that there are proteins which are bound to these pieces of DNA and they actually trigger this the moment of this complex. So, if the race is about to start, all the proteins including RNA polymerase, are sitting near the promoter element, but they are not actively transcribing RNA. They have to be told to start.

And this job of telling them to start is done by a protein which is sitting on a piece of DNA called a enhancer. And the enhancer is different from the promoter in the sense that the promoter is where the initiation complex binds, and the enhancer and the protein bound to the enhancer will come close to this piece of DNA and trigger it. So that is the job of proteins bound to the enhancer.

Student: Sir, we are denoting enhancer as that sequence or that protein which the figure shows?

Professor: So the enhancer is a piece of DNA, but the piece of DNA does not trigger transcription. We know there is a protein sitting on the piece of DNA which triggers transcription.

Student: Okay, sir. One more question. You told about coding and non-coding genes. Besides having evolutionary significance, is there any special advantage of having non-coding genes?

Professor: Well, way beyond evolutionary significance, for example, soon you will have hear of this very important piece of RNA called transfer RNA. And transfer RNA is very, very important as you realize in the translation process, and it is basically RNA and it does a function in itself. Similarly, the ribosome and also you will look at splicing factors all of these RNA enzymes. So these are functional pieces of RNA, which, with executive functions in the cell.

So, always remember that even though we are told that DNA makes RNA makes protein, both RNA and protein have structural and functional roles inside the cell. It is not always protein which does all the work in the cell. RNA also does a lot of work in the cell.

So, RNA is much more than just being a transient information intermediate, mRNA is a transient information intermediate, tRNA is not. And the RNA which is in the ribosome is actually an enzyme. The RNA in splicing complexes is actually an enzyme. So, a lot of information from DNA is transferred to protein, but it is also transferred to RNA, which is the endpoint and that RNA is never made into protein.

Student: Is there a need to change thymine to uracil in RNA?

Professor: So, that is a question which is not completely answered. All we know is that just like you have iPhone 10 and iPhone 14, in nucleic acids in the cell functionally you have RNA and DNA, and they are cousins, but they are different in terms of the sugar and in terms of one of the bases, one of the four bases. And somehow evolution has made it so. This is the way things are at this point. Why has this been the final way in which all life functions is not very clear.

There are of course certain properties which the ribose sugar and uracil give to the RNA. So, RNA is similar to DNA, but subtly different. So, hopefully an answer but not a very clear answer. Questions?

Student: What is the role of the DNA between promoter and enhancer?

Professor: So, you will realize in time that out of the whole human genome 3 billion base pairs, a very small fraction, probably around 10 percent, really codes for protein, another 10 percent may code for other pieces of RNA. A lot of DNA in your bodies, and remember, there is the same DNA and every gene in your body, sorry, in every cell in your body does not seem to have any obvious role at all. And for a long time this huge chunk of DNA which is somewhere between 50 to 70 percent, 50 to 80 percent, was given the term junk DNA.

Now, it is still controversial whether this DNA is junk or whether by just being there it provides it has some role. So, for example, if you look at the so called ratio of coding pieces of DNA to junk DNA, it is really, the fraction is very small for vertebrates and mammals. But for eukaryotes and more simpler organisms, there is very little unused or junk DNA in those organisms. So if

you look at evolution from simple organisms to invertebrates to vertebrates as a trend in evolution, our accumulation of so called non-functional DNA is really high in the higher organisms.

Student: Thank you sir for answering our questions.

Student: I remember reading about sigma factors and rho factors which were needed for initiation and termination of the process of transcription.

Professor: So, which is why I have been careful to tell you that RNA polymerase is not alone. It does the main job, but it has a lot of accessory factors. And in fact, one of the pictures you saw the accessory factors.

Student: Sure sir, sir I have another doubt, after the DNA unbinds and the mRNA is produced, the unbinded portion of the DNA rebinds just by making new hydrogen bombs, right?

Professor: Yes.

Student: Okay thank you.

Student: Sir, I had a doubt in the activator protein part, what is it sir?

Professor: So, let us think of it as an analogy. Take a road which is in front of your college or in front of your house and let us say that one kilometer road has to be tar. And effectively the whole team of having assembles with the raw materials ready to tar the road, 1 kilometer of road, but they do not start work unless a supervisor comes and says start work.

So, this enhancer DNA, which is a little further away, has the ability to tell an assembly of factors which have the ability to do the work, but will not do the work unless they are told start. So, the job of the protein bound to the enhancer element is to say start to the transcription factors and the polymerase which is bound near the promoter region.

Student: Yes sir.

Student: Sir, I had a small doubt. In the video like it showed that RNA polymerase like goes to the promoter region with the help of more proteins and then enhancer activates it. Sir, so how is termination done like with the RNAs running across the DNA how will it exactly terminate?

Professor: So, a simple way you can visualize this is again using the example of the road, tarring the road which I gave you. If somebody is tarring the road and there is no supervisor to tell them where to stop, where will they stop? So, logically, they will stop when the older road runs out, which is not true in the case of DNA, because DNA is kind of continuing. Another point at they will stop is the, if for some reason there is a big roadblock in the middle of the road, if somebody has put a roadblock, then they will automatically stop at the roadblock.

So, what is happening in the terminator regions is that the sequence of DNA over there probably folds into a secondary or tertiary structure which acts as a roadblock. So the polymerase goes and hits that roadblocks, and because it hits it with great force, you can see the polymerase is moving with great force, it is thrown off the sequence of DNA and that is how termination happens.

Student: And sir where does this rho factor come into play?

Professor: Again, I think that is very detailed and let us not go into it immediately for this level of class.

Student: Sir, the promoter region you said that is just a section of DNA, right?

Professor: It is a section of DNA with a certain sequence.

Student: So, is not that an cyclic argument, because you use the promoter to identify the gene, but then to identify the promoter, do you have to scan the DNA first?

Professor: Right, so, the way this is done best is that you take a cell and in that cell you break open the cell and take out all the RNA from the cell, because the RNA you know is nothing but a copy of a gene and you sequence all the RNA. And once you sequence all the RNA, you back trace where the RNA is coming from, from which stretch of DNA. So the sequence of mRNA actually tells you the sequence of the so called gene area between the initiation and the ending points.

Once you map your RNA to your DNA, if you look, let me use the term 5 prime or upstream, you look upstream to the start site which is mapped on to the DNA and you start looking for motifs in that site, because you know that is where the machinery is probably sitting. And when you start doing that, you realize that there are special sequences which are found in that piece of

DNA, which are generically not found within a gene or in random locations in the DNA. And one very famous site is called as the TATA box, nothing to do with the company Tata, but TAATA that is the motif. And that is how we discovered motifs where the initiation complexes sit, especially the transcription factors and the polymerases.

Student: Okay sir thank you.

Student: Sir, when you say that the activator protein activates the RNA polymerase to initiate the transcription, it must be a chemical dialogue between these two proteins, right?

Professor: So, you will realize as you go more deeper into biology, rather than using the term chemical, it is better to use the term contact. So, it is a physical dialogue, it is like a punch. So, both proteins physically come in contact with each other and the physical contact triggers something. Now, you are right in the sense that the physical contact may involve modifying the initiator complex, but that is not completely clear. One way a protein A can modify a protein B is by tagging it with a chemical group. So, in that sense, your comment is right. But that is not completely proven.

What we do know is that proteins, the so called protein sitting on the enhancer complex physically comes and touches some part of the initiator complex and tells it to run. So it is not a long distance interaction. It is a proximal contact interaction.

Student: Yes, sir. Also, sir which is the smallest gene discovered? You told the largest one was dystrophin.

Professor: Yeah, so Google. So, the reason the answer these questions or answers are not completely useful is because I have a feeling that the smallest human gene is not clearly defined. Also, even after 20 years after sequencing the entire human genome, the answers may actually change because defining genes is not as simple as you would think.

And since we are living on Earth with such a huge variety of life, I prefer not to live in a human centric universe, because even cockroaches may have the, so the more interesting question is, which is the longest gene ever in the animal plant in life per se and I do not think we have an answer to that question, simply because we have not really made a dent in sequencing all the

animals, microbial and plant life on this planet. So I am saying it is a reasonable question, but you can just Google it. It is not that interesting.

Student: Sir, how does the spliceosome identify the border between exons and introns?

Professor: So, this is a question which relates to the fact, I am again being a little simplistic, but I want everybody to understand what is going on, you are dealing with only four bases A, T, G and C. You have in each cell of your body 10 raise to 9 such bases and they look like a very simplistic four letter code. It is not actually complicated, like an amino acid code which is a 20 member code.

Now, using these four bases, you have to have information in chunks of your DNA, which are called genes. Not only that, you need to have specific regions for binding, which are called promoters, you need enhancers, you need termination sites, all using this four letter code. So effectively what this comes down to is that along this huge stretch of single, double stranded DNA are these patches of information specifically for let us say binding of DNA polymerase also and RNA polymerase also.

Once the stretch is copied into a pre-RNA, obviously, when we look at a pre-RNA, it is a long strand of single stranded nucleic acid. It does not have green and red markers on it, but they are motifs which are called as intron-exon boundaries. Now, these motifs are recognized by the spliceosome complex. Spliceosome binds over there or pre-spliceosome complexes bind over there and the spliceosome cuts out that piece of RNA.

So, obviously, what I am coming to is that there are intron exon boundary motifs. Now, remember that this whole business which we have been talking about for the last four lectures is about transfer of information from DNA to RNA to protein. And if there is any mistake in either not removing that piece of intron or making one single base pair in that intron that will lead to the production of a protein which will be non-functional, which will then cause disease.

So, all the processes you have been studying including RNA splicing are precise regulated processes and they are happening every second in your body and they all have to be perfect or near perfect, otherwise your body is not going to be able to live. So, I think I answered a little beyond what you asked, but I want to emphasize a few points. Questions?

Student: Sir, what happens to the lariat loops, like after the introns has spliced, are they destroyed inside the cell or they are left as such?

Professor: So, do realize that cell is a independent city. It does negotiate with things outside its boundaries which is the plasma membrane, but a lot of stuff, it is a very green system if I use modern terminology. There is a lot of recycling going on. So, the lariat is basically broken down by RNA hexonucleases and nucleotides recycle.

Student: Sir, I have a question in transcription. Sir, how does the RNA polymerase knows that it has already transcribed this part of DNA and does not transcribe it again?

Professor: So, again, the business of knowing is a very tricky question, simply because let us take three genes, Gene 1, Gene 2, Gene 3. These are genes inside your body and they are being transcribed. Let us say Gene 1 is inside your blood cell which is a red blood cell before it loses its nucleus which is not part of what we are talking about. But red blood cells have to make a lot of hemoglobin.

And if Gene 1 is the hemoglobin gene, it basically is, it needs to be in over production mode, 90 percent of the production of protein in that cell will be hemoglobin protein, whereas Gene 2 maybe less important, maybe something involved in endocytic trafficking or signaling from cell to cell and you basically are making about two or three copies a minute and for hemoglobin you are making 20,000 copies a minute.

So, the rate at which different genes are transcribed is very different from cell to cell depending on need. So, effectively what that tells you is for the 20,000 genes, let us say, take only the protein coding genes in your body, each gene is transcribed at a different rate based on its requirement in that cell at that moment. And if you are starving, if you are eaten too much, if you are scared, all these things are going to affect the ability of transcription to happen.

So, this is a highly regulated system and it is not completely clear what decides the level of, what decides the transcription rates. Remember, and I am going to come to this in genetics, you have two copies of every gene. They are called alleles. So, there are two places in your nucleus where the gene can be transcribed not more than that. So, things are dictated not by the number of

copies of DNA. There are only two copies. But by the rates at which polymerases come, sit, transcribe, come, sit, transcribe.

You can imagine that if a polymerase sits and moves ahead 20 bases, a second polymerase can sit which is right behind the first polymerase. So, do not be under the impression that a polymerase come, sits, the entire gene is transcribed, only then a polymerase can sit and start. A polymerase you can have 400 polymerases sitting and transcribing on the same gene, simply because the gene is something like hemoglobin and has to be transcribed at very high levels in a particular cell which is the red blood cell. I hope this answers your question.

Student: Sir, would not that make many copies of the DNA then?

Professor: Transcription is not making copies of DNA. It is making copies of RNA.

Student: Yes sir, I mean, many RNAs.

Professor: So, you will have many more hundreds and thousands and thousands of copies of RNA in your cell right now. They are continuously being made, because your body continuously needs protein. And to make the continuous amounts of protein, you need continuous amounts of RNA. So, transcription is a continuous, never ending process as long as you live.

Student: Sir, how does the DNA bend for the activator proteins to come together with the transcription initiation?

Professor: So, you notice that for DNA polymerase the animations did not show bending. But for RNA polymerase, the animations tend to show a little bit of a kink or bend in DNA and that extrapolation we have made from crystal structures. So, when we make, when we solve crystal structures of RNA polymerase bound to DNA, we do see that the DNA is bent at an angle and that is where that information is coming from. And bending of DNA is accomplished by the proteins which are bound to that DNA and RNA polymerase is a major player in that particular bending.