Transcription section of the central dogma

ASSIGN BUSTER

- Fulya TÜRKER

TRANSCRIPTION: ONE OF THE KEY POINTS OF THE CENTRAL DOGMA

ALTERNATIVE SPLICING

It is known that the central dogma is the most important principle for an organism to carry on its lifespan. In the central dogma, there are 3 major stages: DNA replication and repair, transcription and translation. DNA replication can be simply described as the duplication of DNA. Transcription is the process of conversion DNA to mRNA. Last step is the translation which means the production of polypeptides from the mRNA, it is simply called as protein synthesis. In this essay, transcription section of the central dogma will be explained with detail.

Why transcription is essential for living organisms? Without transcription there is no way to express genes. In order to synthesise protein, at first the mRNA should be formed from a DNA template. Proteins are the functional units in the cells which determine the phenotype of the living organisms. Since the polypeptide chains are formed from the mRNA, there should be a mechanism to create mRNA since it is not found in the cells initially. The proteins that are used in the determination of the phenotype carries the information of the used template DNA strand. DNA sequence is converted to mRNA sequence which then calls the amino acids according to this information. The amino acids form peptide bond with each other and at the end create one, long polypeptide chain.

In order to initiate the process, there should be opening of two DNA strands. Since DNA is found as the double helix in the cells, when they are tightly

bound to each other, RNA polymerase (the enzyme that adds nucleotides in the 3’ end of the newly formed mRNA strands) can’t bind its template and so transcription can’t initiate. So, the cell must find a way to break the hydrogen bonds between the bases of the DNA strands.

In the process of transcription, the essential element is an enzyme, RNA polymerase. What is the significance of this enzyme? AT first, it helps DNA strand to open up for a specific place. The strands are separated from each other and one of them will be selected by polymerase in order to bind and use the information. The newly formed RNA will be carrying the complementary base pairs of the strand that it binds and will have the same sequence with the other strand of the DNA. These are really important points. Of course the new strand will be RNA and eventually will carry uracil instead of thymine. Secondly, in order to elongate the RNA strand, there should be addition of ribonucleotides. During DNA replication, since we are creating new DNA strand, DNA polymerase is responsible for addition of deoxyribonucleotide. However, in transcription we are concerning about RNA strand production. That’s why our enzyme should use ribonuclrotides (adenine, guanine, cytosine, thymine). The hydrogen bonds that are formed between the RNA and DNA strands are very unstable. That’s why RNA polymerase only allows very small length of binding sequence. Also, another important question is that: how the energy for RNA polymerase to move is provided? Our ribonucleotides are carrying three phosphate molecules. That’s why, they are called adenosine triphosphate or guanine etc. When they are added to the newly formed strand, they will release two of the

phosphate in their bases and provide the energy necessary for RNA polymerase.

There are some significant differences between the RNA polymerase and DNA polymerase except that one is taking role in replication the other is the enzyme for transcription. As indicated before, DNA polymerase is taking place in the addition of deoxyribonucleotides in the replicated DNA, while RNA polymerase adds ribonucleotides in the newly transcribed RNA molecule. Also, DNA polymerase needs a primer in order to initiate the transcription, but for RNA polymerase it is not necessary. Lastly, in the replication, the repair mechanism is highly active since we are duplicating the DNA, the errors shouldn’t be tolerated that much, but during the transcription there might be some tolerance to mistakes since after the transcription the deformed or faulty mRNA will be recognized by a mechanism and be degraded.

There are five major classes of RNAs. First one is the mRNA (messenger RNA). It is so much important since it is the template and carries the genetic information for the protein synthesis. In the cell, it can be found in really small amounts. The importance of the function and the amount of it is inversely proportional in this case. Second one is the rRNA (ribosomal RNA). This is the most abundant RNA type in the living organisms’ cells. As the name is indicating, this RNA type is responsible for the structural design of the ribosomes. Ribosome (which is the place where the translation occurs) is mostly composed of rRNAs and the ribosomal proteins. Ribosome formed in the nucleolus by the association of the rRNA and ribosomal proteins. Third one is also really important for the protein synthesis to occur, it is the most

important adaptor in the living organisms' cells. It is tRNA (transfer RNA); include anticodons which recognises the codons on the mRNA and then attach to the corresponding protein and brings the amino acid to the ribosome structure and helps forming of the correctly ordered polypeptide chain. Fourth one is the snRNA that is responsible for the splicing of premature mRNA. The other types of RNAs are providing various benefits to the cell and take role in the cellular activities.

It is known that the elongation of the new strand should be from 5' to 3'. So this newly formed strand should use the template that is moving from 3' to 5'. By pairing with the 3' end, its initial point will be 5' and it will elongate through 3'. According to the direction of the movement of RNA polymerase, the mechanism will decide which strand (bottom or top strand) will be used by looking for the 3' end of the DNA template.

Since there are different kinds of RNAs are formed, there should also be different type of RNA polymerase in eukaryotic cells. RNA polymeraseI is used in the transcription of specific rRNA molecules. RNA polymerase is used in order to produce genes that carries the information for protein synthesis (including mRNA) and also snRNA. Lastly RNA polymerase III transcribes the tRNA molecules, some rRNA and snRNA.

The mechanism of the transcription is a little bit complicated. Because there are lots of additional proteins are involved in the process, without them transcription can't occur in eucaryotes. There will be comparison between prokaryotic and eukaryotic transcription machinery in the following parts of this essay. Let's start with the transcription of the eukaryotes since it is

much more complicated when it is compared to the prokaryotic system. At first, for the initiation of transcription in the eukaryotic gene has a specific sequence which is called as promoter and composed of TATA sequence. This sequence has an essential role in the transcription ÅŸn order to initiate the process. This TATA sequence is around 25 base pairs upstream from the initiator site and RNA polymerase II recognizes this site and can binds there with the help of protein complex which is called as transcription factors . The first general factor that is used in the process is TFIID and contains a specific part that is called TBP (TATA binding protein). TBP will fit the TATA box and activates the addition of other general transcription factor binding. When they are added to the complex RNA polymerase (contains other transcription factor on it) will be able to bind to the start site. Another important factor in the transcription initiation is the activator proteins. Those areTFIIH has a key role in this process since its ability to give the signals for the unwinding of the DNA strands by hydrolysis of the ATP. This ATP usage by TFIIH causes some modification to occur on the RNA polymerase (mostly phosphorylation). This process changes the RNA polymerase’s shape and allows the detachment of the transcription factors from the complex so from now on the initiation of the transcription ends and elongation process is ready to start.

For the transcription initiation to carry on, there are some specific sequence on the genome that are called “ enhancers”. Enhancers have a specific property, they are the site for activator protein binding. Enhancers may be thousands of base pair away from the RNA polymerase binding site however it has a specific ability to bend over and find the RNA polymerase so

eventually let the activator proteins to interact with the other transcriptional factors on the start site or on the RNA polymerase.

In addition to that in order to create a chance for the protein complex to bind to DNA, DNA must be loosely packed. Since, in normal conditions, DNA is found in a very strictly packaged conformation, this must change. Proteins need some sequence to be bound, however if the DNA is packaged strictly in the nucleosome, the transcription factor binding is impossible. So, there are some complex processes that are responsible for the change the packed conformation of the DNA and increase the approachability of the DNA by the transcriptional factor and RNA polymerase. There are mainly two ways to accomplish this aim: chromatin remodelling complex and histone modification. Remodelling complex separates the histones from the DNA strands slightly and the DNA will have a loose conformation. Histone modification is the second way for the increasing of the deforming the packed DNA. Histone acetylation is the best known technique. Histone acetylation causes the histone proteins on the nucleosome to release the DNA slightly and make protein binding to DNA possible.

Second step of the transcription is called the elongation process. There are some elongation factors which provide the attachment of RNA polymerase to DNA throughout the transcription process. Also, they carry out the RNA polymerase and increase its tolerance to the different sequences that should be transcribed. In the elongation stage, the ribonucleotides will be added to the newly formed RNA strand and at the end there will be a termination signal which causes RNA polymerase-DNA interaction breakage and lead to the product which is called precurser mature mRNA (pre-mRNA).

Eukaryotic pre-mRNA needs to be exposed to some modification and of course alternative splicing. Since our genome, most of the eukaryotic organisms' genome is composed of coding (exon) and non-coding (intron) regions, in order to translation process to occur, the pre-mRNA must cleaved from the intron sequences. In addition to that, pre-mRNA needs to be modified and the 2 ends of the pre-mRNA must have some additional feature. This is important because the translational process can't occur without the cap modifications. Those modifications marks the mRNA as a healthy and usable product and also help the mRNA to be transported to the cytoplasm (protein synthesis occurs in the cytoplasm) from the nucleus.

Firstly, let's indicate the cap modifications. In the pre-mRNA, there are 2 caps: 5' cap and 3' cap. 5' cap should be modified by the addition of 7-methylguanosine. This procedure is activated by phosphates enzyme, guanyl transferase (GTP to GMP+PP) and methyl transferase. By addition of methylguanosine, the mRNA product is separated from the other RNA molecules and also mRNA, now, will be able to transported to the cytoplasm. 3' cap is also exposed to additional modification: Poly Adenine tail. At the end of the mRNA product, there will be addition of adenine ribonucleotides and this sequence will prevent the degradation of the mRNA.

However, the most exciting and different process that the eukaryotic pre-mRNA is faced with is the splicing. At the beginning and end of each intron, there are a specific sequences that indicates that the machinery is dealing with an intron. The 5' end of the intron mostly contains GU and 3' end of it contains AG. Also we have specific base in the middle of the intron, Adenine, which is also called as a branch point and gives the signal for 5' end binding

and the formation of the lariat with the help of the snRNPs. What are the snRNPs and what is their role? At first, as it is claimed in the earliest pages of this essay, there is a specific RNA type which is known as snRNA (small nuclear RNA) which are the important factors in the RNA splicing. In the splicing theory, the 5 of them play an active role: U1, U2, U4, U5, U6. Those particles recognise the exon and intron end and start points and can distinguish them so help the splicing process a lot. Each of them contain at least seven proteins and form snRNPs (small nuclear ribonucleoprotein) which afterwards creates a structure that is called a “ spliceosome”. Now, in the following section each snRNP that is actively join the splicing process will be explained.

At first, there is a BBP protein which binds to the branch point (mostly Adenine). U2 snRNP recognizes BBP binding and replaces this protein with itself and form interaction with the branch point. U2 pushes A to the outside of the sequence and allows the attack of the 5’ site to this specific base. U1 initially recognises the 5’ end of the intron. U4-U5-U6 joins the process as a triplet. U4 and U6 is dissociated from each other and U6 removes the U1 snRNP and sits onto the 5’ end. U6 and branch point interacts and come closer (the process is called first phosphoryl transfer-reaction). At the end of this process lariat formation occurs. Lastly, U5 causes exon-exon interactions and second phosphoryl-transfer reaction takes place. At the end of this process the RNA is spliced and the mature mRNA is formed. In mature mRNA, there is 5’ cap, 3’ poly a tail and no introns. From now on, this mRNA is ready to synthesise the protein (the process of translation).

Also, there is a theory of alternative splicing which should be mentioned in the discussion about the RNA splicing subject. After the removal of the introns, some exon can leave the sequence and cause alternative sequences. In this process, no shuffling of exons is allowed, but some of them leave the track. The first exon, which carries the start codon AUG, can’t change. It must be always found in the first position of the mature mRNA. The last exon of the spliced RNA must also contain one of the stop codons (UAA, UAG, UGA). But, other exon sequences are allowed to change without shuffling. Exon orders must be preserved. The below diagram indicates the process of the alternative splicing:

Lastly, it is important to mention about the differences between the eukaryotic and prokaryotic transcription. In prokaryotes, we have a special term for the DNA which is called an operon. Operon carries the information for more than one gene and transcribed by the single promoter, eventually create a single mRNA which contain more than one gene. This single mRNA will be translated and eventually produce more than one protein, this characteristic of the prokaryotic DNA is called as “ polycistronic” gene. However, in eucaryotes, there is only one gene that is transcribed at once. That’s why eukaryotic organisms are called as “ monocistronic”.

Prokaryotic transcription is only dependent to one factor in order to hold the RNA polymerase on the DNA while the eukaryotic transcription needs so many transcription factor, additional proteins and mediators. Also, since the prokaryotic DNA is found in a

loose conformation in its original form, there is no need to use additional modification to destroy the packed structure as in the eukaryotes. So prokaryotic transcription machinery doesn’t use any chromatin remodelling complex or histone modifications. Prokaryotes don’t have intron in their pre-mRNA so there is no splicing in their mRNA after it is synthesised. They are free from introns, their mRNA is composed of more than one gene.

Lastly, prokaryotes don’t need additional cap modification after the transcription of the mRNA. the mRNA can be easily transcribed as soon as they are synthesised.

References:

Belfort, M. (1989). RNA: Catalysis, splicing, evolution. Amsterdam: Elsevier .

Belotserkovskaya, R., Saunders, A., Lis, J., & Reinberg, D. (n. d.). Transcription through chromatin: Understanding a complex FACT. Biochimica Et Biophysica Acta (BBA) – Gene Structure and Expression, 87-99.

Dahlberg, J. (1989). RNA processing. San Diego: Academic Press.

Darnell, J., & Lodish, H. (1990). Molecular cell biology (2nd ed.). New York: Scientific American Books :.

Elliott, D., & Ladomery, M. (2011). Molecular biology of RNA. Oxford: Oxford University Press.

Goodrich, J. A.; Tjian, R. (1994). “ Transcription factors IIE and IIH and ATP hydrolysis direct promoter clearance by RNA polymerase II”. *Cell* .

Latchman, D. (2004). Eukaryotic transcription factors (4th ed.). Amsterdam: Elsevier/Academic Press.

Locker, J. (2001). Transcription factors. Oxford: BIOS ;.

Ohyama, T. (2005). DNA conformation and transcription. Georgetown, Tex.: Landes Bioscience/Eurekah. com ;.

Raven, Peter H. (2011). *Biology* (9th ed.). New York: McGraw-Hill.

Stamm, S. (2012). Alternative pre-mRNA splicing theory and protocols. Weinheim: Wiley-Blackwell.

Wagner, R. (2000). Transcription regulation in prokaryotes. Oxford: Oxford University Press.