Cell Biology: Cellular Organization, Division and Processes

Prof. Shikha Laloraya
Department of Biochemistry
Indian Institute of Science – Bangalore

Lecture – 4
DNA: The Genetic Material, Part II

Hello everyone. In this lecture we will discuss how genetic information in DNA is decoded.

(refer time: 00:30)

As you know, base pairing in DNA is specific. However, there is no constraint on the sequence of the successive or adjacent bases along one DNA strand, as can be seen here from the sequencing output in the background. In fact, this sequence of bases constitutes the genetic information; there can be any combination of successive basis forming the sequence. The sequence of bases corresponds to the genetic information and it codes for other macromolecules, such as RNA and proteins.

(refer time: 01:13)

So, DNA stores genetic information within the sequence of its nucleotides. We know that the gene is a basic unit of heredity, and now we know that it is a sequence of nucleotides, encoding a product such as an RNA or a protein. But earlier, although it was known that DNA is the genetic material, how it codes for information was not understood. One of the first evidence for information flow from genes to amino acid sequences in a polypeptide, came from studies on the disease sickle cell anemia, which is an inherited disease.

So, sickle cell anemia is linked to the S-allele of the beta globin gene. It was found that in the protein there is a glutamic acid to a valine substitution. Therefore, it was likely that the S allele was coding for this altered protein sequence and this implied that there is a flow of information from DNA to the protein. So how would this happen? DNA can't be a direct template as it is removed from the site of protein synthesis in cells. Protein synthesis occurs in the cytoplasm as you know, and DNA is in the nucleus, in eukaryotic cells. So, it is suggested to scientists that there must be some other molecule that conveys genetic information, which is encoded in DNA, from the nucleus to the cytoplasm.

(refer time: 03:04)

It seemed likely that the molecule which was conveying genetic information from DNA to protein is RNA or ribonucleic acid. The reasons for this were that this is just another class of nucleic acid, which resembles DNA to some extent, and it was possible to imagine therefore, DNA serving as a template for its synthesis. Also RNA was known to be abundant in the cytoplasm.

RNA or ribonucleic acid is an unbranched polymer of ribonucleotides A,U,G or C, which are linked by 3 prime - 5 prime phosphodiester linkages. So, in contrast with DNA, it contains a ribose sugar that has a hydroxyl group at C2 prime. It lacks thymine but instead it has Uracil, which is quite similar to thymine except that it lacks the methyl group at the 5 position in the pyrimidine ring, which is replaced in this case by a hydrogen.

In the cell, RNA is usually present as a single stranded molecule as shown in this animation here. Double stranded helixes can be formed intramolecularly by pairing of bases within different regions of the same RNA molecule.

(refer time: 04:44)

So, in 1953 after the structure of DNA was solved and a copying mechanism for genetic information was evident from the proposed model, it was hypothesized that chromosomal or genomic DNA, it functions as a template, not only for its own synthesis, as was proposed by Watson and Crick, but also for RNA molecules that then move to the cytoplasm where they somehow determine the amino acid sequence present in proteins.

This pathway of flow of information was referred to at that time, as the central dogma, by Francis Crick. So, the process of RNA synthesis from DNA was one of the aspects of the central dogma and this process was referred to as transcription, and RNA template directed protein synthesis was referred to as translation. Now, these arrows are unidirectional, and this implies that according to this theory, the information flow in this direction occurs, and it cannot proceed in the reverse direction, which would mean that the protein sequence does not direct or template RNA synthesis and likewise, neither the RNA could be a template for DNA, as per the central dogma.

(refer time: 06:30)

So, how would RNA code be a template for protein synthesis? This was a question that was on the minds of many of the scientists and based on structural considerations alone it seemed unlikely that RNA would be able to distinguish between very similar amino acids for it to be a template for protein synthesis. So, Crick came up with the adapter hypothesis, that is, he hypothesized that an amino acid must be first attached to specific adapters that recognize the base sequences in the messenger RNA and he thought that they might be small RNAs, but did not have evidence for this. But later on another scientist Paul C. Zamecnik, while studying in vitro protein synthesis requirements, found that amino acids are first attached to transfer RNA molecules, before being incorporated into polypeptides in the in vitro translation reactions.

A mature transfer RNA is shown here; it is a short sequence of bases and particularly note this region here, the anticodon, that binds to bases in messenger RNA to specify amino acid sequence. So a set of 3 bases, which specify the amino acid that should be added, which is referred to as a triplet code and this part of the mRNA is referred to as a codon. The full set of codons corresponding to the amino acids or in other words, the genetic code was determined by the efforts of several other scientists at that time: Nirenberg, and Matthaei, and also, Har Gobind Khorana who was working then at the University of Wisconsin-Madison. So, in this way, the transfer of information from mRNA to protein could be understood.

(refer time: 08:57)

However, later on, one aspect of the central dogma was modified by the discovery of reverse transcriptase, an enzyme that makes DNA from RNA. This discovery was made by Howard Temin who was a professor of Oncology at the McArdle lab for cancer research, where I also did my PhD incidentally. So, he studied an RNA tumor virus, Rous Sarcoma Virus. As you may know, many viruses have RNA genomes present within the virions. These RNA viruses, they are also common in plants, and people studying Tobacco Mosaic Virus found that purified RNA from TMV could infect host cells and it could produce infectious virus particles. So, these viruses coded for an enzyme that could synthesize RNA from an RNA template, using mechanisms, which are more or less similar to transcription. However, this mechanism of RNA-directed RNA synthesis could not explain the replication of the family of animal RNA viruses that Temin was studying, the RNA tumor viruses.

RNA tumor viruses cause cancer in infected animals. For example, the Rous Sarcoma Virus could also in fact, transform normal cells to cancer cells in vitro and this provided a nice quantitative assay to study transformation in vitro at that time. It was observed that the characteristics of transformed cells were determined by genetic information in the viral genome, but the information was transmitted to progeny cells, even in the absence of viral replication.

So this implied that the viral genome was stably inherited, and this form of the virus was referred to as a provirus by Temin. He observed that Actinomycin-D, which is known to be an inhibitor of DNA directed RNA synthesis, or transcription, could inhibit virus production by infected cells, suggesting that the provirus was in a DNA form. However, the virus itself was an RNA virus.

And he also observed that the early stages of cell infection by Rous Sarcoma Virus could be blocked by inhibitors of DNA synthesis. Hence, it was inferred that DNA synthesis was required early, and DNA-directed RNA synthesis was important late in the life cycle of the Rous Sarcoma Virus. So based on these observations, Howard Temin proposed the DNA provirus hypothesis, that is the provirus is a DNA copy of the viral genome.

Now this was somewhat heretical, because this would involve the synthesis of DNA from RNA, which was against the central dogma of molecular biology that was popular at that time. Hence, his idea was met with disbelief and sometimes even ridiculed by other scientists.

(refer time: 12: 53)

But he continued and he persevered trying to find more evidence in support of his theory, while he was working at the McArdle lab for Cancer Research at UW-Madison. Incidentally, he was also a member of the committee for my qualifying preliminary exam for PhD candidacy, and I remember him very well, as he was quite encouraging towards students. So, while clear evidence for integration of the viral DNA in the host genome could not be obtained at that time, although now it is pretty easy, but due to technical limitations at the time they were doing this work, Temin and Mizutani obtained direct evidence for their idea by discovering a viral enzyme that could synthesize DNA from RNA. This enzyme is now known as reverse transcriptase, and this discovery had revolutionized molecular biology research. His ideas also advanced the understanding of the retroviral life cycle that was immensely useful in rapidly understanding the lifecycle of another retrovirus, HIV, when it emerged as a very tough human pathogen.

(refer time: 14:22)

So, his work is a great example I should say, of the importance of curiosity driven basic research that often leads to big discoveries of biomedical significance. Temin was awarded a Nobel Prize for the discovery of reverse transcriptase, along with David Baltimore who simultaneously discovered this enzyme. So as per the provirus hypothesis, viral RNA is copied into viral DNA, and this viral DNA then integrates into the host genome and it is stably inherited. It transcribes viral RNAs that are translated and then they are also packaged along with the virion proteins to release new virus particles.

The discovery of reverse transcriptase enzyme was indeed an unambiguous proof in support of the provirus hypothesis and thus a key aspect of the central dogma had to be revised, based on these findings.

(refer time: 15:40)

So, it was now understood that not only RNA viruses could copy their RNA, but some of the RNA viruses could also make DNA from RNA using reverse transcriptase. Hence, the information could flow from RNA to DNA as well.

(refer time: 16:13)

This slide shows the enzymatic synthesis of messenger RNA from DNA, which is catalyzed by RNA polymerase enzyme; the nucleotides ATP, GTP, CTP, UTP, are used as precursors. Now, in the DNA there are 2 strands but only one of the 2 strands is used as the template, and the addition of nucleotides occurs in this region where the 2 strands have separated, forming kind of a transcriptional bubble. The synthesis of RNA proceeds in the 5 prime to 3 prime direction.

(refer time: 17:02)

In eukaryotes, messenger RNA is synthesized in the nucleus and it moves to the cytoplasm where its message is translated by the ribosomes. So this in fact, was first demonstrated by a pulse-chase experiment by Prescott, using tritiated cytidine in *Tetrahymena* cells that were exposed to tritiated cytidine for a few minutes, 15 minutes or so, and chased with cold cytidine. It was seen that the label could be detected inside the nucleus at about 15 minutes but after an hour of chase, using the cold cytidine, the label had moved to the cytoplasm and this demonstrated that transport of mRNA occurs from the nucleus to the cytoplasm.

The enzyme RNA polymerase is required for transcription; it binds at specific sites known as promoters. Various other factors are also required for this highly complex process, such as transcription factors, enhancers, which are cis-acting activating sequences, and binding factors, which are activators, and many other factors, in fact.

(refer time: 18:36)

So, general transcription factors, they are class of proteins that bind to a promoter on DNA and they activate its transcription. Enhancers of transcription are regulatory sequences that can be located far away from the gene as well, and they activate transcription several fold. Without enhancers the expression level remains only basal. The enhancer region in the DNA, it recruits activators, which are responsible for stabilizing promoter binding through DNA looping, and interactions with a mediator.

Shown here is the RNA polymerase and it is got a C-terminal domain that is quite important for its function. So the unphosphorylated form of the CTD -C terminal domain, of RNA Polymerase 2 binds to mediators. Whereas its phosphorylation can release the mediator and stimulate promoter escape of the enzyme. Phosphorylation of CTD also helps in elongation, and these elongation factors prefer binding to the phosphorylated form of the CTD.

The C terminal domain, it is quite long and it lies near the channel through which the newly synthesized RNA exits the enzyme, and it recruits elongation factors and processing machinery and it brings them to the newly made RNA. So, phosphorylation of the CTD occurs by TFIIH at serine 5 and this leads to mediator dissociation from the polymerase during initiation, and promoter escape of the enzyme. Kinase P-TEFb, which is recruited by transcription activators, phosphorylates the serine at position 2 of CTD repeats, and this correlates with elongation.

(refer time: 21:10)

This animation provided by Drew Berry and Etsuko Uno shows the process of transcription by RNA polymerase or Pol2. During initiation Pol2 together with initiation factors binds a promoter. Various other factors also bind with the complex for example enhancers can bring about activation by looping. Phosphorylation of the C-terminal domain causes the escape of the polymerase from the promoter. The CTD frees itself from other proteins of the promoter and it escapes and enters into the elongation phase. You can see Pol2 zipping along the DNA transcribing RNA in yellow as it goes along. This happens in the complex milieu of the nucleus, on open chromatin. Other regions may have nucleosomes and a folded chromatin organization. Pol2 uses NTPs as precursors and efficiently elongates the transcript. After the elongation is complete, the mRNA will be released.

The transcript also undergoes some post transcriptional processing such as capping, splicing and polyadenylation, which are not shown here. Upon termination of transcription, the mRNA is released and eventually it is translocated into the cytoplasm from the nucleus.

(refer time: 22:35)

This slide shows some of the basic requirements for translation, that is a formation of protein from RNA. So these are of course, messenger RNA, which is shown here it is single stranded, its got a 5 prime cap and a 3 prime poly-A tail, and of course the main region of the transcript, which has the coding information; it also requires ribosome. The ribosome has 2 subunits, a large and a small subunit. It also requires another type of RNA, transfer RNA or tRNA, and an Amino acyl tRNA synthetase enzyme that charges this tRNA with a specific amino acid, that is it recognizes the specific transfer RNA to which a specific amino acid has to be attached and it brings about the covalent binding of the amino acid to the 3 prime end of this transfer RNA. So simply for the process of translation, ribosomes associate with the mRNA and transfer RNAs bring the amino acids to this site, where they are read by the anticodon in the transfer RNA. The code is read by the anticodon present in the transfer RNA and this ensures the correct sequence of amino acids as specified by the genetic code, is linked up to form the polypeptide shown here.

(refer time: 24:22)

Now shown here is the structure of a transfer RNA in the open clover leaf form that is a consequence of complementary base pairing between different parts of the molecule. Transfer RNAs are quite

small; they are generally 70 to 80 nucleotides long and to function as an adapter two parts of the transfer RNA are quite important. So these are: one is the sequence CCA which is present at the 3 prime end, amino acids are covalently attached to the ribose of the terminal adenine present here. And the other important region is the anticodon loop at the other end; this associates with the cognate codon by complementary base pairing. Amino acyl tRNA synthetases shown here are enzymes that are important for the attachment of the amino acids to specific transfer RNAs. So, each of these recognizes a single amino acid specifically, as well as the correct transfer RNA to which it should be attached. So, the enzyme recognizes specific nucleotide sequences including the anticodon region unique to each species of the tRNA and of course, this is very important if the fidelity has to be maintained. tRNAs also have modified bases, which are not shown in this cartoon, but there are several modified bases present in transfer RNA such as methyl guanosine and methyl cytosine, dihydrouridine, ribothymidine, pseudouridine etc.

And in this interesting animation of a 3-dimensional structure of a transfer RNA you can see that the transfer RNA folds into a compact L-shaped structure and this was demonstrated by X-ray crystallography. So, this folded shape is important for transfer RNAs for fitting into the ribosome during the process of translation; you will have seen in one of the earlier animations that most of the transfer RNAs appear to be depicted in this type of shape. Here in this animation, the anticodon is shown in yellow over here, and the amino acid group is attached to the 3 prime end, it is shown in blue in this slide, come around- Yeah.

(refer time: 27: 18)

The basic steps in protein translation are shown in this slide. Ribosomes are the sites of protein synthesis and they have 2 subunits: large and small. These are 50 and 30S in case of prokaryotes and 60 and 40S in case of eukaryotes. Each of the subunits consists of various ribosomal proteins and ribosomal RNAs. Ribosomes are very abundant in cells. *E.coli* has about 20,000 ribosomes, which is about 25% of the dry weight, while rapidly growing mammalian cells have about 10 million. Translation starts with methionine at the AUG codon. The process of translation is actually much more complex than explained in this slide and the description I am giving in today's introductory lecture. For example, not shown here are initiation factors, which are important for the initiation of translation.

The ribosome has got three tRNA binding sites, the P site, the A site and the E site. P for peptidyl, A for amino acyl and E for exit. These sites are located between the entry channel for the mRNA and this channel is wide enough for entry only of the unpaired mRNA to pass and enter into the decoding centre where it is not engaged in any other interactions which will prevent it from being read.

First the initiator methionine tRNA is positioned at the P site, the one in the middle, and the A site is vacant. The second amino acyl tRNA is brought into the A site by EF1 alpha and the tRNA is inserted into the A site coupled with GTP hydrolysis. Now the peptide bond is formed, resulting in the transfer of the methionine to the acyl tRNA in the A site. The ribosome then moves 3 nucleotides along the mRNA in which it is held by EF2, also coupled to GTP hydrolysis. This translocation positions the peptidyl tRNA in the P site and the uncharged tRNA is now in the exit site, which is then released. So, now the A site is again vacant and a new amino acyl tRNA can enter there and the process can be repeated for the addition of the next amino acid. So, this process is continued elongating the polypeptide chain until the stop codon enters into the A site.

So, there are protein release factors or RFs, which recognize the stop codon by a tripeptide motif referred to as a peptide anticodon. In addition, there is a GGQ motif in the release factor. This triggers the hydrolysis of the protein and tRNA bond. So, for example, when RF 2 binds to the ribosome with a stop codon, at the A site, it undergoes a conformational change such that the domain with the conserved peptide anticodon interacts with the mRNA at the decoding centre of the ribosome, and the GGQ containing region comes in contact with the peptidyl transferase centre and then it triggers the hydrolysis of the peptidyl tRNA linkage. So, in this way the polypeptide chain can be released, and translation is terminated.

Now it is also interesting to note that the ribosome is a ribozyme, that is the peptidyl transferase reaction is catalyzed at a site in the ribosome where ribosomal proteins are notably absent, as per the X-ray crystal structure. And also rRNA and the P site tRNA contribute to the catalysis because mutations that remove the 2 prime OH of the adenine residue at the 3 prime end of the P-site tRNA, reduce the catalysis rate by (10^6) 10 to the 6th fold. So, this is an interesting example of substrate assisted catalysis.

(refer time: 32:34)

This is an animation of the next step in protein synthesis after the mRNA has translocated out of the nucleus and it reaches the cytoplasm, it associates with ribosomes shown here in blue, and here you can see the entry and the exit channel of the mRNA and transfer RNAs floating around. So, you can see the enzyme, ribosome which actually is a ribozyme, doing its business and here is a close up view of the entry channel. So you can see the RNA entering in the single stranded form. This is a view of the inside of the ribosome showing the formation of the polypeptide and the recognition of the codons in the mRNA. And now, you can see sort of a close up view of the codon - anticodon recognition site. So, the transfer RNAs are of different types and they have different amino acids which are attached to them. And they find their way to the ribosome you can see here an enlarged version of one of the tRNAs with this pink amino acid. Here is a view of the site where the polypeptide chain exits the ribosome. So, as the polypeptide chain is being elongated inside it comes out of this channel and this process goes on, and as it goes on and the chain becomes longer, it might also start folding, it may undergo some level of secondary folding because you know that is dependent upon its primary sequence of the polypeptide. Once the process of translation is completed, termination of translation will happen. This polypeptide is released and it forms a folded protein. After this it may undergo further folding or post translational modifications and it will be transported to its correct destination within the cell. Some of these processes may even happen co-translationally.

(refer time: 35:23)

So, recall that prokaryotes do not have a nucleus. So far we have been discussing how the mRNA is formed and it comes out of the nucleus and meets up with ribosomes and translation ensues. But, we have discussed earlier another class of organisms the prokaryotes, that they do not have a nucleus. okay. So, there is no compartmentalization between the DNA and where the ribosomes are present.

So, here are some interesting points I thought should be discussed about this aspect of translation in prokaryotes considering the lack of compartmentalization in prokaryotes. So what happens is that in

prokaryotes, the ribosomes start building proteins as the mRNA is being transcribed, okay, and transcription elongation factors, they actually link the RNA polymerase to the ribosome, holding it close but not too close. So they give some space for the movements of the ribosome, which are required for it to translate. The RNA polymerase and the ribosome and all of the other associated molecules together form an expressome, or a transcription translation complex, or TTC. This is shown here, and it is actually quite fascinating: here is the entire central dogma being depicted within one structure!

So here in this structure, the green is the RNA polymerase, the DNA is shown in yellow where it has come apart forming the bubble and the messenger RNA is shown in red, it is being produced. And here it is being translated: so this blue blob is the ribosome both the subunits in 2 different shades. You can see the transfer RNA in magenta and it is got this little white dot, which is actually a phenyl alanine amino acid which is attached there. So, this is quite fascinating actually so I thought I should bring this up and I strongly suggest you to visit this link which is provided here to get more details regarding the structural aspects of co-transcriptional translation, which goes on in prokaryotes. Thank you.