Basics of Biology
Professor Doctor Vishal Trivedi.
Department of Biosciences and Bioengineering,
Indian Institute of Technology, Guwahati
Lecture - 29
Translation (Part-I)

Hello everyone. This is Dr. Vishal Trivedi from Department of Biosciences and Bioengineering, IIT, Guwahati. And what we were discussing? We were discussing about the different properties of the living organisms, and in this context, so far what we have discussed, we have discussed about the classifications, evolutions, we have discussed about the prokaryotic as well as the eukaryotic cells.

And in the previous module, we have also discussed about the different types of biomolecules. And in the current module, we are discussing about the central dogma of life or the central dogma of molecular biology. And in that context so far what we have discussed, we have discussed about the DNA dependent, DNA synthesis and that process is called as the replications.

And while we were discussing about the replications, we have discussed about the different steps and how the replication as it is been responsible for the development of a technique which is called as the polymerase chain reactions. And then we have also discussed about the different types of applications of the polymerase chain reactions.

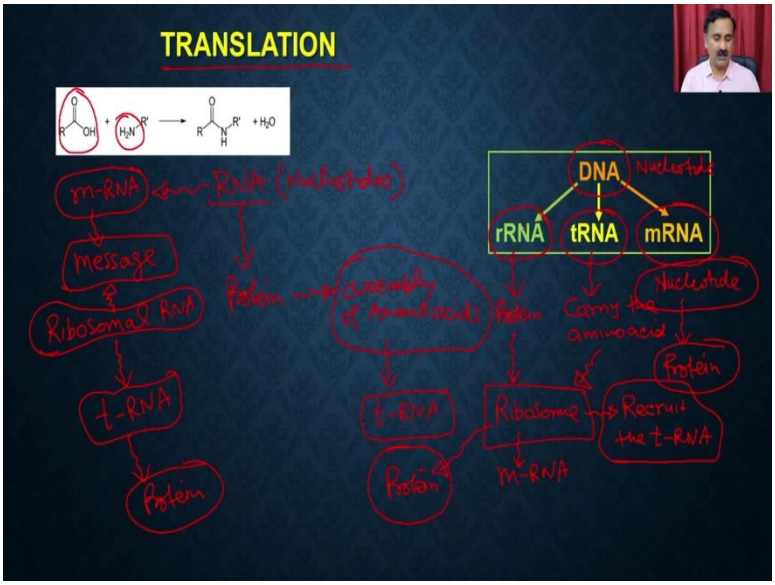
In the previous lectures, we have discussed about the transcriptions, so DNA dependent RNA synthesis and that is being catalysed by the RNA polymerase and this process is known as the transcription. And we have discussed about the detailed steps, what are involved in the transcription of the prokaryotes or the transcription of eukaryotes.

Subsequent to that, we have also discussed about the post transcriptional modifications in the all the RNA species, we discussed about the capping, we have discussed about the tailing, and we have discussed about the splicing in the messenger RNA, and then we also discussed about the different types of post-transcriptional modifications into the RNA and as well as in the ribosomal RNA.

So, in the today's lecture, we are going to discuss about the third topic of the central dogma of life or the central dogma of molecular biology, and that process is known as the RNA dependent protein synthesis or in general it is called as the protein synthesis. And this process

is called as translations. So, let us start discussing about the translations, and we are also going to discuss about the post-translational modifications into the protein as well.

(Refer Slide Time: 03:33)



So, as the name suggests the translation means that you are actually going to use the information from the one individual or one source and then you are actually going to translate into the new form. So, in this case, we were having the primary information in the form of DNA and that primary information of, that was stored in the form of nucleotides.

And this primary information of the nucleotide is been then transcribed into the secondary information into in the synthesis of the different types of RNA molecules like the ribosomal RNA molecules, tRNA molecules, and the messenger RNA molecules. Now, these molecules are also having this transcribed information in the form of the nucleotides.

And now what we have to do is, we have to take the information from this particular nucleotide sequences what are present in the messenger RNA and then we have to synthesize the proteins. So, as you recall, when we were discussing about the proteins, the proteins are the polymer of the amino acids where the amino acids are linked to each other by a peptide linkage.

So, what you can see here is that, if you have two amino acids, like the amino acid which has the the carboxyl group and the amino acid which has the amino group so when these two groups are being combined they will actually going to form, or they are going to be combined

by a peptide linkage and this is a process which are actually going to be achieved during the translation.

So, what translation mean is that you are actually going to take the information from the RNA and this RNA, this information in the RNA is going to be in the form of nucleotides and this nucleotide information is going to be interpreted to synthesize the protein by the assembling, the different types of amino acids. So, protein is going to be synthesized by the assembly of, or by the combination of the amino acids.

So, what you see here is that you are actually going to use the very sophisticated machine, so that you can be able to interpret the information what is given on to the RNA molecules. So, in this case the RNA what we are going to use for that purpose is called as the messenger RNA and as the name suggests. It is called as messenger RNA because it is con, it is actually securing a message that message has to be interpret by the machinery and this machinery is then going to utilize that to assemble the amino acids.

So, you see that the translation is actually a multi-step process, where you require a machinery, so that it can actually be able to interpret the information what is given in the messenger RNA. So, messenger RNA has the message, this message is actually telling you that in what sequence, you are actually going to provide or you are going to assemble the amino acids, but that message has to be interpreted by the machinery.

So, you require a system, so that you can be able to interpret this message and that things comes from the ribosomal RNA. So, you are actually going to use the ribosomal RNA to interpret this message, and then in by interpreting this message it is actually going to assemble the amino acids.

And that assembly of the amino acid, in what sequence the assembly of the amino acids are going to be performed that is going to be done by the transfer RNA, which means it is actually the RNA which is going to transfer the amino acids as per the message what has been given on to the messenger RNA and the ribosomal RNA is also going to take part.

So, ribosomal RNA is going to be on the interface, where on one site it is actually going to read the message on to the messenger RNA and then on the other hand it is actually going to assemble the tRNA into a sequence, so that you are ultimately going to have the protein

molecules, or you are also going to have the assembly of these amino acids in a particular sequence and it is going to be responsible for the synthesis of protein molecules.

So, what you see here is that from the DNA you are going to have the 3 different types of RNA molecules, ribosomal RNA molecules, tRNA molecules and messenger RNA molecules and all these RNA molecules are going to have their own discrete role in the in the process of translations. So, a ribosomal RNA is actually going to assemble with the protein and that is how the ribosomal RNA is going to be responsible for the synthesis of the ribosome.

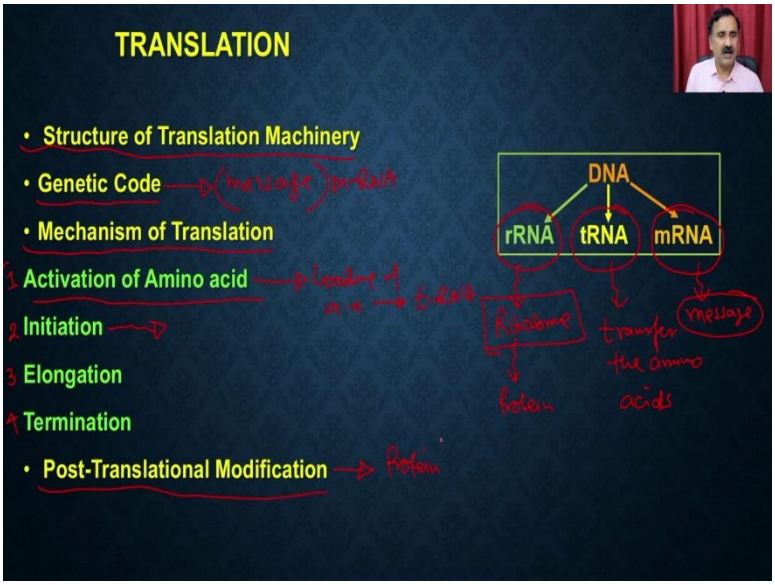
And the ribosome is actually going to interpret the information what is what is given on to the messenger RNA. And on the other hand, and on the other hand it is actually going to recruit the tRNA which is going to be take up the amino acids and that is how the ribosomal RNA is actually ultimately going to synthesize the proteins.

What you see here is transfer RNA, so transfer RNA as the name suggests transfer RNA is going to transfer the amino acids. So that is why it is actually going to have the it is going to carry the amino acids and it is going to carry and transfer that amino acids as per the instruction or as per the information provided onto the ribosomes.

Then messenger RNA, messenger RNA as the name suggests it is going to have the messages and that messages are in the form of nucleotide and that message is going to be interpreted by the ribosomal machinery and ultimately it is going to give you the synthesis of the proteins.

So, if you would like to understand the whole process of translations, what we have to understand is that we have to first understand about the machinery, so machinery which includes the ribosomal RNA, tRNA, and messenger RNA, and then we have to understand the different types of processes. So, let us see what are the different machinery what has been involved into the process of translations.

(Refer Slide Time: 10:18)



So, as far as the machinery is concerned, we have the machinery like the different types of RNA molecules like the ribosomal RNA molecules, tRNA molecules, and messenger RNA molecule. As I already explained in the previous slide that the ribosomal RNA is going to be responsible for the synthesis of the ribosome and that ribosome is actually going to be the central machinery what is going to be responsible for the synthesis of the proteins.

Transfer RNA is actually going to transfer or it is going to carry the amino acid, so it is going to carry the amino acids and it is going to transfer the amino acids, as per the information given on the messenger RNA and as well as, as per the instructions from the ribosome. Messenger RNA, it is actually going to contain the message and that message is going to be interpreted by the transfer RNA as well as the ribosomal RNA and then third, the ribosome is going to give you the protein.

Apart from the, so we are going to understand the structure of the some of these RNA molecules and then we also going to understand about the genetic code. So genetic code is actually nothing but that message which are actually going to be present on to the messenger RNA, so that also is very important to understand. And then we are also going to discuss about the mechanism of translations.

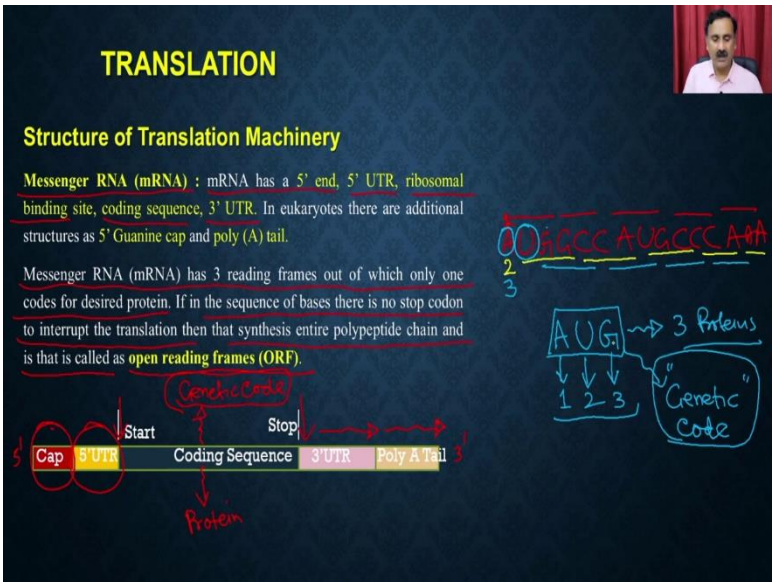
In the mechanism of translation the first step is that we are going to understand about the activation of the amino acids, which means we are going to understand about the loading of the messenger, loading of the amino acids onto the tRNA. And once you are actually going to

load the amino acid onto the tRNA, you are actually committing that amino acid for the protein synthesis.

Then we also going to discuss about the initiation, so just like as we remember that other two processes like the replication or the transcription which also have the similar kind of steps like initiation, elongation and terminations. Here, also we are going to have the process of initiation elongation and terminations, so we are going to have the initiation and then we are going to have the elongation and terminations.

And ultimately, we are also going to discuss about the post translational modifications into the protein which are actually going to have the very significant role in terms of the many of the biological function associated with that particular protein. So, let us start discussing about the machinery, so we are going to start discussing first with the messenger RNA.

(Refer Slide Time: 12:55)



So, as the name suggests the messenger RNA or the mRNA has a 5 prime end, 5 prime UTR and the ribosomal binding site, then it also has the coding sequence and it has the 3 prime UTR. So, what you see here is that the messenger RNA is going to have this is the 5 prime end and this is the 3 prime end. So, on the 5 prime end it is going to have a cap structures.

And next to the cap it is going to have the 5 prime UTR, then it is going to have the translational start site and then it was going to going to have the translational stop site and the nucleotides what are present between the start and the stop site are going to be called as the

coding sequence. So, these are the folding sequence which are actually going to direct the protein synthesis machinery like the ribosomes or the tRNA to synthesize the protein.

So and that is being done because it is actually going to contain the genetic code, and these genetic codes are actually going to provide the informations in what sequence you are actually going to assemble the amino acids and that is how it is actually going to be give you the correct proteins.

So, in this, is the in, so and then we have so messenger RNA has the 3 reading frames out of which only one codes for the desired proteins, if in the sequence of these bases there is a there is no stop codon to interpret the translation then the synthesis of the entire polypeptide chain and that is called as the open reading frame, so this is also called as the coding sequence or the open reading frames.

Then next to that stop codon, you are going to have 3 prime UTR and then you are also going to have the poly (A) tail. So, this poly (A) tail I think remember, when we were discussing about the transcription, we said that that you require a cap and you require a poly (A) tail also for the proper functioning of the messenger RNA, both the poly (A) tail and as well as the cap is actually required for the for the messenger RNA to bind to the ribosome and that is how they are actually going to participate into the protein synthesis.

Now, what it is mean is that, when you are actually going to have the coding sequence the coding sequence could actually be able to interpret into the 3 reading frames. So, let us take an example how it is actually going to be done. So, for example, if you have a coding sequence like this, you have a coding sequence like AUGGCGAUGCCCAAA.

Now, if this coding sequence can be interpreted in the 3 reading frames. So, number one reading frames, you can start from the one, and you can have this reading frame, you can have like this, so this is the reading frame number one. Now, if I change the colour, so you can have the reading frame number 2, so reading frame number two is going to be like this.

So it is going to left the first residues and it is going to have the reading frame number two which is UGC, then CCA, then UGC then CCA and A, so these are not going to be part of the reading frame because you only have two nucleotides. And this is all we are going to discuss in that. And then you are going to have the third reading frame, so that is also going to be there, so you are going to have the third reading frame.

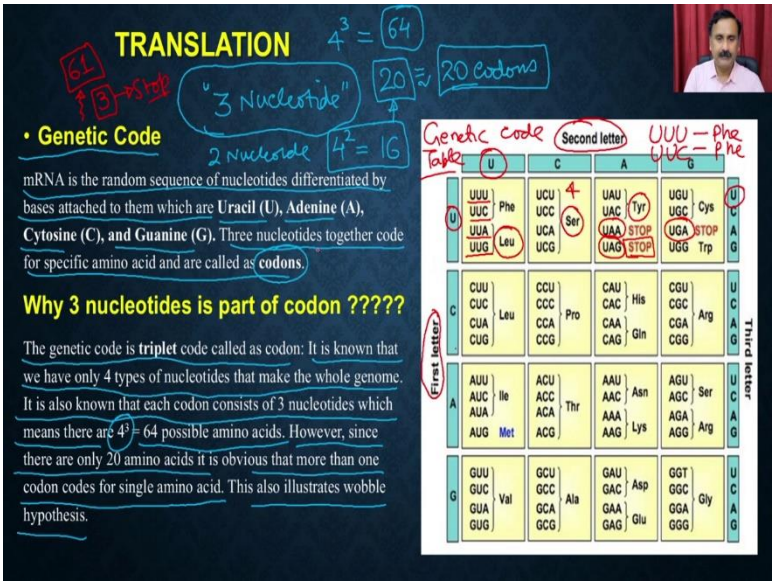
So, what is going to be, so first we just started with this residue, the second one we started with this residue, and the third one we are going to start with this residue, so it is going to be like this. Will you see one a is left, so this is the what happened, so this means if the 3 reading frames are actually can start from, so first reading frame is going to start from A, the second reading frame is going to start from U, third reading frame is going to start from G.

So, that is why technically, a single RNA molecule or single messenger RNA molecule is potentially be able to give you the 3 proteins, but that the, when we started with the reading frame number one the everything was settled, so you are getting the all the nuclear all these codons are being used, so that is why that is actually going to give you a protein, whereas when you go with the reading frame number two what you see here is first is, first nucleotide you have left.

And there are two nucleotides which are left at the back, so there you there will be a truncated protein which is going to be synthesized, and as it is as you can see that the residues, the nucleotides what we are using for the synthesis is also going to be different, so that is why, the different types of proteins can be synthesized if you go with the reading frame number 1, 2 and 3.

And ideally every gene is actually defined that which reading frame they are actually going to use, so and that is how they are actually been able to provide the correct or they will be able to help in synthesis of the correct proteins. This combination like this combination of triplet is actually been called as genetic code because it actually is going to give you the information about the which amino acid is it is actually going to code and that is how it is going to be called as genetic code. So, let us discuss about the genetic code.

(Refer Slide Time: 18:42)



So, genetic code is actually the messenger RNA, is having the random sequence of the nucleotide differentiated by the bases attached to them which are uracil, adenine, cytosine and guanine and the 3 nucleotides together code for a specific amino acid and that are called as the codon. Now, the question comes why there are only 3 nucleotides, why not 4 or why not 2.

So if you go with the calculations what you see here is that the 3 nucleotides are actually going to give you the sufficient number of codons, so that you can be able to satisfy or you can be able to code all the 20 amino acids, so that we have 20 amino acids. So, if you want to code for the 20 amino acids the minimum number what you require is actually the 20 codons.

So, 20 codons, how you are going to get the 20 codons. So why there are only 3 nucleotides as a part of the genetic code, the genetic code is a triplet code called as codon. It is known that we have only 4 different types of nucleotide that make the whole genome, so it is known that the each codon consists of 3 nucleotides which means that 4 to the power 3 is actually going to give you the 64 possible codons for the different types of amino acids.

However, we have only the 20 amino acids, so it is obvious that the more than one codons are coding for the single amino acid and this is also illustrated by the verbal hypothesis. So, if you go with the 2 nucleotides, if you go with the 2 nucleotides, then how many numbers is going to be 4 to the power 2. So, what will be the number is going to be the 16 nucleotides.

So if it is 16 is actually lower than this number, and if you go with the 3 nucleotides, then the number is going to be the 4 to the power 3, and the 4 to power 3 is actually the 64. So, 64 is sufficient number which actually be able to code for the all the amino acids what are present into, what is being utilized by the ribosomal machinery for the synthesis.

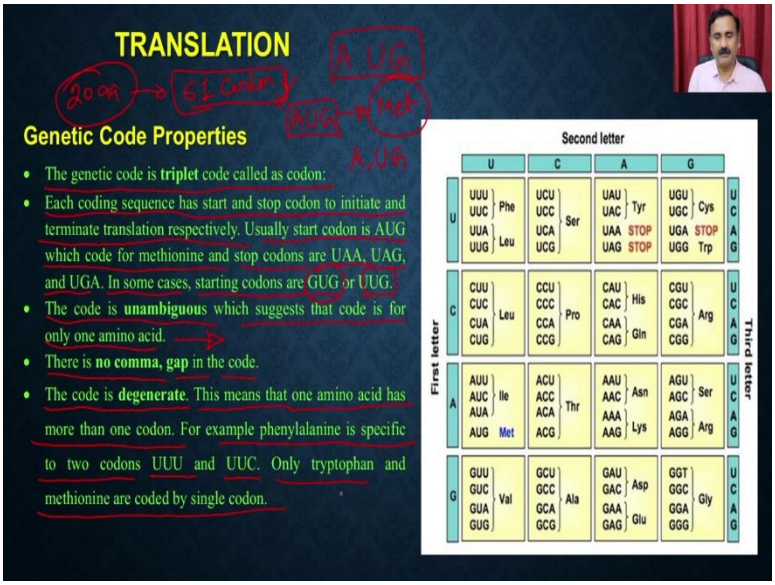
So, what you see here is that this is the table of the genetic code and the what you see here is so this is called as the genetic code table, and you can use this table to calculate to know which amino acid is actually going to incorporate so what you can go with the is the first nucleotide is actually of these then the second nucleotide and the third nucleotide.

So, for example, if I say U as a first, U as second, and U as third, then what will be the genetic code, is going to be U, U and U which is actually been present here. So, this is the first nucleotide. Then we can have the UUC, so U U, and you can have C from this site. So, it can be phenylalanine, so from UU is actually coding for phenylalanine, you can see that UUC is coding for the phenylalanine, but the other amino acid, the other combinations are actually coding for the leucine.

So, you can see that the for the single amino acids, the phenylalanine you have the multiple amino codons, similarly you can have, you can see that the for serine you have the 4 nucleotides similarly for the tyrosine, you have the 2 different types of nucleotides and so on. Apart from these, you can also have these some of these top codons like the you have the UAA, UAG and UGA.

And all these are called stop codon because they do not code for any of the amino acids. So, ideally for the genetic codes what you have is you have the 61 genetic codes, which are coding for the amino acids and the 3 genetic codes, which are coding actually are called as stock codons. So, let us see what are the general property of the genetic code.

(Refer Slide Time: 22:56)



Genetic codes is triplet, so genetic code is triplet called as codons, so these are going to be triplets, like for example AUG, so this and it is always been read as in triplets, so you cannot have like really in the any other combinations. Each coding sequence has start and stop codon to initiate and terminate the translation, respectively. For example, the start codon is AUG which code for the methionine and the stop codons are UA UAG and UGA.

In some cases, the starting codons are GUG or the UUC, so these are in the bacterial cases in the prokaryotic cases. You can also have the GUG or the UUC as a start codon. So what is mean by the start codon is that whenever the translation will start the AUG is going to be the first codon which is present onto the gene.

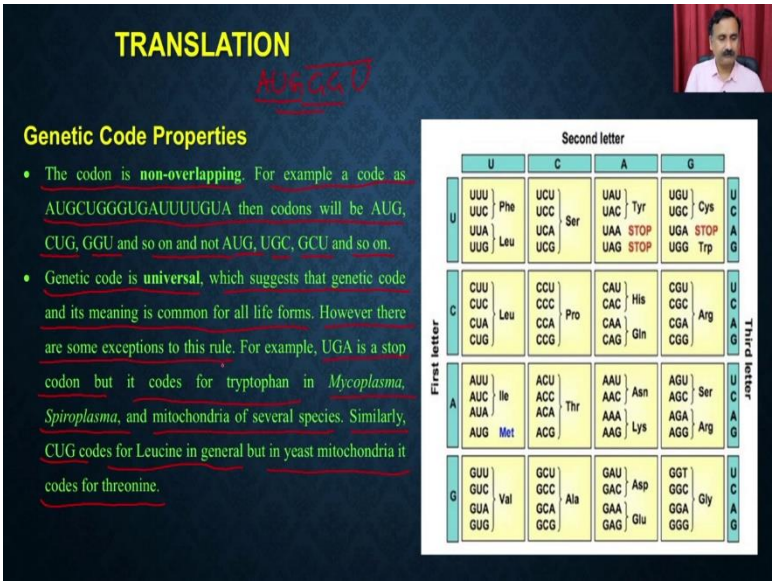
The code is unambiguous which suggests that the code is for only for one amino acid, which means code is unambiguous means it is not, it is not like that AUG is a codon for the methionine in E.coli but the AUG is a is a is a codon for the lysine or leucine or arginine in the other organism so it is actually a constant codon, it does not vary from the species to species or organism to organisms, that is called as unambiguous.

So that the AUG is fixed AUG is fixed for the methionine, it could be, and irrespective of the organisms. Then when you have the genetic code, there is no comma, there is no gap in the code, which means there is no like you cannot write, you cannot have the codon like this or you cannot have the codon like a, then there is a gap and then you can have AUG, so this is going to be a problem in the codon.

So it always going to be a continuous three triplets which is going to be constitute, which is going to be, going to make the codons. The codon is degenerate this means that the one amino acid has more than one codon, for example, the phenylalanine is specific and it is code by the two codon UUU and UUC. So, the codons are degenerate which means for the single amino acids you are going to have the multiple codon.

You might have seen that we have the 20 amino acid and we have the 61 different codons, so that is why you can all, you can ideally for every amino acid, you can have either two or 3 amino acids except there are exception that in for the tryptophan and as well as for the methionine, they are coded by the single codon.

(Refer Slide Time: 25:55)



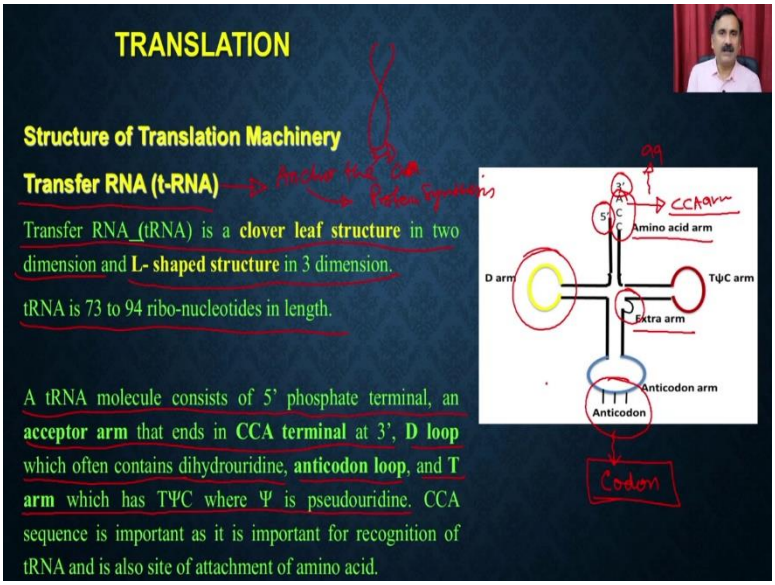
So, then the codon is non-overlapping, for example a code as the AUG this and that and then codon will be AUG CUC GUG and so on and not like AUG, UGC, CCU and so on. So it is not like, you can have the codons like, so it is not like, so for example, you have a sequence like this AUG, GGU. So, it is not like you can have a codon like first codon like this, then the second codon like this or third codon like this.

So you cannot have the overlapping codons, so they are not overlapping this means AUG will be like this then the second codon is like this, it is not they are not going to have the overlapping nucleotides. Genetic code is universal which suggests that the genetic code and it is meaning is common for all life forms, for however there are some exceptions to this rule.

For example, UGA is a stop codon, but its codes for the tryptophan, in the mycoplasma, is spiroplasma and the mitochondria of some species. That means is that genetic code is universal, which means if the AUG is being used for the methionine, it is going to be used for methionine and it is going to be used for methionine irrespective of the organism or irrespective of the gene actually.

But there are exceptions like for example UGA is a stop codon in all other species but it codes for the cryptophone in the case of some of the organisms like the mycoplasma, spiroplasma, and in some of the mitochondrial species. Similarly, you have another example like CUG, CUG codes for the leucine in general but in east mitochondria it codes for the theine. So, there are exceptions in the biology and that is why the biology is so much complicated and diversified.

(Refer Slide Time: 27:56)



Now, let us discuss about the next machinery, so the next machinery is the transfer RNA and we have said already that transfer RNA is actually going to anchor the amino acid and then it is actually going to deliver the amino acid or it is going to transfer the amino acid during the protein synthesis, as per the instruction from the ribosome.

So, transfer RNA is a clover leaf structure in the two dimensional and it has the L-shaped structure in the 3 dimensionals. So it is going to have the L-shape structure like this something like this in the 3D structure and it is going to have the cloverleaf structure like this in a two dimensional structure.

The total length of the tRNA is 73 to 94 ribonucleotides and it is going to have the different shape, so you are going to have the 5 prime end, you are going to have 3 prime end and the tRNA molecule is consist of the 5 prime phosphate terminal and acceptor arm that is ends in the CCA terminal at the 3 prime end. So, on the 3 prime end is going to have the CCA arm which is going to also called as the amino acid arm, so this is called as CCA arm.

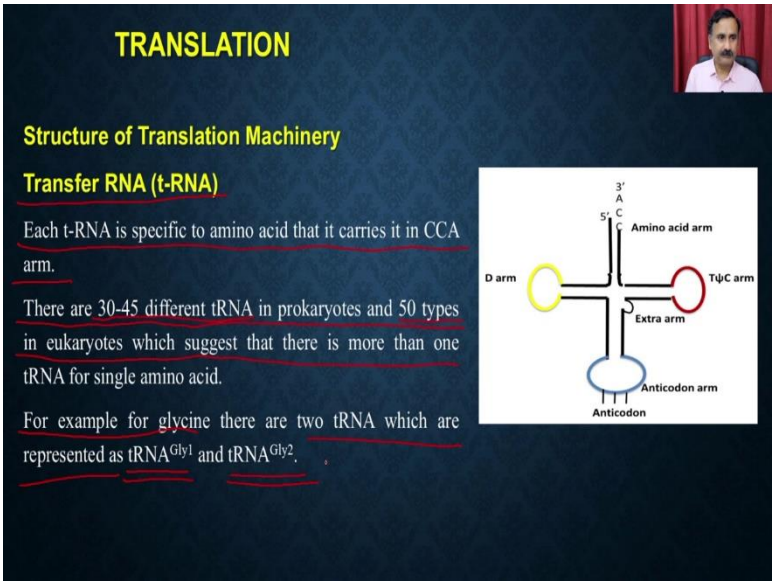
And this is the arm which is actually going to receive the amino acid. Then you also have the D loop, so this D, this is the D loop, so you are going to have the D loop, which often contains the some of the modified amino acids, nucleotides like the dihydrouridine, then you also have the anticodon loop. So, anticodon loops are actually going to have the anticodon and the purpose of the anticodon is that, it is actually going to interpret the codon.

So, depending on the anticodon, the amino acids are going to be tagged to the amino acids to the tRNA. So, the amino acid whatever is actually going to be tagged onto the 3 prime end with on to the this CC end the anticodon is going to be present, so that anticodon is actually going to check the codon and then accordingly, it is actually going to bring that particular amino acids.

Then, we also have the T arm and or which has the T psi C where size is actually, what is called as pseudouridine. So, you also have the T psi C arms and then you also have the extra arm. So, this is the extra thick arm, which actually is being taking care of the extra nucleotides which are present into the rRNA, so whatever the extra nucleotides are present that are being put into this extra arm.

And the most important region of the tRNA is the two region one is called as the CCA arm, which is responsible for the coupling of the amino acid, and the other is the anticodon arm which is actually going to interpret the codon onto the messenger RNA, and that is how these tRNA are actually going to bring the messenger, the amino acids for the coupling.

(Refer Slide Time: 31:00)



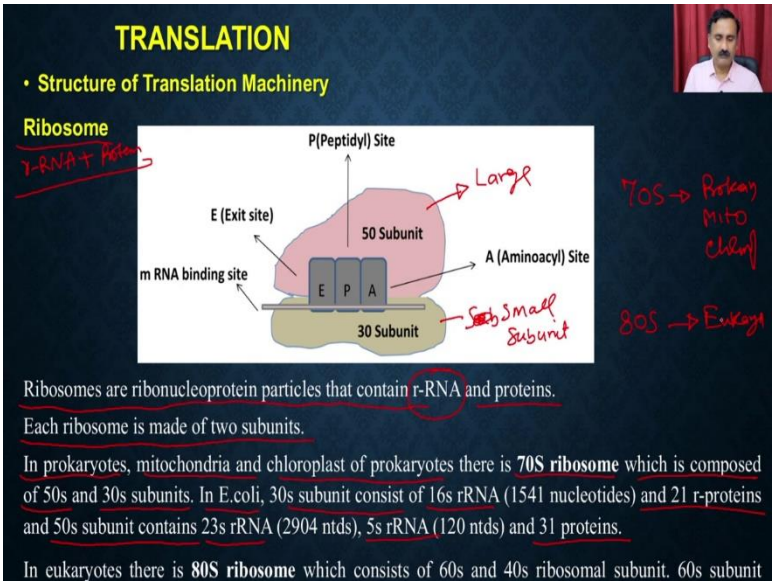
Then we have the transfer RNA, so transfer RNA, the each transfer RNA is specific to amino acid and it as, it carries on to the CCA arm. There are 30 to 45 different types of tRNA in prokaryotes and the 50 different types of tRNA in the eukaryotes. And, for example, for the glycine there are two tRNAs which are represented as the tRNA gly1 and tRNA gly2.

So, remember that we have the different types of codons, accordingly the tRNAs are also going to be different, right. So, we can also have the multiple copies of the tRNAs and these tRNAs are actually going to have the two information, one they are going to carry the amino acids, so they are going to have the information about the amino acid.

The second is they are also going to have the information about the codon which they are going to use so that they are going to do with the help of the anticodon. And that is why they we have the 30 to 45 different types of tRNAs in the prokaryotes, whereas the number of tRNAs into the eukaryotic system is even more like, you have the 50 different types of tRNA.

Which means for single amino acids you can have the multiple types of tRNA, one of the examples is that for the glycine it has the two different types of tRNA molecules, one is called as the Gly one, the other one is called as the tRNA Gly two.

(Refer Slide Time: 32:34)

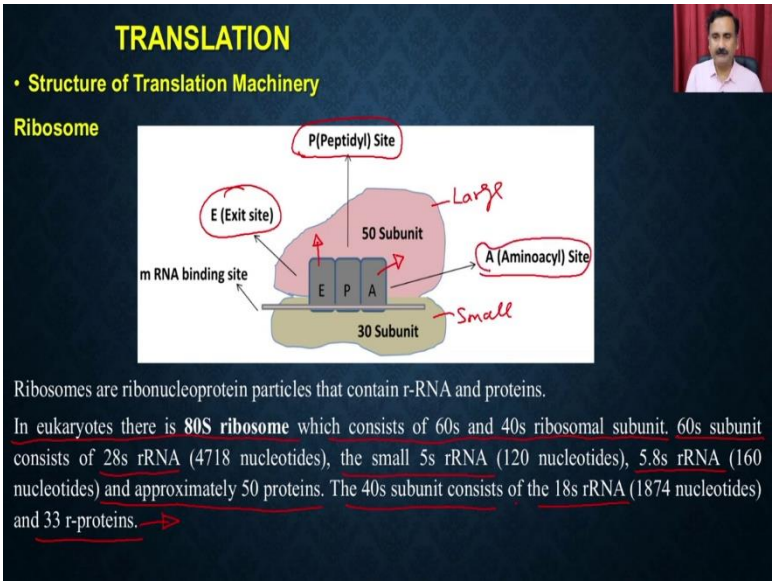


Now, let us move on to the next machinery and the next machinery is called as the ribosomes or ribosome or the ribosomal RNA, which is coupled with the protein to give you the ribosomes. So, the ribosomes are the ribonucleotides particles that contain the RNA, rRNA, and the proteins. Each ribosome is made up of the two subunits; you can have the large subunit, and the small subunit.

In prokaryotes, the mitochondria and the chloroplast of the prokaryotes, there is a 70s ribosomes, which is composed of the 50s and the 30s subunit. In the E.coli, the 30s subunit consists of the 16s ribosomal RNA, and the 21 ribosomal protein, whereas the 50 subunits actually contains the 23s ribosomal RNA, 5s ribosomal RNA and the 31 different types of proteins.

In eukaryotes it also contains the 87s, so you can have the two different types of ribosomes, you can have the 70s ribosome, which is actually present into the prokaryotes or it also contains into the mitochondria or the chloroplast, then you can have the 80s ribosomes, these are going to be present into the eukaryotes.

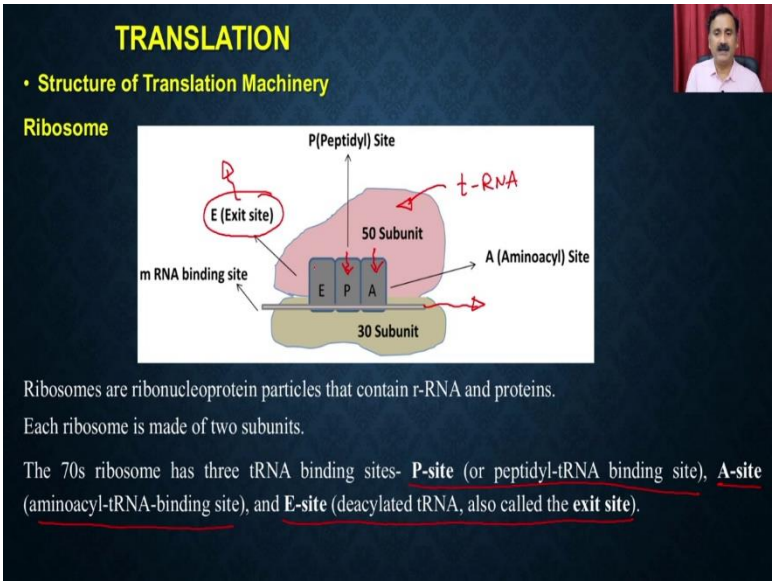
(Refer Slide Time: 34:11)



In eukaryotes, there are 80s ribosomes which consist of the 60s and as well as the 40s ribosomal subunits. 60s subunit consists of the 20s ribosomal RNA, the small 5s RNA and the 5.8 ribosomal RNA, whereas and it also contains the approximately 50 different types of proteins. The 40s ribosomal RNA, 40s small subunit consists of the 18s ribosomal RNA and the 33 ribosomal proteins.

So this is the composition of the ribosomal, ribosome, where you can have the large subunit, you can have the small subunit, and this large and small subunit, when they combined they are actually going to give you the 3 different types of activity sites. You can have the E site which is also going to called as exit site, you can have the P site which is called as the peptide site, and you can have the A site which is called as the A site or the amino site.

(Refer Slide Time: 35:22)

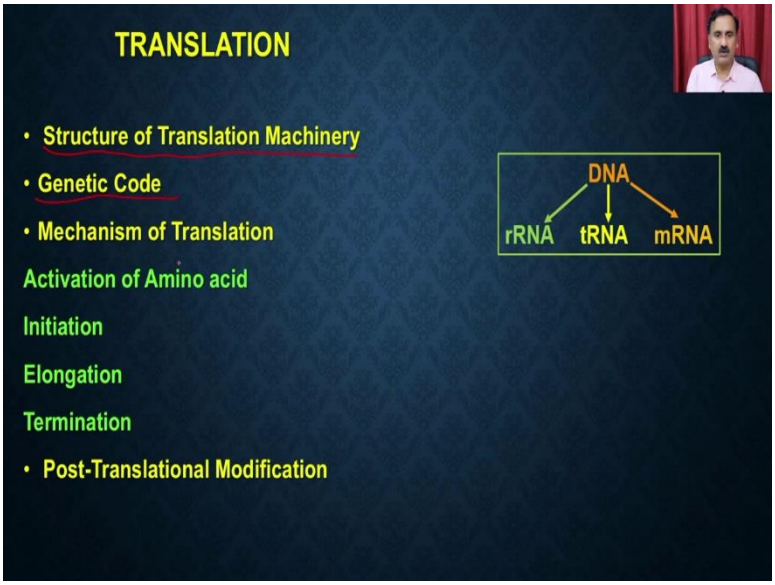


So, the function of these sites are very different. So, you can have the P site which is for the peptidyl tRNA binding site, A site which is actually going to be called as amino acid RNA binding site, and the E site from where the D isolated tRNA is going to exist, so that is why you have the E site from where the deacylated amino acid tRNA is going to removed.

Then you have the P site at where the tRNA is going to sit, so P site where you have the peptidyl tRNA which is going to bind and the A site where the incoming tRNA is going to come. So, this means the ribosome is actually providing the necessary infrastructure and framework for the protein synthesis.

On one site it is actually going to have the entry and exit of the tRNA, so tRNA is actually going to have the entry and exit onto the ribosomes, and on the other end it is also have the messenger RNA binding site, so it is actually going to interpret the message what is present on to the ribosome, onto the messenger RNA, and that is how it is actually going to be central machinery, where the protein synthesis is going to take place.

(Refer Slide Time: 36:42)



So, what we have discussed? We have discussed about the structure of the translational machinery, we discussed about the genetic course. So, with this I would like to conclude my lecture here, in our subsequent lecture we are going to discuss some more aspects related to the living organisms. Thank you.