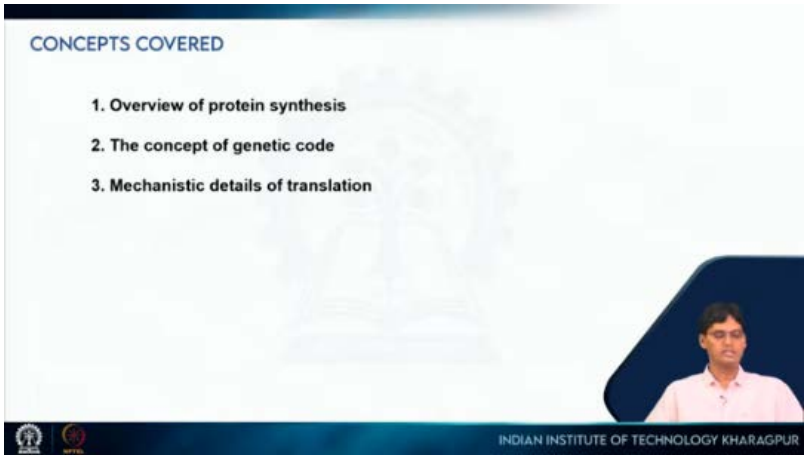
Introduction to Complex Biological Systems
Professor Dibyendu Samanta and Professor Soumya De
Department of Bioscience and Biotechnology
Indian Institute of Technology, Kharagpur
Lecture 8
Translation: Decoding the message of an mRNA

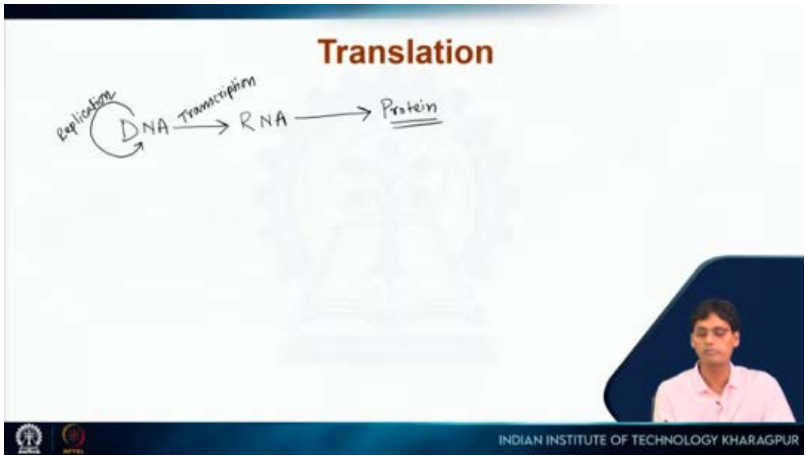
Hello everyone, today this is the third lecture of module 2. Today I am going to discuss translation, how the information present in mRNA is decoded by complex machinery. So, that is the essence of translation. So, mostly I will restrict myself here by discussing the overview of protein synthesis, the concept of genetic code, and the mechanistic details of translation.

It is a very complex process, but I will mostly highlight fundamental concepts about translation and some insightful ideas regarding this translation process. So, here translation means the synthesis of protein inside our cell. So, as I was discussing replication as well as transcription before in the previous classes. So, the DNA molecule, which actually contains the information in terms of nucleotide sequence, gets converted into RNA, and this process is called transcription. DNA goes to DNA, which also I mentioned before, is replication.



Now in both cases, whether we are synthesizing DNA or we are synthesizing RNA, in both cases we are basically synthesizing a polymer of nucleotides where the nucleotides are attached by phosphodiester bonds and the basic chemistry of DNA or RNA synthesis is basically the same. But now here is the next step of the central dogma. We will be getting proteins. So, here the protein molecule and this process is called translation. If you

remember when I was discussing replication or transcription, we mentioned the enzyme, the major enzyme involved in these processes, like DNA polymerase enzymes that catalyzes or that polymerizes deoxyribonucleotides.



Similarly, when we are getting RNA through transcription, the enzyme is called RNA polymerase, the polymerase, which polymerizes ribonucleotide triphosphates. But in the case of translation, logically the name should be protein polymerase, but there is no protein polymerase. So, translation is happening, or I would say translation is being conducted by a big machine that is called the ribosome. The ribosome helps in the process of translation. Now, in this case, in protein, amino acids are linked by peptide bonds.

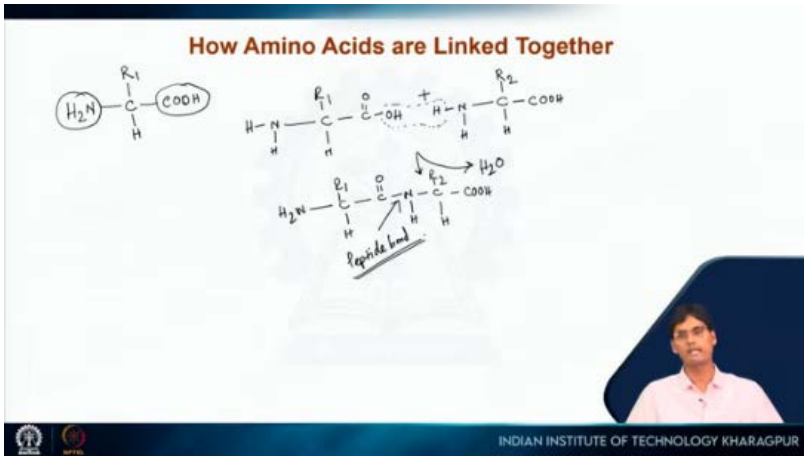
So, let us learn what a peptide bond is and how it is formed. So, how are amino acids actually linked together? This is the question. So, since proteins are polymers of amino acids, I should first introduce a little bit about amino acids. So, if we look at amino acids, this is the basic structure of an amino acid. They have one amine group and one carboxyl group.

The carboxyl group and here you have the side chain and another hydrogen atom attached to the central carbon. So, that is why this is called an amino acid, as both the amine and the carboxyl group are present there. So, we have 20 different types of side chains present. For example, if I put here R1, this is one specific type of side chain, and that should be designating a particular amino acid. But now, how are these amino acids getting attached together by a peptide bond? That I am going to show now. So, if I draw this structure in

this way: H N H C R1 H. So, this is the same amino acid, right? And now, another amino acid if I draw in this way, just nearby amino acid.

This is a different amino acid. I am putting R2 as the side chain so if R is just a hydrogen atom, then it should be glycine, which is the simplest amino acid. So, now here COOH. So, if those two amino acids react together, they make something like this structure here: H2N C R1 H C O, and this OH and this H from the next amino acid can leave as a water molecule. So, as a result of that, what we are getting here is NH C R2 H COOH.

So, now you can see these two amino acids are now attached together, and this bond is called a peptide bond. I discussed so many things about phosphodiester bonds when we discussed nucleic acids, but this is a peptide bond, and by which 20 different types of amino acids naturally present in our body get attached and make a specific type of protein molecule. Now, one important thing is to make a peptide bond in the lab chemically, which is not a very simple job. So, chemically, if we need to make a peptide bond, we need multiple organic solvents and also much higher temperatures compared to our body temperature.



But if you see in our body, inside our cells, the temperature is much lower, the ambient temperature or body temperature, and also we have an aqueous solution, not an organic solvent, but still everything is happening properly. Protein synthesis is constantly happening and that is why we are healthy and happy. So, let us learn how this synthesis happens. This is a complex process, but I already told you that I will stick to the very basic and fundamental aspects of protein synthesis. So, translation, the overall concept of

translation means protein synthesis inside the cell, that is all. Here, if I start from the beginning.

So, I am just writing something here, say, for example, if I write something here. So, something just written using the English alphabet. Now, if we know how to read English, if we learn the English language, then we can actually decode the information. Similarly, in our cells, inside our genetic material, that is DNA. So, information is written in such a way, but in this English literature, English language, I would say that

there are many alphabets, but in our genes, we have just 4 alphabets, A, T, G, and C, and just the combination of that, it will make some kind of sense, and we will get protein and that is the functional molecule, and that will be doing functions like everything happening in our body. So, whatever is written here, if we know English, then we can try to read this. So, I just intentionally did not give any space or anything because in our genes, in our DNA, those nucleotides are present like that, just continuously it is present. Here, if I just break them just after 3 letters. So, now, we are getting some kind of meaning that the sun had set, but was not out yet.

So, similarly if we decode the information present in genes, finally in the intermediate stage, it will make the RNA, particularly mRNA, the messenger RNA, and then after translation, we will get some meaningful product. I am just showing here, the sun had set, but was not out yet. So, also, I intentionally used only 3-letter words. So, 3 letters, what I used.

In living systems, those nucleotide sequences that are also read by just three nucleotides consecutively, and those three nucleotides present consecutively mean something that is called a codon. So, I will explain in more detail. So, see if I write here something like this, some sequence here, say, AGCUG, GAUGA. AAGGCCCA, and we can put many, many sequences here, and then, for example, UAAA. GC, and then 3'.

So, this is our 5' end and 3' end. So, what is this? So, this is nothing but some mRNA sequence, messenger RNA sequence, just I am putting it to explain the process of translation. Now, as you can see here, this is mRNA, and it should be decoded by the ribosome, as I already mentioned, the major machine which makes protein, that is called the ribosome.

So, as a result of that, when the ribosome will bind to the 5' end of this mRNA, it will sit on the 5' end of the mRNA, and there is some kind of signaling sequence, some specific sequence. There is a little bit of variation in eukaryotes and prokaryotes. But, I will discuss that a little later, but the ribosome will sit there and start reading this sequence which is present in this mRNA and while the ribosome is, so this is the ribosome. When the ribosome is moving in this direction, from the 5' to the 3' direction, then whenever it encounters the first AUG, this is only AUG, the first AUG. So, if I start from the 5' end, it is written AGCUGG. So, then you can see this AUG.

So, whenever it will get this sequence AUG, from there the protein synthesis will be started, and this is called the start codon. So, these three nucleotides later together we say a codon. So, this is the start codon, the start codon. So, whenever the ribosome will find the start codon, after that, every three nucleotides will be another codon. It is just going like this. What is the function of the codon? The codon will code for a specific amino acid and that is all.

So, a codon will bring some amino acid into the machinery within the ribosome where the translation is happening, and there will be some catalytic reactions and finally we will get the protein. So, as a result of that, AUG, the start codon, always brings one amino acid called methionine. So, we have 20 different amino acids. So, methionine is the starting point of a protein. So, AUG is the start codon, and I must mention here that this kind of

code is very much universal. So, methionine or the AUG should be the start codon in every mRNA during every protein synthesis, not only that but for every organism, whether it is bacteria, plant, or animal, that does not matter. The start codon should be AUG. If you ask me why? I do not have the specific answer. I would say this is because of millions of years of evolution, and life chose this thing in this way, that is all.

So, AUG is the start codon. Then, the next amino acid, for example, the codon here is AAA. So, it codes for lysine. You do not need to remember which codon codes for what amino acids, not declared. So, we generally use some kind of table that is called the genetic code by which we can understand which codon refers to which amino acids.

Then, because of the GGC, we will have glycine, another amino acid, then CCA, it will code for proline, and it is going on like that. So, the ribosome is basically adding amino acids one after another based on the sequence present in mRNA, that is the catch, that is the main thing in translation. So, you are making a specific product which is a polymer of amino acids based on the mRNA sequence, and the mRNA sequence actually came from the DNA. So, as a result of that, the actual information is present in DNA but through several steps, we are making the product which is a protein.

When the ribosome is moving in this way from the 5' to 3' direction, whenever it encounters this codon, which is UAA. So, this is a stop codon, so no matching amino acid will come. So in this case, the protein synthesis will stop. Thus, protein synthesis will stop at the stop codon. Now, I would like to discuss a little bit more about this genetic code and codon. So, I am just mentioning a few codons here. So, how many codons should there be?

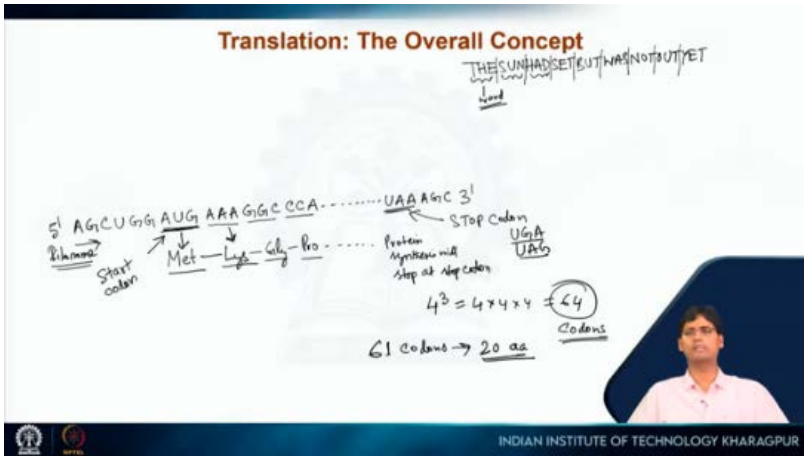
So, this is a very simple, a very logical answer to how many codons should there be? Because, as I already mentioned, we have just four nucleotides, and three nucleotides together make a codon. So, if you do different types of combinations, then you will get how many 4 to the power of 3, right? So, that is $4 \times 4 \times 4$, equal to 64 codons present.

So, this is the total number of codons present in the living system—just 64 words, in that way. So, now out of these 64 codons, there are 3 stop codons. I just mentioned one, UAA, apart from that, 2 more stop codons are UGA and UAG. So, those 3 are the stop codons and the start codon I already mentioned, AUG. So, as a result, those 3 stop codons do not

specify any amino acid, but now if we just ignore those 3 codons, then we have 61 codons left. So, now those 61 codons are responsible for how many amino acids? Because I told you, naturally in our cells, we have 20 different types of amino acids. So, they code for 20 different types of amino acids. I have given a few examples while I was explaining here, for example, methionine, lysine, glycine, proline, etc., and their corresponding codons. So, if we have 61 codons and 20 amino acids, that means some kind of degeneracy is there. That is, I would say, multiple codons actually specify a single amino acid. In a few cases, multiple codons specify, like, a single amino acid.

So, that is why this number relation, the 61 codons and 20 amino acids. So, whatever I have mentioned until now is a very, I would say, methodical and simplistic way of how the ribosome reads the sequence present in mRNA and it makes protein, but the actual scenario is much more complex, right? So, that is why, if you see the complexity of the biological system, it is really a complex scenario.

So, if I start a little bit on that, the first question here is how a particular amino acid is coming because of this AUG. For example, I told that AUG codes for methionine, but how does methionine come? Similarly, AAA brings lysine to the translation site, but how does it come here? So, that is the major question. and at the time when all those things were being discovered, it was a really challenging question. At that time, many scientists thought that those three nucleotides present might be making some kind of chemical pocket, and that chemical pocket could attract the side chain of a specific amino acid. But it is really hard to imagine how this will bring 20 different types of amino acids and how this process is happening.



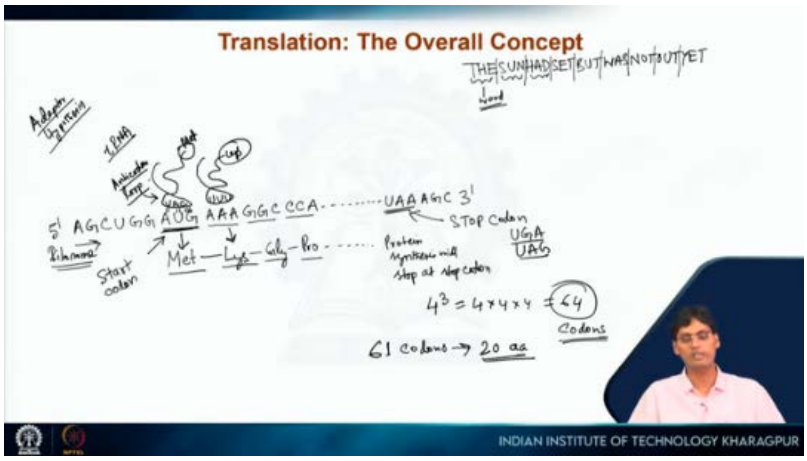
So, actually, in this scenario, Francis Crick brought forward some hypotheses. He suggested something that is called the adapter hypothesis. So, the adapter hypothesis means Crick was suggesting that maybe inside our cells, we have some kind of smaller molecules, smaller polynucleotides, like some kind of nucleic acid molecule. which may act as adapters. Like all of us know what an adapter is, sometimes we use adapters to charge our devices, for example, a laptop. Similarly, what Crick was suggesting is that if we have some adapter molecule made up of nucleotides, then it can recognize the sequence present in mRNA, the codon sequences can recognize and, by different methods, it can bind to a specific amino acid. So, that particular adapter can bring a specific amino acid to the translation site. Remember, when he proposed this idea, tRNA had not been discovered. tRNA stands for transfer RNA. It was not discovered, it was just a hypothesis. But scientists later found those small RNA molecules called tRNA, which actually bring amino acids to the site of translation. So, if you see here now, this is just a cartoon diagram of some tRNA. So, tRNA, transfer RNA, is also a polymer of ribonucleotides. Now, if you see, this is made up of a nucleotide sequence. So, in this, there is one loop. So, there is one loop present in this tRNA, and here you have a complementary sequence to the codon.

So, for example, in this tRNA, the sequence will be UA C. Why? Because UAC actually recognizes AUG through Watson-Crick base pairing and hydrogen bonding. So, as you can see, UA and a triple bond here. So, this way, this tRNA can recognize this particular codon and this particular tRNA can bind to a specific amino acid, which is methionine. That is why whenever you have the codon AUG, it will always bring the amino acid methionine. That is all. Similarly, the next tRNA will bring another amino acid, as I already mentioned

here. So, it is bringing lysine. So, the sequence in the loop should be U U U, so that it can recognize the codon sequence AAA present in mRNA.

Now, this loop is called the anticodon loop. Why? Because it binds to the codon sequence present in mRNA. So, the anticodon loop.

So, codon-anticodon recognition is actually dropping the correct amino acid at the site of translation, and then some catalysis is happening. Let us see, in the next slide, I am going to discuss a little bit more. I will add something more here, but this is the basic schematics, a basic explanation of protein synthesis and how it is happening. So, here are the translation, the involved machineries, and the processes. So, as you can see, the template for translation or the template for protein synthesis is definitely, as all of you know mRNA, the major machine I mentioned is the ribosome, and the ribosome is made up of rRNA and proteins, many proteins involved in making this ribosome particle and then we have tRNA, those adapter molecules, transfer RNA, that bring amino acids to the site of translation.

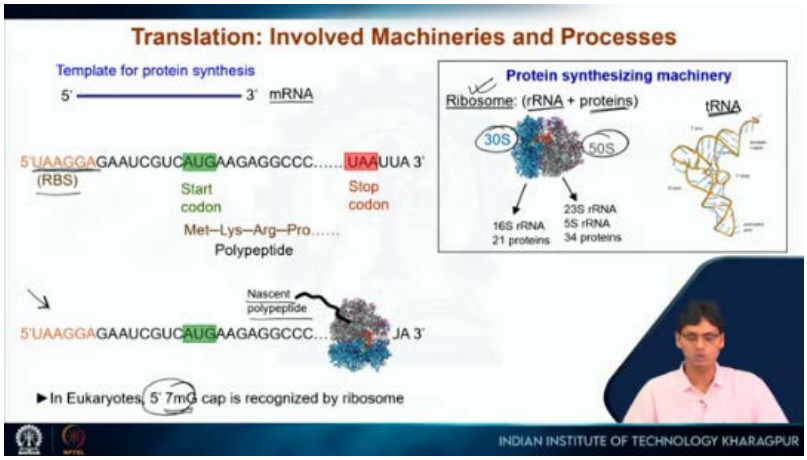


Now, see what is happening? If you see this ribosome, we have 30S and 50S, those are the different subunits of the ribosome. In the case of E. coli or bacteria, they have 30S and 50S subunits. Similarly, in the case of the eukaryotic system, they have 40S and 60S subunits, but those details are not required to understand the fundamental aspects of protein synthesis. Now, if I put this sequence particularly here, this is an mRNA sequence, and if you see whatever I explain, the first AUG should be the start codon, and then it will go like this. After AUG, every three letters will be bringing some specific amino acids, and we are getting a polypeptide. So polypeptide and protein are very similar things. Polypeptide is

generally what we say when the length of this polymer is much smaller and when it is much bigger, and has some kind of three dimensional conformation, we generally refer to that as a protein. Otherwise, they are the same in terms of chemistry, as amino acids are linked together by peptide bonds. Now, if you see how the ribosome will actually bind to this mRNA.

So, here, for example, in the case of the bacterial system, they have some specific sequence at the 5' end of the mRNA, and that will attract the ribosome. The ribosome will get attached to this site. This is again some kind of Watson-Crick base pairing because I already mentioned that the ribosome is made up of rRNA and proteins. So, that means whatever the ribosome binding site I am mentioning here in this sequence, for example, 5' UAAGGA. So, some complementary sequence is present in some ribosomal RNA, which is present inside the ribosome, and by virtue of that, the ribosome will come and bind here and then it will move towards the 5' to 3' direction, and then the first AUG, whenever it is encountered, will start protein synthesis.

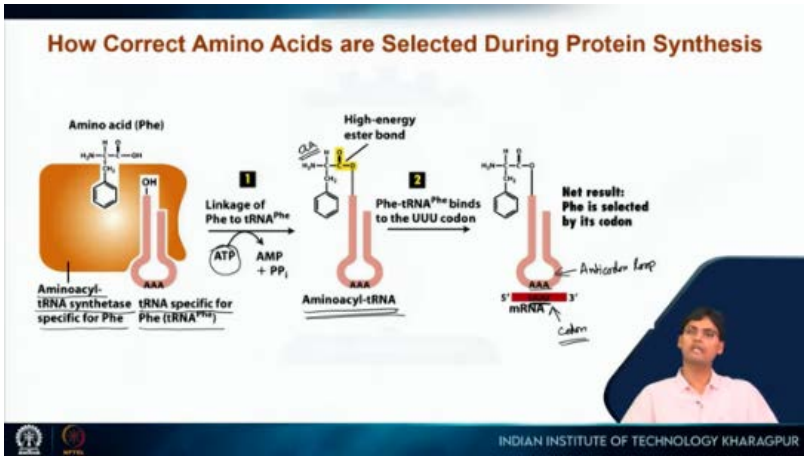
So, as you can see here, this is the ribosome. It is attached in this sequence. Now, it just recognizes this sequence and will move from the 5' to 3' direction like this, and the polypeptide or the protein will be synthesized here. I am mentioning that nascent polypeptide, which means the polypeptide is just being synthesized. So, in the bacterial system, I mentioned this kind of sequence present, which is called the ribosome binding site. Similarly, in eukaryotes, when I was discussing the processing of eukaryotic mRNA, I told that there is some kind of special cap called the 7mG cap present at the 5' end of eukaryotic mRNA. So, in eukaryotes, this 5' 7mG cap is recognized by the ribosome, and it will help in translation. Now, the next question is how a specific amino acid should be attached to a specific tRNA, right?



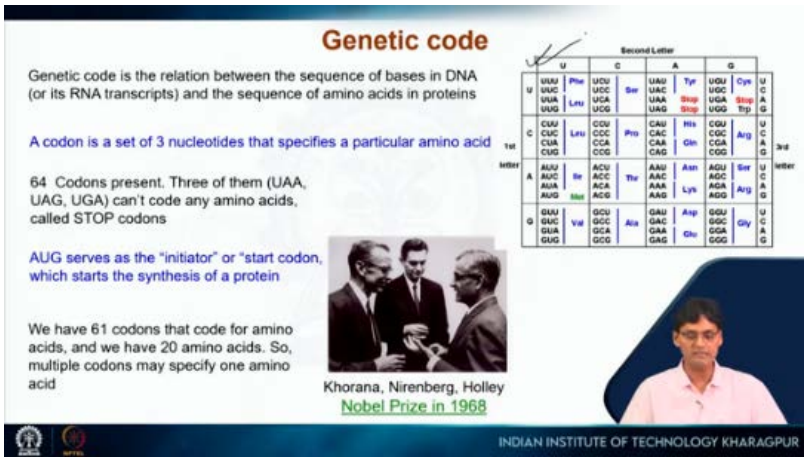
So, tRNA brings the correct amino acid to the site of translation. So, in this slide, I am trying to explain that. So, this is the tRNA you can see here that this is tRNA, and it is specific for phenylalanine. Phenylalanine is an amino acid. So, the system is that for a specific tRNA, we have some enzymes present in our cell; in this case, I would say this enzyme, whose generic name is Aminoacyl tRNA synthetase.

So, they will add a specific amino acid to the tRNA. So, in this case, the Aminoacyl tRNA synthetase is specific for phenylalanine. So, as a result, for other tRNAs, they have some other enzymes, and they will add a specific amino acid to that tRNA. So, you need some energy to carry out this reaction. The ATP is here and then finally, you will get aminoacyl tRNA, which is sometimes called charged tRNA. So it is aminoacyl tRNA, where the phenylalanine has already been added to the tRNA.

This tRNA will go and bind to the mRNA sequence during translation, as you can see in the mRNA sequence. Now, if you have U U U, this is the codon, then AAA is present in this anticodon loop, and this is our codon now. So, as a result, this phenylalanine is coming into the site of translation. So, this process will continue, and as a result, the correct sequence of amino acids will be added based on the sequence present in mRNA. You should always remember that is why we are actually translating from the mRNA template.



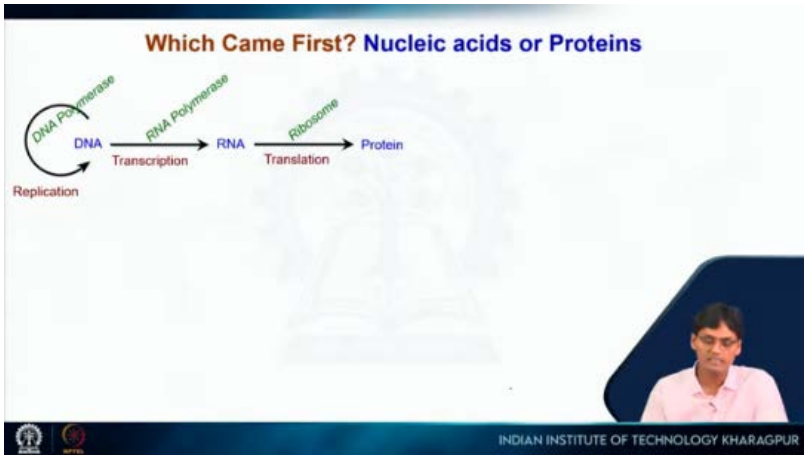
Here, some information is given, which I already discussed about the genetic code. So, as I already told, it is very difficult to remember which codon codes for what. So, here is the genetic code table. So, all 64 codons are here, the genetic code, and their corresponding amino acids are mentioned here. We do not need to remember that, but just for an idea and everything we discussed is briefly mentioned here.



As you can see, this is a fundamental step in the protein synthesis mechanism, and this is the same. It is universal everywhere in plants, animals, and bacteria, as I already mentioned. Because of this process, Har Gobind Khorana, Nirenberg, and Holley received the Nobel Prize in 1968 for their beautiful work on protein synthesis. Here, I would say some stimulating discussion is still remaining, and I would like to explain this thing because if you go into much more detail about translation in a textbook, you will get around 50 pages of different types of discussion there. So many factors, so many protein molecules, and so

many things are involved to make this process more efficient. But the basic logic, as I already explained, is how protein synthesis is being carried out inside.

But now the question is that if you see this central dogma here, the protein is basically what we are getting at the end. This is the last step of this central dogma. But now, before that, we need to replicate our DNA, and then we have to make the RNA through RNA polymerase. So, if you see this DNA polymerase or RNA polymerase, all are proteins, right? So, they are enzymes, but they are basically proteins, the protein enzymes. The question is then how those proteins or enzymes are formed if protein formation is at the end? How did life started? This is a kind of difficult question to answer. It is, I would say, a chicken and egg problem.



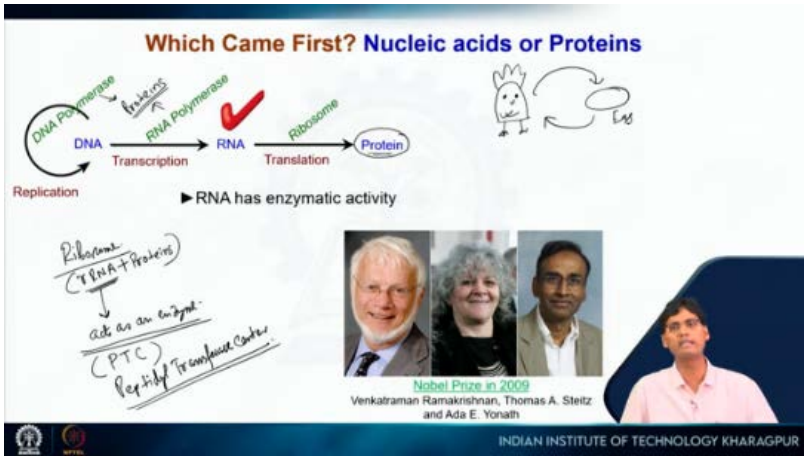
So, this is the egg and the chicken, the chicken-and-egg problem, because if we do not have the protein, then how are we getting this RNA polymerase, DNA polymerase? Not only the DNA polymerase or RNA polymerase, I mentioned some other enzymes also, and all of them are proteins. So, as a result of that, the idea is that RNA has enzymatic activity. So, RNA can perform some enzymatic activity. So, this can be discussed somewhere else also. The RNA World hypothesis. The RNA came first because, as I already mentioned in some other classes, RNA can be genetic material also, as you can see in some cases, in some viruses, where they have RNA as their genetic material. and it can also catalyze some reactions.

So, as a result of that, RNA has enzymatic activity. So, maybe at the beginning of life formation, maybe during that time, RNA was the genetic material, and it could act as

genetic material as well as having some kind of enzymatic function also. Similarly, if you see the ribosome, which is the major machine for protein synthesis. So, the ribosome and the ribosome are made up of rRNA, ribosomal RNA, plus so many proteins.

Many scientists work for a long time on their structural functions, the structural and functional relation of the ribosome, how they catalyze the reaction, and so many other things. Interestingly, they found that this catalysis, the protein synthesis, is performed by rRNA, not by protein. So, rRNA actually acts as an enzyme. This enzyme is actually acting as an enzyme. RNA, particularly what we call the PPC, the peptidyl transferase center. So, this is the region of ribosomal RNA that catalyzes the reaction. Particularly, I would like to mention here that this is such a fundamental step in protein synthesis, but it is being catalyzed by a non-protein enzyme. Protein is not involved there.

As a result of that, it actually suggests the RNA World hypothesis. This is such a fundamental thing that protein synthesis is happening every second in every life form, but it is still being catalyzed by ribosomal RNA, or you can say just RNA, but not by protein. Because of their beautiful work, not just these three scientists, but many other scientists actually contributed a lot. In 2009, Venki Ramakrishnan, Thomas A. Steitz, and Ada Yonath received the Nobel Prize in chemistry for their beautiful work, particularly understanding the structure-function relation of the ribosome and how it helps in translation, and so many different types of insights into the ribosome. Now, this is the last slide here. Translation machinery, particularly the ribosome, is an attractive target for therapeutics. If you see many antibiotics, which are some kind of drug or medicine that we generally take to prevent bacteria, antibiotics inhibit bacterial growth.



Now, if you see many antibiotics, for example, tetracycline. This is the generic name. In commercial terms, for example, some doxycycline, tetracycline. Those antibiotics actually target the ribosome. So, tetracycline and doxycycline bind to the 30S ribosome and block the binding of aminoacyl tRNA to the A site. The thing is, translation is happening in both bacteria as well as humans. I told you it's a fundamental process, but still, there are some differences. Many scientists, when they worked on understanding the ribosome, actually elucidated what those minor differences are.

Based on that, scientists can design some antibiotics that will specifically stop the protein synthesis of bacteria. Then those antibiotics, like they are targeting the ribosome, but they are not creating problems for eukaryotic ribosomes, or particularly human ribosomes. As a result, they are very good medicine to prevent bacterial growth. If there is no protein synthesis, bacteria will not survive; they will not divide in our body. We have more examples, like chloramphenicol, which is present in many of our eye drops. That also prevents the peptidyl transferase reaction, which means this peptide bond formation on the 50S ribosome. And then, streptomycin, erythromycin, those are some other examples of antibiotics. There are many more, but they actually block different stages of the translation process.

And this is one I would say an interesting question you may try to look into it. What could be the possible answer to why ribosomes are an attractive target for the development of antibiotics? Because if you try to buy 100 different types of antibiotics from a medicine shop, you will find many of them are actually targeting the ribosome, but what is the reason

for that? So that is all. You can follow any of these textbooks, all books are very good, so any textbook you feel comfortable with, you can go with that. Thank you very much.