Comprehensive Molecular Diagnostics and Advanced Gene Expression Analysis

Prof. Aritri Bir

Dr. B.C. Roy Multi-Speciality Medical Research Centre

Indian Institute of Technology Kharagpur

Week 01

Lecture 04: Protein Synthesis (Translation)

Namaskar. Welcome back to the NPTEL lecture series, Comprehensive Molecular Diagnostics and Advanced Gene Expression Analysis. So, we were discussing the very basics and fundamentals of molecular biology and molecular diagnostics, where we are discussing the central dogma that we have discussed replication, we have discussed transcription. Today we are going to discuss translation that is protein synthesis. So, here in this lecture class we are going to discuss the details of the mechanism of translation also a bit of post translational modifications. So, before starting the mechanism we need to know what is genetic code because that is the code which is going to help to synthesize all the proteins present in our body.

So, basically genetic code is nothing, but the collection or dictionary that identifies the correspondence between the sequence of nucleotide present in DNA as well as RNA and the sequence of amino acids. So, that is a link. So, basically this is comprises of like dictionary has multiple words similarly the genetic code has multiple codons, codons which are formed by 3 successive nucleotides or triplets of nucleotides considered as word in the dictionary of genetic code. Now, the codons have specific characteristics it is it has degeneracy, it has unambiguity, universality, non overlapping and comma less.

So, let us discuss what are these. Now, what is the basically if we see this is our mRNA where all these nucleotides are present. So, these are the nucleotides which are present in successive manner. Now, codons are rate from the very first amino acid which is taking part in the protein or amino acid synthesis. Now, these are the considered these are 3 nucleotides which are considered as codon.

So, they are read in a non overlapping manner. The next codon this is the first codon. So, basically the next codon is this. So, even if this is the first codon and this is the second codon in between them there is no comma or there is no punctuation. So, basically the frame is read in a continuous and comma lessness manner.

Now, the characteristics of these codons are these 3 triplet of nucleotide basically indicates 1 amino acid. Similarly, these 3 indicates another amino acid. But what happens that 1 amino acid can be coded by different types of codons more than 1 codons can be representing 1 single amino acid that is the degeneracy of codon whereas. So, these are the first codon and the second codon they can be they can give rise to 1 amino acid. If we read it from the back way these codon cannot give rise to multiple amino acid.

So, number 1 codon can only give rise to number 1 amino acid. Similarly, number 2 codon can give rise to this same amino acid. But if you see from the perspective of the amino acid, amino acid can be read can be rather coded by different types of codon. So, there is degeneracy, but there is universality means this 1, 2, 3 triplet of nucleotide will always give rise to 1 type of amino acid only in all the species in the universe. So, that is our universality.

Now, there are other types of codon a special type of codon that is known as initiation codon. So, initiation of translation or the very beginning of translation starts with this initiation codon that is AUG which at represents methionine a amino acid methionine. Similarly, there are other types of codons like termination codon these are our terminations codon which are very specific UAA, UAG, UGA these are the very specific questions in MCQs basically. So, a reading frame or the mRNA triplets suppose this is once again our mRNA. So, this reading frame is continuously read till we encounter suppose this is 1 UAA.

So, 1 termination codon is reached in that case the translation stops. Now, an open reading frame we consider when this reading of mRNA continues till 50 or more codons. So, a reading frame which does not have any termination codon till 15 amino acids are formed those are known as open reading from frame. Now, remember for a successful translation we need open reading frame and the more the sequences are there without the termination codon the longer protein synthesis or and the successful translation can happen. So, you can see here are multiple types of codon here.

So, this this is PH is phenyl alanine, phenyl alanine can be represented by this triplet UUU and UUC whereas, UUU can only gives rise to the amino acid phenyl alanine. So, I think the concept of universality and degeneracy are clear. Here you can see AUG which represents the initiation codon representing methionine remember in the proteins in the protein chain wherever methionine is it is always represented by AUG only, but the initiation codon must start with AUG or methionine. Now coming to global hypothesis which has which is actually postulated by our very own Crick Francis Crick. So, it says that just we discuss that codons can degenerate means 2 or more codons can represent or can code for one single amino acid.

In those cases it has been seen if we go to the previous slide you will see the difference in those type of codons are basically in the third base. So, the first those 2 or more codons which represents one amino acid they are different in the third codon. Similarly, if you see the histidine, histidine is decoded by coded by CAU and CAC. So, the third codon is different again arginine here you can see the first 2 are CG where are the fourth are CUCAG. So, the third codon differs.

Based on that the postulation is described as the third base of most codons pairs loosely with the corresponding anticodon means the third base wobbles. Now corresponding anticodon is located if you remember that is in tRNA. Now if you remember the very structure of the tRNA there is one anticodon arm and there are 3 nucleotides or 3 triplets of nucleotides are presents sorry a triplet of nucleotides are present which basically base pairs complementary with the codons. So, they too bind codons present in mRNA and anticodon in tRNA they are basically complementary to each other. Now in those cases where the degeneracy occurs the third base basically differs.

So, it has been postulated in that third base the binding is loose. Now based on that 4 other concepts have been proposed or postulated that comes under the wobble hypothesis. So, what are those the first 2 bases of an mRNA codon always from strong base with the anticodon whereas, the third base binds loosely. Second the first base of the anticodon present in tRNA determines the number of codons recognized by the tRNA. Now basically tRNA anticodon can identify or can bind with different types of codon and that is the reason of our degeneracy.

Because if you remember once again that tRNA every amino acid has their own specific tRNA to which they bind. Now based on that choice of the tRNAs anticodon to bind the codon present in mRNA degeneracy occur. So, the first base present over the anticodon determine the number of codon that can be recognized by the tRNA. How? If the first base of the anticodon is C or A the base pairing is specific and only one whereas, if it is U and G, G means is one in uracil I am talking about the nucleotides. So, the binding is less specific in that case 2 different codons can be read whereas, if the first base of the anticodon is ionosine first nucleotide present in the anticodon then 3 different codon can be recognized.

So, this is the second one. Next when an amino acid is specified by several different codons the codon that differ in either of the first 2 bases require different tRNAs. So, if the amino acid needs multiple it can be coded by different codons then I mean then in that case the first 2 bases of the codon require different tRNAs. Then as I told you minimum 32 tRNA are required to translate all 61 codons 31 why because 31 amino acids are present and 32 because plus 1 for the initiation of initiation codon that is AUG

a separate tRNA is present. Now, why all this problem? What would have been if all the bases of the codon engaged in strong Watson-Crick pairing with the 3 bases of the anticodon? Remember this question very important because we always say that there is a strong base pairing complementary base pairing of hydrogen bonds with the codon and the anticodon or the complementary nucleotides.

So, if instead of Wobble which says that the third base pair binds loosely if it would have been like all the 3 base pairs are strongly bound what would have been the problem? The problem lies in delivering the amino acid over the mRNA. So, the function of tRNA is to deliver it attach with that codon and anticodon sequence it attach to the mRNA and delivers the amino acid from the cytosol over the mRNA to form the protein chain. After delivering the tRNA needs to be detached and take part in bringing that same amino acid from other parts in the protein chain. So, that dissociation would have been very much problematic if 3 of them would have been binding in a strong way. So, next going directly to the steps of protein synthesis.

So, the translation is also having initiation elongation and termination, but along with that the very beginning is activation of amino acid without this protein synthesis cannot occur. And after the synthesis of the protein there is another modification that is folding of the protein in a proper way to gain the amino acid chain gain it secondary tertiary or quaternary structure and also some post translational processing. Now, coming to the activation of amino acid that is based on 2 very fundamental requirement that is activation of the amino acid in its carboxyl group to facilitate the formation of peptide bond. And a link for establishment between amino acid and mRNA these are 2 very important thing which has required and those 2 are achieved by 1 single event that is attaching the amino acid to its very specific tRNA. Now, that tRNA remember this binding of tRNA with the amino acid does not occur over the ribosome where the protein synthesis occurs.

This binding of amino acid with the tRNA remember it occurs in cytosol with the help of the enzyme aminoacyl tRNA synthetase which is a magnesium dependent enzyme. Now, this binding of amino acid with the tRNA this term is also known as charging of tRNA the tRNA is charged with the amino acid. Now, remember that the tRNA has different terms amongst that this is our anticodon arms which is complementary to the codon over the mRNA along with that there was another arm that is amino acid arm where this sequence is present. The here amino acid binds with the tRNA. So, this amino acid arm can carry a specific amino acid esterified by its carboxyl group to the 2 prime or 3 prime hydroxyl group of the A residue adenine residue at the 3 prime end of the tRNA.

So, here the amino acid binds with the tRNA and each 20 amino acid covalently

attached to its 20 types of specific tRNA and of course, this process is energy dependent and here ATP is hydrolyzed. Now binding of this codon and anticodon this is by the enzyme aminoacyl tRNA synthetase is known as the second genetic code because this is how the proof reading in the synthesis of protein is maintained. How the carrying of amino acid by tRNA to the mRNA or in the protein chain is dependent on a 2 step reaction. How initially the enzyme aminoacyl tRNA synthetase in a in a site which is not the active site in that site the enzyme is helps the amino acid to bind with the ATP. So, what is formed one amino acid AMP conjugate is formed over the enzyme aminoacyl tRNA synthetase.

Now these conjugate now then transferred to the active site of the enzyme where the amino acid is transferred to the growing polypeptide chain or it is bound to the it is forming rather thus am I charged tRNA or aminoacyl tRNA with the release of this AMP. Now these 2 step reaction basically helps to identify the specific amino acid which is specific to that tRNA only and delivery of the specific or right amino acid in the protein chain because this is a 2 step reaction. If any mistake has happened in these step that aminoacyl adenylate complex if there is any wrong amino acid is incorporated when it comes to the active site aminoacyl tRNA synthetase has the property to hydrolyze this bond and release the wrong amino acid. So, this is how proofreading can be done in case of translation. Next going to the initiation of translation where a very specific amino acid that is methionine is utilized to initiate the protein synthesis.

Now protein synthesis begins in case of the protein chain protein synthesis begin from the amino terminal end whereas, consecutively in a successive manner the specific amino acids are attached and a carboxyl end is growing. So, AUG is the initiation codon representing methionine. Now remember there are as I told you 32 because one AUG which is the initiation codon or the initiation methionine is carried by one tRNA whereas, all the other methionine which are present in between the protein chain those are carried by the separate type of tRNA. And then in initiation there is formation of initiation complex which we are going to discuss. Now in bacteria the first methionine is a modified version that is formyl methionine whereas, in eukaryotic cell the specific tRNA for AUG present that is not because in bacteria there is a tRNA which is specific for formyl methionine.

So, definitely two types of tRNA which carries formyl methionine and which is present only methionine which is present inside the peptide chain. But in eukaryote there is no such formyl methionine, but there is specific tRNA which is present which is carrying the first methionine of the protein chain. Now let us see step by step what happens. Now remember ribosome is the organelle where protein synthesis occurs. In the ribosome suppose we are discussing the prokaryotic translation now and we discuss we will discuss what are the differences between prokaryotic translation and eukaryotes.

Now this is the 30th subunit of the ribosome here initiation factors attach. What are the initiation factors? Initiation factors 1, 2 and 3 take part in bacterial prokaryotic initiation. So, here you can see initiation factor 1 and 3 they attach to the 30th subunit of the ribosome and over that mRNA this green representing the mRNA which has the codons it is attached. How it is attached? So, this 30th ribosome has 3 sides or the rather the ribosome has 3 side P side, A side and E side. In the P side the first initiation codon that is A ug is attached or aligned the mRNA is aligned in such a way that the A ug falls over the P side and these specific binding is helped by a sequence that is known as signed algarno sequence which is present in mRNA.

Now signed algarno sequence is a 4 to 9 purine residue sequence which is present 8 to 13 base pair 5 prime side of the initiation codon. So, initiation codons 5 prime side there is our signed algarno sequence that signed algarno sequence basically base pair with the complementary pyrimidine rich sequence near the 3 prime end of the 16 S rRNA. So, the signed algarno sequence is present over the mRNA and it binds with a pyrimidine rich sequence present over the 16 S rRNA over the ribosome. So, signed algarno sequence finds and attaches over this region in such a way that A ug falls over the P side. Now along with 30th subunit there is 50th subunit of the ribosome.

So, basically initiation factor 3 prevents the binding of the 30th and 50th subunit whereas, in initiation factor 1 basically blocks this is our A site blocks the A site why the blocking is required? We can see in the next step the initiation codon that is the formyl methionine carrying tRNA with its anticodon binds over the P site. So, it must not get attached over the A site and that A site to prevent a mistake is blocking and the is blocked by initiation factor 1. Now, charge tRNA with formyl methionine is carried by another initiation factor that is initiation factor 2 which is bound to GTP. So, basically this complex with 30 S ribosome then here is our A ug along with that formyl methionine carrying tRNA with initiation factor 2 GTP it forms the initiation complex. And that is helped by attachment of the 50 S subunit also.

So, in the initiation complex the final initiation complex what we get all the initiation factors 1, 2, 3 they all are released because the initiation factor 3 is released basically 50 S ribosome can bind with this complex and also the bound GTP here you can see is hydrolyzed and released as GDP and inorganic phosphate. So, this is basically one energy required stage. So, we have initiation complex where the first very first methionine the first amino acid is already attached over the P site. Now, along with this P site and A site I told you there is another site that is E site which represent exit. Now, this E site is basically present in 50 S subunit.

E site can only accommodate one uncharged tRNA means tRNA which is not attached

to any amino acid whereas, P site and A site can accommodate only charged RNA tRNA charged tRNA that tRNA which is attached to specific amino acid. Now, in case of eukaryotic translations initiation there are different initiation factors which are present and they form complexes. Now, remember the reading frame reading frame means that part of mRNA which is getting translated that reading frame is basically bound from both the sites 5 prime to 3 prime site and form a complex at the 3 prime end there is a poly A binding protein. So, you here you can see at the 3 prime site there is a poly A binding protein at the 5 prime end there is a complex that is eukaryotic initiation factor 4 F complex and that complex contains 4 E 4 G and also 4 A. 4 A is the helicase is having the helicase function which helps in translation.

Now, this complex helps in association of eukaryotic initiation factor 3 which helps in attachment with the 40 S ribosome this pink is the 40 A ribosomal sub unit. And instead of the signed alguernos sequence in case of eukaryote the identification of 5 prime AUG the prime the first codon is basically identified by scanning. So, the whole reading frame after forming this complex is scanned by this eukaryotic initiation factor 4 F complex mostly held by the factor 4 B. Next coming to elongation, elongation there is binding of an incoming aminoacyl tRNA. So, initially the first for myelomethionine has already been attached to the P site.

Now, we have a vacant A site where the charge tRNA is carried by a complex of GTP bound elongation factor T U it binds in the A site. So, here you can see in the A site the second amino acid AA 2 has bound. After binding what happens the GTP is hydrolyzed and the elongation factor T U GDP is released. So, this T U GDP is released and it again can be regenerated to T U GTP with the help of another elongation factor that is elongation factor T S.

T S basically exchanges the GDP with the GTP. So, in the P site we have the first amino acid methionine in the A site we have the second amino acid whatever that can be. Now, the next stage is formation of the bond peptide bond between those two amino acid. Now, what happens the first amino acid that is N-formyl methionine group of the amino acid it transferred to the tRNA of the second amino acid present in the A site. So, here you can see there is a transfer of the this blue is basically the formyl methionine group. So, it has been transferred and there is a bond formation with the peptide bond formation with the help of the enzyme peptidryl transferase.

So, here you can see a tRNA hybrid binding state is formed. So, basically what happens this suppose this is our 30 S ribosome this is the P site and this is the A site. So, the alignment is in this way. So, if we draw this even if this is our P site and this is our A site the A site tRNA is basically reaching the approaching the P site and there is a hybrid status is formed.

Similarly, this P site tRNA is aligned to the E site. So, this hybrid now needs to be modified for that what we need is translocation. Translocation forms the realignment how the ribosome moves one codon towards the 3 prime end of the mRNA. So, to relieve that stress ribosome moves towards the 3 prime end in such a way that the here you can see the A site tRNA is now situated over the P site. So, the second amino acid is taking place or in case of the A site not the second amino acid rather the third amino acid because the second amino acid is already attached to the dipeptidyl tRNA 2. So, similarly because of the shifting what happens this AUG because you can see this is the direction of the ribosome movement.

So, basically the ribosome has moved in such a way that the third amino acid falls here in the A site the second amino acid from the A site has shifted to P site and the very AUG codon carrying the tRNA has occupied now the third site that is the E site. So, here the uncharged tRNA because the first tRNA has already delivered its formylmethionyl group over the second amino acid it is now occupying the E site. Now, this movement of the ribosome is known as translocation and is helped by an enzyme translocase and that property is present in the elongation factor G which is helped by GTP hydrolysis. Now, the deacetylated tRNA or the uncharged tRNA now is released from the E site this happened in prokaryotes. In case of eukaryotes the elongation factors are of different in present in different name that is eukaryotic elongation factor 1 alpha 1 beta gamma and 2 they represent elongation factor Tu Ts and G respectively and in case of eukaryote there is no E site.

So, there is direct expulsion of the uncharged tRNA. Now, termination the termination signal is given by the termination codon. So, the mRNA or the ribosome keeps on moving till in A site there is one termination codon UAG and that termination codons are identified by termination factors. So, termination factors in case of prokaryotes are releasing factors releasing factors 1 and 2, 1 recognizes the termination codon UAG and UAA whereas, 2 recognizes UGA and UAA. There are 3 termination codon, but the releasing factor 3 its function is not properly identified. Once the releasing factors occupy the A site in the termination codon the peptidyl transfer is what it what it does it transfers all the growing polypeptide chain not to another amino acid rather it transfer it to the water molecule.

So, basically there is you can see there is release of the growing polypeptide chain. So, this is our mRNA and all the factors now are released and tRNA also released and here you can see the growing polypeptide chain is also released. Now, in case of eukaryote there is one releasing factor only eukaryotic releasing factor that takes part in releasing or terminating the chain elongation process. Now, coming to post translational modification the nascent proteins basically biologically inactive till it gets its proper

secondary tertiary or quaternary structure. So, that is done by protein folding along with that there is different N and C amino and carboxy terminal modifications.

The nascent polypeptide chain can be cleaved to shorter forms sometimes the inactive form which is known as zymogen can that inactive form the polypeptide chain is cleaved in such a way that the active form is generated. There can be formation of multiple disulfide bond intra chain in a single polypeptide chain like this the disulfide bonds can be formed the one very important example is insulin which has intra chain disulfide bond along with that there can be some covalent modification of amino acid by addition of different groups like phosphate groups which is known as phosphorylation. You know that there are different enzymes which are activated or inactivated by adding the phosphoryl group there is hydroxylation extensive hydroxylation is present in collagen, carboxylation carboxyl group added in different proteins like our coagulating factors which are activated by carboxylation also different proteins have glycosyl ends and also prosthetic groups are attached in different enzyme for the activation of that protein. Also all these are post translational modification present in a protein then there are different types of inhibitors drugs which inhibit the protein synthesis. Now, here are examples of such drugs like tetracycline streptomycin these antibiotics basically binds with the 30 S ribosomal subunit and it causes the aminoacyl tRNA synthetase activity is blocked by this enzyme also the first initiation codon methionine can be misinterpreted the reading can the misread the code can be misread by these drugs.

In case of chloramphenicol chloramphenicol basically it remember all these drugs are related with the proteins in the translation inhibition. Then chloramphenicol it binds with binds with the 50 S subunit erythromycin and cyclohexamide they prevents the translocation. So, basically the growing chain cannot be elongated because there is no translocation and then puromycin. So, puromycin basically represents as amino acid analog because in the A site puromycin can be attached and in that case premature chain termination of the translation can occur. So, this is all about the translation or the protein synthesis process.

So, if we come to the summary that activation of amino acid is done by the very important enzyme aminoacyl tRNA synthetase. The first amino acid methionine is represented by AUG it binds in the P site of the ribosome elongation occurs via peptide bond formation the enzyme is peptidyltransferase termination signal is given by the termination codons and also there are releasing factors which is helping in this process. There are multiple initiation factor and elongation factors as well then protein for being biologically active there are different types of modification extensive folding as well as covalent and non covalent modification for the biological activation of the protein. So, this is all about translation and proteins these are my references. Thank you. And, see you in the next class.