2

Molecular biology sheet

Doctors 2022 أثر | medicine | MU

DOCTOR:

Prof. Samir Mahgoub

DONE BY:

Ibrahim Amjad

CORRECTED BY:

Layan Husam
Rahaf Mohamed

Protein Synthesis

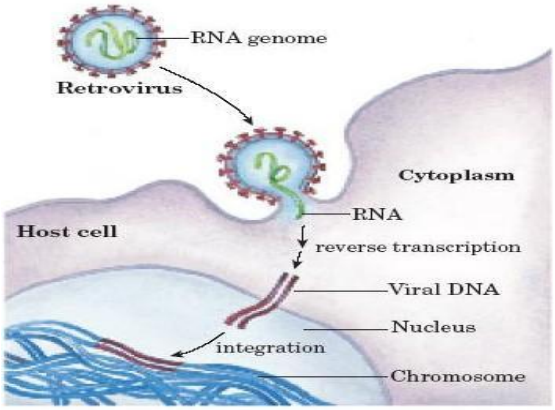
Many thanks to زوج batch each by their name for making our life easier ☺.

Last lecture we talked about transcription; which is synthesizing mRNA from DNA template.

Today's lecture (reverse transcription) is the exact opposite, which happens in certain organisms.

Reverse transcription

The genetic information carrier of some biological systems is ssRNA (single stranded) instead of dsDNA (double stranded) (such as ssRNA viruses) → reverse transcription happens in such systems. Retrovirus: viruses that contain the genetic material RNA instead of DNA and it is ssRNA→ such as HIV (human immunodeficiency virus), the one responsible for AIDS (acquired immune deficiency syndrome).



The virus will first attach itself to the surface of the host cell, then it will inject its genetic material into it, now we have two scenarios: 1) The genetic material of the virus is RNA: no problems will happen as the virus will directly attach its generic material into it to become one unit, then the host will order protein synthesis for the virus. 2) The genetic material of the virus is single strand RNA: there is a problem as a single strand can't bind a double strand no matter what (the figure):

- **The information flow is from RNA to DNA, opposite to the normal process. This special replication mode is called reverse transcription.**
- **HIV has an RNA genome that is duplicated into DNA.**
- **The resulting DNA can be merged with the DNA genome of the host cell. If the genetic material of the host cell consists from double strand DNA and the virus genetic material consists from single strand RNA so no integration can occur, until the virus genetic material (ssRNA) converts to dsDNA by reverse transcriptase to fuse and integrate with host genetic cell.**
- **The main enzyme responsible for synthesis of DNA from an RNA template is called reverse transcriptase (RT), it has a major problem (to the host) of lacking (no)the proof-reading ability (can't correct any errors that may happen while functioning→ gives it a huge variety). That's why if someone contracted influenza viruses (most of them are retroviruses), it is almost the same of contracting 200,000 different types of viruses all- together. Similarly, if someone has AIDS it is the same of contracting 11- 12 billion different viruses→ can't be cured in both cases. It has the following activities (3 in 1 virus):**

1. **RNA-dependent DNA polymerase activity: It uses RNA template to synthesize the first strand**

of the DNA which is called complimentary DNA cDNA.

2. DNA- dependent DNA polymerase activity: the complimentary DNA synthesized by #1 will be its template and it will use it to synthesize the 2nd strand, then
3. RNase or ribonuclease H activity will disintegrate the RNA as it will no longer be needed.

To sum up, the virus will attach itself to the host surface, injects its genetic material, if it's RNA the reverse transcriptase enzyme will use its activities to synthesize the 1st strand (number [#]1 activity), then the 2nd strand (#2 activity) will be synthesized using the freshly made complimentary DNA. Now, RNA will no longer be needed (degraded by #3 activity) and a dsDNA is formed which can be integrated by another enzyme called integrase into the host DNA. How many enzymes have we mentioned so far? Only 2 not 4 (RT with 3 activities and integrase). Note that #2 activity is behind the inability to find some viral pathologies' treatments as we mentioned.

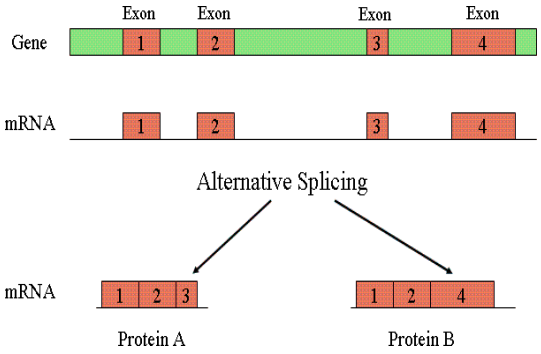
- In the case of HIV, reverse transcriptase is responsible for synthesizing a complementary DNA strand (cDNA) to the viral RNA genome.
- An associated enzyme, ribonuclease H, digests the RNA strand, and reverse transcriptase synthesizes DNA complementary strand to form a double helix DNA structure.
- This DNA is integrated into the host cell's genome by integrase enzyme causing the host cell to generate viral proteins that reassemble into new viral particles. However, in retroviruses, the host cell remains intact as the virus buds out of the cell but in the case of HIV, the host cell undergoes apoptosis.
- Some eukaryotic cells contain an enzyme with reverse transcription activity called telomerase enzyme. Telomerase is a reverse transcriptase that lengthens the ends of linear chromosomes. Telomerase carries an RNA template from which it synthesizes DNA repeating sequence, or "junk" DNA the parts at the ends of a DNA molecule, which is usually 6 bp (base pairs) long. This repeated sequence of DNA is important because, every time a linear chromosome is duplicated in cell division, it is shortened in length, if it happens so many times (may reach 20 thousand hundred times), a great loss will happen. So, telomerase makes sure to constantly compensate the junk DNA parts which will do nothing if lost but protection.
- Telomerase is often activated (highly active) in cancer cells (almost 100% more active than normal cells), to enable cancer cells to duplicate their genomes indefinitely and rapidly without losing important protein-coding DNA sequence (the discovery of RT enriches the understanding about the cancer-causing theory of viruses, where cancer genes in RT viruses, and HIV having RT function). Activation of telomerase could be part of the process that allows cancer cells to become immortal (shows immortal feature [can't die], as seen in an X person who has a small cancerous lesion, if this lesion has RT activity what can we see after let's say 6 months? A huge growth in size due to the extreme RT activity).

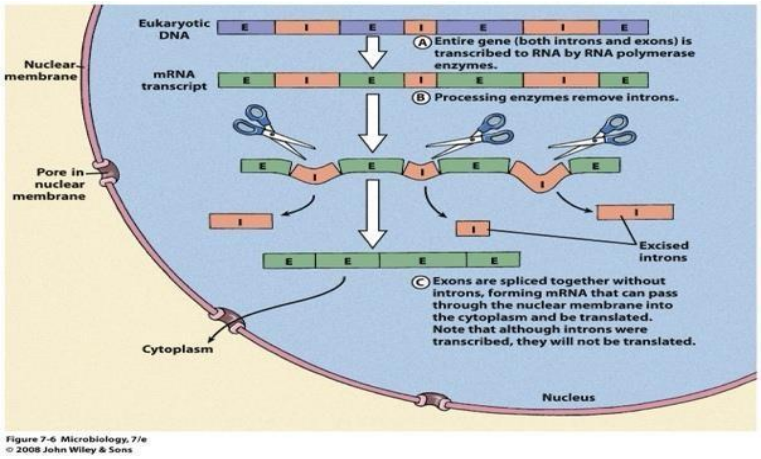
Alternative splicing (eukaryotes only):

The number of genes in our bodies are up to 50 thousand genes but the number of proteins is much

more, this is due to the alternative splicing feature, where splicing – as we know- means 1) removal of introns (introns are removed) and 2) joining of exons (exons are "coding" regions). The thing is, we have many options when it comes to joining exons together:

In the left following figure, a gene (DNA) has 5 exons (green colored) and 4 introns (orange colored), after forming the mRNA the introns were removed, now the remnant exons can join together to form many different proteins: protein A: 1, 2& 3, protein B: 1, 2& 4, protein C:1& 2 only, protein D: 1, 2, 3& 4, protein E: 2& 4 only, protein F: 1& 3 only, protein G: 3& 4 only… etc. So, an infinite (endless) # of proteins possibilities are there due to alternative splicing. (If the introns were not removed, when the strand goes to ribosome, it won't read the introns and will just read the exons which form a disconnected (intermittent) protein → the strand will be longer which causes to be degraded easily by enzymes that found in cytoplasm so that is why the life span of long RNA molecules is shorter than other RNAs).





Note that the formed proteins don't necessarily have to have the same or similar features even if they were made from the same exons. For e.g.:

1. Calcitonin hormone (released from thyroid gland [along with T3 and T4] where it acts on calcium balance; it ↓ it when it is ↑ to avoid developing hypercalcemia and consequently cardiac arrest).
2. Parathyroid hormone [PTH] (released from parathyroid gland and plays a role in calcium metabolism too).
    - Genetic material in thyroid gland, in parathyroid gland, in hair cells, in skin cell, in bone cells ..etc are all the same! The thing is, a certain gene will be active (switched on) in some places while inactive (switched off) in other ones. That being said, Calcitonin coding gene will only be active in thyroid gland and PTH coding gene will only be active in parathyroid gland regardless their presence in every other cell in our bodies, forming a different product like protein for smell sense in brain [has a major role in CNS].

To sum up, the same genetic material can be found in all our cells but there are specified cells that are "on" on some cells and "off" on the other cells. Each gene will yield more than 1 protein; the # of proteins is many many many more than the # of genes.

- Different combinations of exons form different mRNA resulting in multiple proteins

from the same gene, that being said, humans have 30 to 50 thousand genes but are capable of producing 100,000 proteins.

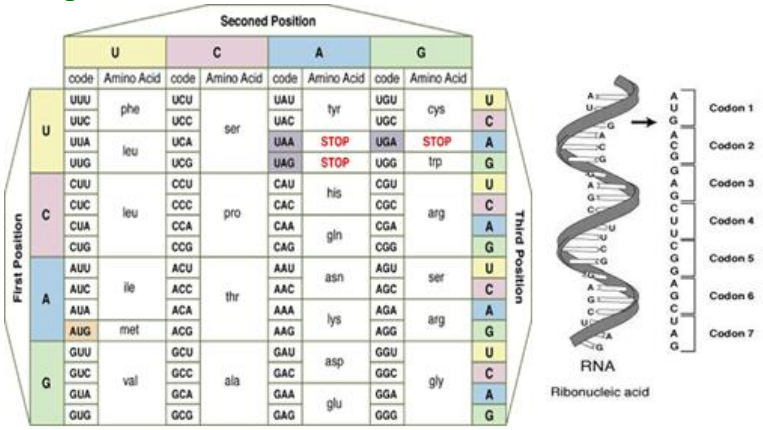
Genetic Code:

-It is the set of rules by which information encoded in genetic material (DNA or mRNA sequences) is translated into proteins (amino acid sequences) by living cells.

-With some exceptions, a triplet (formed of 3 bps or ribonucleotides) codon (we have a total of 64 possible different codons) in a nucleic acid sequence specifies a single amino acid (AA).

-Because the majority of genes are encoded with exactly the same code, this particular code is often referred to as standard genetic code, though in fact there are many variant codes, for example, protein synthesis in human mitochondria relies on a genetic code that differs from the standard genetic code.

- We have 3 stop codons (UAA, UAG and UGA) and 1 initiation or start codon (MOSTLY methionine [met] AA with AUG code), so there are 60 codons left from the 64 ones that will code for 19 AAs (met is already counted out). Note that some AA have more than one codon, for e.g., his (histidine) has 2 codons, gly (glycine) has 4 codons and arg (arginine) has 6 codons. So, more than 1 AA have more than 1 genetic codon, and genetic codons # ≠ # of AAs→ it ↓ the mutations chance.



Characteristics of genetic codes (codons): very important

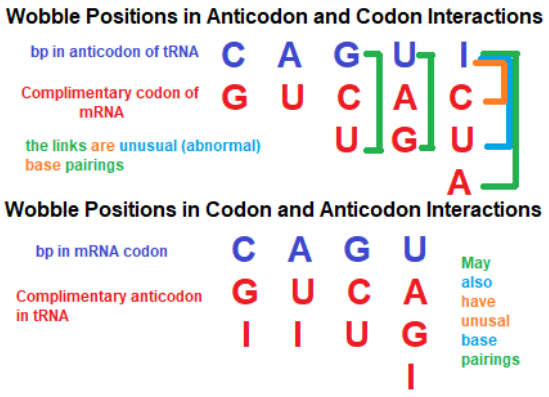
1. The genetic code is composed of ***nucleotide triplets***. In other words, three nucleotides in mRNA (a codon) specify one amino acid in a protein. No 2 AAs can have the same genetic codon. In other words, some AAs have more than one genetic codon but genetic codons can't be shared between any 2 AAs, never ever. The 64 codons are called standard genetic codes.

2. The code is ***non-overlapping***. This means that successive triplets are read in order and each nucleotide is part of only one triplet codon, in other words, each nucleotide is a part of 1 genetic codon and it is not shared to any other codon. The genetic code is read in groups (or "words") of three nucleotides. After reading one triplet, the "reading frame" shifts over three letters, not just one or two. In the following example, the code would not be read GAC, ACU, CUG, UGA... Rather, the code would be read GAC, UGA, CUG, ACU...

3. **The genetic code is *degenerate or redundant* (many genetic codons for the same AA). In contrast, some amino acids can be specified by more than one codon. There are 64 different triplet codons, and only 20 amino acids, there are 64 different triplet codons: 3 of them are stop (non-sense) codons and one initiation codon, for the 19 AAs there is 60 standard codons. Unless some amino acids are specified by more than one codon, some codons would be completely meaningless, therefore, some redundancy is built into the system: some amino acids are coded for by multiple codons, but it must be having a single genetic codon. In some cases, the redundant codons are related to each other by sequence; e.g.: (check the table below) leucine is specified by CUU, CUA, CUC, and CUG. The codons are the same except for the 3rd nucleotide position. This third position is known as the "wobble" position of the codon. This property allows some protection against mutation, if a mutation occurs at the third position of a codon, there is a good chance that the amino acid specified in the encoded protein won't change. Note that it doesn't mean that mutations can't occur in the 1st or 2nd position, the thing is, the 3rd position is kind of protected. (understand the general idea; it will be repeated with more details).**

Standard Genetic Codes for Leucine AA:		Second Position U		It's a part of the full table in page #4	
		Code	Amino Acid		
First Position	C	CUU	Leu (leucine essential AA)	U	Third position (Wobble Position)
		CUC		C	
		CUA		A	
		CUG		G	

This is because in a number of cases, the identity of the base at the third position can wobble (move shakenly= change), and the same amino acid will still be specified.



- This figure is important, the 4 abnormal nucleotides that are found ONLY in tRNA (thymine, pseudouridine, dihydrouridine and inosine [I]). I can be found in the anticodon region; the complimentary one that will make base pairing with mRNA.
- Abnormal (unusual) base pairing is evident; the more we have them between anticodon of tRNA and codon of mRNA, the lower the mutations.

4. **The genetic code is *unambiguous*. Each codon specifies a particular amino acid, and only one amino acid. In other words, the codon ACG codes for the amino acid threonine, and only threonine and never shared with any other AA.**

5. **The code is *nearly universal*. Almost all organisms in nature (from bacteria, parasites,**

protozoans, fungi, animals, plants and to humans) use exactly the same nuclear genetic code (nuclear DNA) that specifies same AAs. The rare exceptions include some changes in the code in mitochondria (mitochondrial DNA), and in a few protozoan species. Some differences:

Criterion/ type of DNA	Nuclear DNA	Mitochondrial DNA
Size (in base pairs)	3.5×10^9	16569 (almost 16 K)
Inheritance	Mendelian inheritance	Maternal inheritance (only from the mother)

But why are there differences between species (let's say brothers, in color, hair, body build...etc) if the source is the same (same mother)? The rule says: the source of DNA is the same (sure thing) but there is variation in almost 1% of the total genes, which is responsible for these differences among humans. That being said, there is a different between nuclear and mitochondrial genetic codons and the rest is universal.

Reading Frames: this is also important

Translation starts @5' end of mRNA and ends @3' of it. However, it won't start until the ribosomes read the initiation or start codon (AUG of methionine) and won't stop until it reads a stop codon (UAA, UAG or UGA). Each strand of DNA will have 3 reading frames to be read a base-by-base, but the only one to be translated into a protein is the open reading frame; the one that starts with the initiation codon and ends with a stop codon without being interrupted in the middle by any other stop codons.

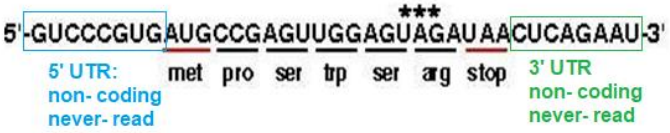
- If you think about it, because the genetic code is triplet based, there are three possible ways a particular message can be read, as shown in the following figure:

- Clearly, each of these would yield completely different results. Genetic messages work much the same way: there is one reading frame that makes sense (called the open reading frame), and two reading frames that are nonsense.
- The code contains signals for starting and stopping translation of the code. The start codon is AUG, which codes for methionine and encountered signals for translation to begin and subsequent triplets are read in the same reading frame. Translation continues until a stop codon is encountered. There are three stop codons: UAA, UAG, and UGA. To be recognized as a stop codon, the triplet must be in the same reading frame as the start codon. A reading frame between a start codon and an in-frame stop codon is called an open reading frame.
- Translation can take place considering the following sequence:

  5'-GUCCCGUGAUGCCGAGUUGGAGUCGAUAACUCAGAAU-3'

- First, the code is read in a 5' to 3' direction. The first AUG read in that direction sets the reading frame, and subsequent codons are read in frame, until the stop codon, UAA, is encountered.

- **In the following example:**



- **We have 37 bps, by simple math 37/3= 12, a 12 AA protein as a translation result. However, it is not the case in here. It will only form a 6 AA protein as ribosomes will only start translating at met (AUG) codon and will stop right before UAA. Moreover, in this sequence, there are nucleotides at either end that are outside of the open reading frame. Because they are outside of the open reading frame, these nucleotides are not used to code for amino acids. This is a common situation in mRNA molecules, where the region at the 5' end that is not translated is called 5' untranslated region (5' UTR) (it is always before the start codon) and at the 3' end is called the 3' untranslated region (3'UTR) (it is always after the stop codon). These sequences, even though they do not encode any polypeptide sequence, are not wasted: in eukaryotes these regions typically contain regulatory sequences that can affect when a message gets translated, where in a cell an mRNA is localized, and how long an mRNA lasts in a cell before it is destroyed. At the 5' UTR (un translated region) we have the boxes (prokaryotes 2 boxes and eukaryotes 3 boxes) while at the 3' UTR we have terminators (in prokaryotes Rho dependent and Rho- independent of mRNA synthesis while in eukaryotes are not well- understood).**

Different Types of Codons Positions:

According to the position of a codon we have different four type of degenerate- sites; think of degenerate as # of repetitions (different genetic codes) we can use to get the same AA:

1) **A position of a codon is said to be a *fourfold degenerate site or synonymous* if any nucleotide at this position specifies the same amino acid, e.g. the third position of the glycine codons (GGA, GGG, GGC, GGU) is a fourfold (4 genetic codons yielding the same AA) degenerate site, because all nucleotide substitutions at this site are [synonymous]; i.e., they do not change the amino acid. So, only the third positions of some codons may be fourfold degenerate. A change in position 3 form a different genetic codon, yet it specifies the same amino acid. No mutations occur (0% mutation and 100% protection always).**
2) **A position of a codon is said to be a *twofold degenerate site* if only two of four possible nucleotides at this position specify the same amino acid. For example, the third position of the glutamic acid codons (GAA, GAG) is a twofold (2 genetic codons yielding the same AA) degenerate site. There is 50% possible for the mutation causing by changing 3rd position (with C or U) and 50% protection.**
3) **A position of a codon is said to be a *non-degenerate site* if any mutation at this position results in amino acid substitution (any change in the 3rd position of the genetic codon will give a genetic codon for another AA, 100% mutation and 0% protection)→ significant mutation risk.**

4) **There is only *<u>one threefold degenerate</u>* site where changing to three of the four nucleotides may have no effect on the amino acid (depending on what it is changed to), while changing to the fourth possible nucleotide always results in an amino acid substitution. This is the third position of an <u>isoleucine</u> codon: AU<u>U</u>, AU<u>C</u>, or AU<u>A</u> all encode isoleucine, but AU<u>G</u> encodes the start codon methionine. (AA who has 3 genetic codons, but only one change will yield a mutation 75% protection and 25% mutation).**

To sum up, what are the characters of genetic codons?

1) **Triple nucleotides.**
2) **Redundant or degenerate (many codons for the same AA).**
3) **Non- overlapping.**
4) **Unambiguous.**
5) **Nearly universal except for mitochondrial DNA (for e.g., a genetic code of UUU may translate to phenylalanine in nuclear DNA but to methionine in mitochondrial DNA).**

Protein Synthesis:

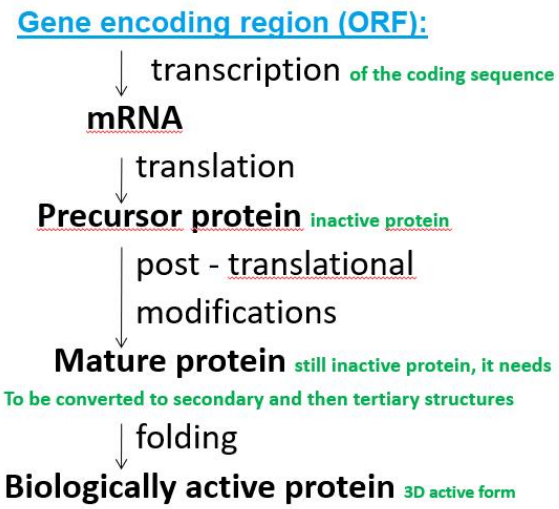
- **It starts with mRNA binding to ribosomes -in both prokaryotes and eukaryotes- at <u>mRNA binding site or groove</u>, the mRNA should be properly oriented; translation should start @ 5' end of mRNA which will represent the amino terminal end of the resultant protein molecule and it ends @ the 3' end of mRNA which will represent the carboxyl terminal of this protein. However, there is a certain space before the initiation codon, at which some proteins may start translation to form some AAs before met, those AAs are not needed and they may even affect the function of the produced protein→ it should be removed by proteolytic enzymes activity. For e.g., for zymogen (pepsinogen), a # of AAs should be removed to activate it into pepsin. Similarly, trypsinogen to trypsin and chymotrypsinogen (-2 dipeptides) to chymotrypsin and so on; it's called <u>trimming.</u> So we have: <u>Preproprotein</u>: immature protein, Pre is represented by what we called <u>signal sequence</u>, and pro is represented by what we called <u>inhibitory sequence</u> like <u>proenzyme,</u> two sequences (peptides) will be found.**
- **Protein synthesis is the process in which cells build proteins (a multi-step process, beginning with amino acid synthesis and transcription of nuclear DNA into messenger RNA, which then decoded by the ribosome to produce proteins).**
- **When a protein must be available on short notice or in large quantities, a protein precursor is produced (<u>proprotein</u>). A proprotein is an inactive protein containing one or more inhibitory peptides that can be activated when the inhibitory sequence is removed by proteolysis during posttranslational modification <u>(trimming)</u>. A <u>preproprotein</u> is a form that contains a <u>signal sequence (an N- terminal signal peptide)</u> this signal sequence specifies its insertion into or through membranes, i.e., targets them for secretion. The signal peptide is cleaved off in the endoplasmic reticulum. Preproproteins have both sequences (inhibitory and signal) still present. Throughout the whole journey (ribosomes [protein synthesizing machinery]→smooth ER→Golgi Apparatus). It's very important for the secretion of protein, an enzyme called <u>signal peptidase</u> (removes) cleaves off the <u>signal peptide (sequence)</u> in**

endoplasmic reticulum from preproprotein [can't be activated] and converts it to proprotein [can be activated], this conversion must happen before going to Golgi Apparatus's secretory vesicles to be stored. The inhibitory sequence remains with both (preproprotein and proprotein both have one or more inhibitory sequences yet only the 1st one has the signal peptide).

- For synthesis of protein, a succession of tRNA molecules charged with appropriate amino acids have to be brought together with a mRNA molecule and matched up by base-pairing through their anti- codons with each of its successive codons.
- The amino acids then have to be linked together to extend the growing protein chain, and the tRNAs, relieved of their burdens, have to be released.
- These whole complexes of processes are carried out by the ribosome, formed of two main chains of rRNA, and more than 50 different proteins.

Gene Expression: the whole story

- **Transcription**: is synthesis of an RNA that is complementary to one of the strands of DNA.
- **Translation:** when ribosomes read a messenger RNA and make protein according to its instruction.



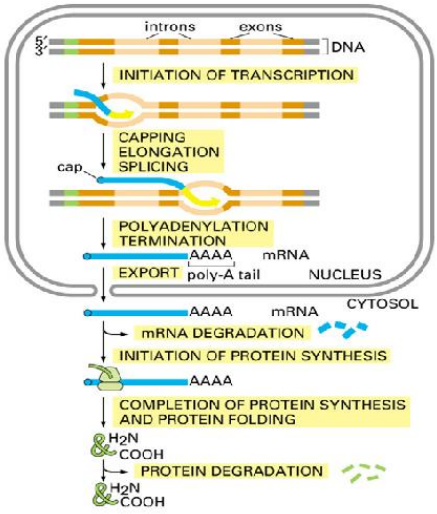
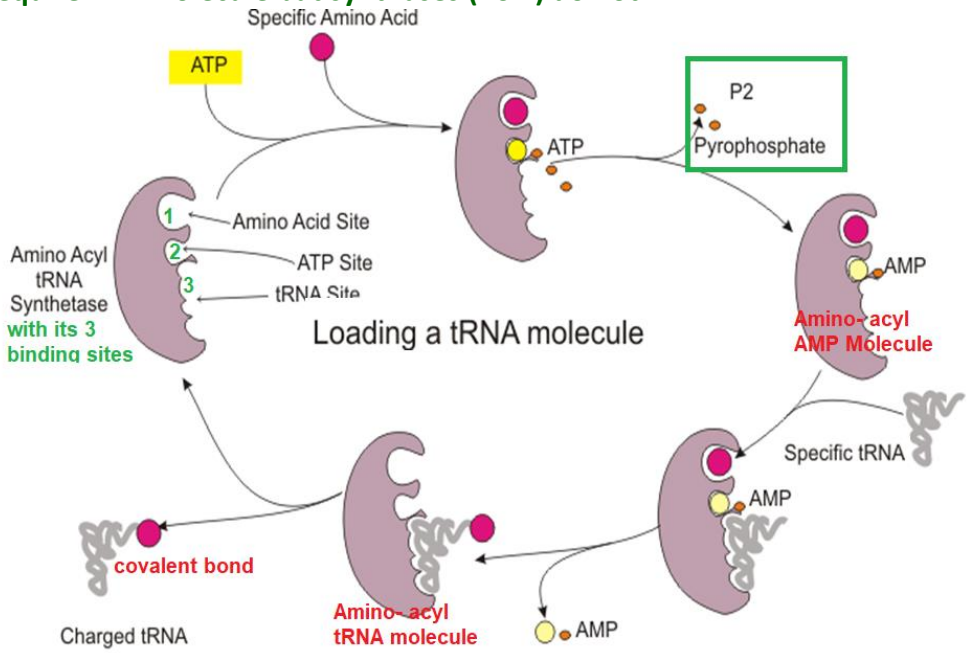


Figure 6–90. Molecular Biology of the Cell, 4th Edition.

The 1st step in protein synthesis and the only one that happens in the cytosol is called:

1. **Charging (loading) the tRNA (binding specific AAs to specific tRNA to participate in protein synthesis):**
   - tRNA acts as a translator between mRNA and protein, each tRNA has a specific anticodon and an amino acid acceptor site (arm), this arm, must be identified by the addition of CCA as we discussed in the previous lecture so that AAs can bind the 3' end of tRNA to be transferred to the protein synthesizing machinery → it regulates protein synthesis by only allowing the correct AA to bind the acceptor site of the correct tRNA.
   - Each tRNA also has a specific charger protein (aminoacyl tRNA synthetases) which can

only bind to that particular tRNA and attach the correct amino acid to the acceptor site. The energy to make this bond comes from ATP. The general rule is: each AA has a specific enzyme→ we have @ least 20 different enzymes. Note that synthetase (with T) require ATP molecule but synthases (no T) do not.



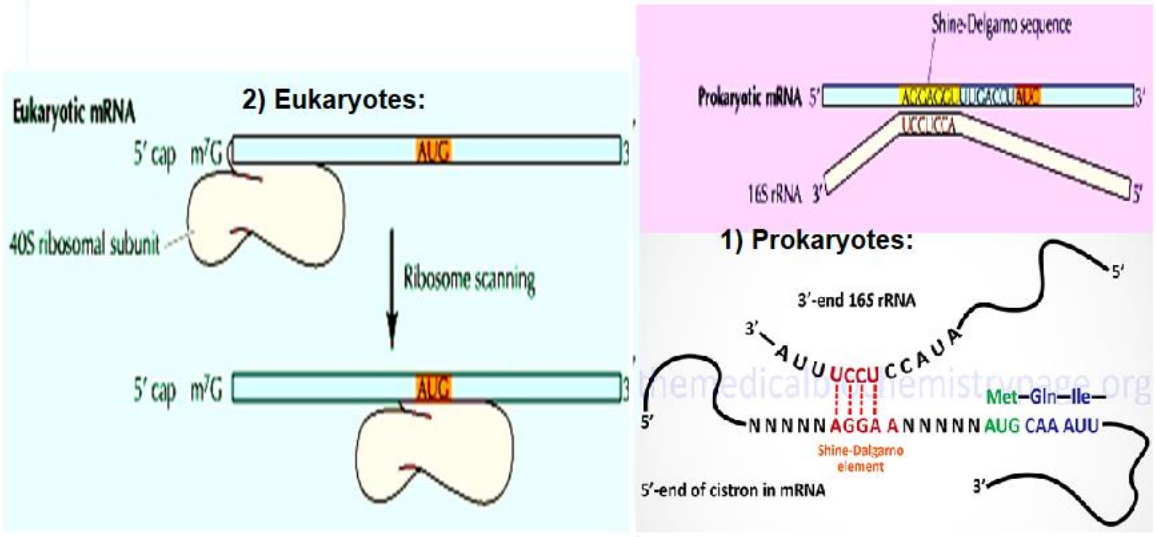
- It is done by binding 1) AMP to AA on the enzyme then by 2) adding tRNA to the complex molecule. It ensures that this suitable amino acid is bounded to the acceptor site of tRNA after the hydrolysis of ATP molecule (as 2 inorganic phosphates [= 1 pyrophosphate, a high energy molecule that is equal to breaking 2ATPs to 2ADPs] was degraded, it is considered a high regulation reaction. This binding process is irreversible, so it should be done with high fidelity, in order to guarantee a correct complementarity between the codon and anti-codon with the intended amino acid; which will determine the way the reaction will take: a correct protein synthesis or a wrong protein synthesis.
- Aminoacyl tRNA synthetase is a one enzyme that catalyzes two reactions: a) activation of AA: by AMP binding to AA to form the 1st complex molecule. B) Charging (loading) tRNA: now it removes AMP after the specific tRNA is successfully bound to form the second molecule.
- There are 20 different synthetases one for each amino acid that can catalyze the covalent bond between the amino acid and tRNA
- A single synthetase may recognize multiple tRNAs for the same amino acid specified by the mRNA codon to which the tRNA anticodon binds (more than one genetic code for the same AA more than one tRNA [each carry an anticodon complimentary to the genetic codon on mRNA], however, the main determinant of synthetases # remains the AA).
- Two classes of synthetases, differ in
- 1) the 3-dimensional (3D) structures (configurations),

- **2) which side of the tRNA they recognize and 3) how they bind ATP**
  - ❖ **Class I – monomeric (single polypeptide chain), acylates (recognizes) the 2'OH on the terminal ribose and it will only form a tertiary protein structure.**
    Arg, Cys , Gln, Glu, Ile, Leu, Met, Trp, Tyr, Val
  - ❖ **Class II – dimeric (double polypeptide chains), acylates (recognizes) the 3'OH on the terminal ribose and it will form a quaternary protein structure.**
    Ala, Asn, Asp, Gly, His, Lys, Phe, Ser, Pro, Thr
- **Wobble: it is defined as a change in the 3rd positioned- ribonucleotide in a genetic codon, this change limits the # of mutations that may happen in this position. The result will be mRNA to be translated later on in eukaryotes, what about prokaryotes? Nope, remember that both transcription and translation occur simultaneously in it; since it does not have a nuclear envelope to separate the two processes.**
  - **In prokaryotes: mRNA binds the small subunit of ribosome (remember, the whole ribosome is 70S, the large subunit is 50S while the small one is 30S. While in eukaryotes; 80S, 60S and 40S, respectively).**
  - **If there was one tRNA for each mRNA codon, there would be 61 different tRNAs but there are fewer. Some tRNAs have anticodons that recognize 2 or more different codons**
  - **Base pairing rules between the third base of a codon and its tRNA anticodon are not a rigid as DNA to mRNA pairing, for e.g., U in tRNA can pair with either A or G in the third position of an mRNA codon. This flexibility is called wobble**
  - **There are two levels of control to ensure that the proper amino acid is incorporated into protein through:**

1- The reaction of amino acyl tRNA synthetase for charging the proper tRNA
2- Matching the specific tRNA to a particular codon of mRNA

- **In prokaryotes: mRNA translation will start by its binding to a 16s rRNA (remember, small subunit of ribosomes in prokaryotes is formed of a 16S rRNA and 21 proteins). Now, specific sequences in the mRNA around the AUG codon, called Shine – Delgarno sequences (know it), are recognized by an initiation complex consisting of a Met amino-acyl tRNA, initiation factors (IFs) and the small ribosomal subunit, this sequence at 5' end of the mRNA will be identified by a complimentary sequence at 3' end of the 16s rRNA. Lastly, the ribosomes will now know where the 5' end is and translation starts. Remember, in prokaryotes there is no capping, poly A nor splicing→ posttranscriptional processing of RNA.**
- **In eukaryotes: no Shine- Delgarno sequence, how can ribosomes detect the 5' of mRNA? there is a process called ribosome scanning, where mRNA binds the 40S ribosomal subunit then it will be moving along the small subunit of ribosome till finding the codon of initiation (AUG) of methionine to be located in the P site. (Ribosomes have 3 special sites: P, A and E, to be discussed later).**

Detection of 5' end of mRNA in:

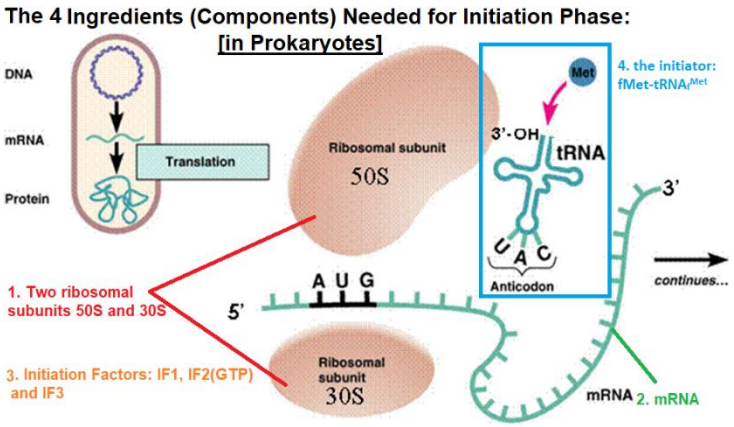


2. Translation: translation now begins with its 3 steps: a. initiation, b. elongation and c. termination.

a. Initiation:

- The only stage in protein synthesis that takes place on a small ribosomal subunit. That beings said, it can't work on the whole ribosome and dissociation must occur. So, large subunits must leave to allow the initiation of protein synthesis.
- This phase of protein synthesis results in the assembly of a functionally competent ribosome in which an mRNA has been positioned correctly so that its start codon is positioned in the P (peptidyl) site and is paired with the initiator tRNA.
- The following ingredients are needed for this phase of protein synthesis:

  1- Two ribosome subunits: 30S and 50S (prokaryotes) 2- The mRNA 3- Three Initiation Factors - IF1, IF2 (GTP, Guanosine Triphosphate, ATP will only be used in charging tRNA after that all the processes that need energy will use GTP) and IF3 and 4- The initiator molecule or initiator AA: fMet-$tRNA_f^{Met}$: [formyl/methionine {methionyl}- transfer RNA] in the prokaryotic cells (formyl group is required to recognize the 5' end direction of mRNA [along with Shine- Delgarno sequence]) while in eukaryotes there is no need for formyl group due to the presence of the cap [5' cap in ribosomal scanning, so Met-$tRNA^{Met}$ (methionine [or methionyl] tRNA) is used instead. (extra: formyl is the most basic aldehyde −CH=O)



- **The following steps take place in initiation: very important with its figures**
  1. **In prokaryotes:**
     - **A- Binding of the ribosome 30S subunit with initiation factor (IF3)1) promotes the dissociation of the ribosome into its two component subunits.**

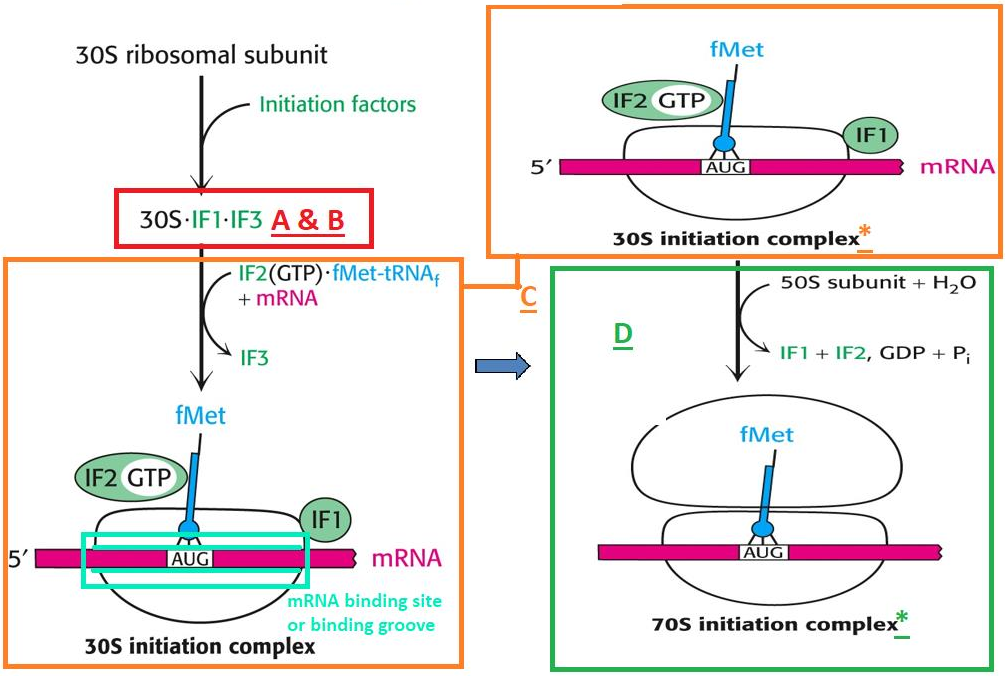
     - **B- The presence of IF3 2) permits the assembly of the 30S initiation complex (helps binding the mRNA on the mRNA binding site or groove on the 30S subunit) and 3) prevents binding of the 50S subunit prematurely (before the initiation is done), IF1 1) assists IF3 in some way, perhaps by increasing the dissociation rate of the 30S and 50S subunits of the ribosome.**
       The first IFs to arrive are IF1 and IF3; to separate the small ribosomal subunit by dissociating the two subunits (as initiation can't happen on the whole ribosome).

     - **C- Binding of the mRNA and the fMet-tRNA$_f^{Met}$:**
       - IF3 assists the mRNA to bind with the 30S subunit of the ribosome so that the start codon is correctly positioned at the peptidyl site (P- site) of the ribosome.
       - The mRNA is positioned by means of base-pairing between the 3' end of the 16S rRNA with the Shine- Dalgarno sequence immediately upstream of the start codon. (Shine- Delgarno sequence @ the 5' end of mRNA is needed to make complementarity between it and the complimentary sequence @ the 3' end of the 16S rRNA→ this leads to positioning the AUG of methionine on the P- site; where it's always the starting point). Shine- Delgarno sequence is needed her as prokaryotes lack the 5' cap that allows mRNA in eukaryotes to move along the ribosome till it reaches the p- site in the ribosomal scanning process.
       - IF2(GTP) assists the fMet-tRNA$_f^{Met}$ to bind to the 30S subunit in the correct site - the P site. This will lead to inserting the tRNA inside the P- site to make complementarity between the genetic codon of mRNA for met (AUG) and anticodon of tRNA also for met (i.e. it binds the mRNA with tRNA after the mRNA is successfully placed on the P- site), tRNA itself is bound to IF2[GTP]. At this stage of assembly, the 30S initiation complex is complete and IF3 can dissociate.

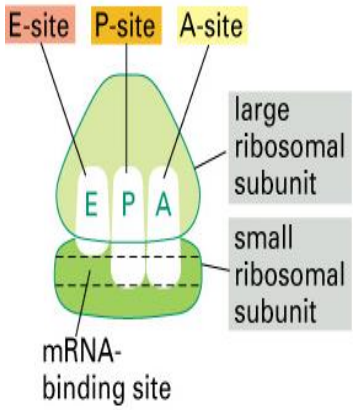
D- Binding of the ribosome 50S subunit and release of Initiation Factors:

- Three events now happen "simultaneously". As the 50S subunit of the ribosome associates with the 30S initiation complex, 1) GTP hydrolysis occurs on IF2 (this hydrolysis will release energy as GTP decomposes into GDP +inorganic phosphate, the energy will mostly be used to 2) uncouple the remaining IFs (1 and 2), this uncoupling will lead to re- associating the large and small subunits together [when the re-association happens it means that initiation phase is done]). This hydrolysis may be helped by the 3) L7/L12 ribosomal proteins rather than by IF2 itself (those proteins help in the re- association and the release of Ifs from the 30S ribosomal subunit). After all these events happen, the 70S initiation complex is formed.

Initiation Phase in Prokaryotes:



In addition to the APE sites (P- site in the middle, A- site on its left and E- site on its right) there is an mRNA binding groove (site) that holds onto the message being translated. The A site (accepting site) binds an aminoacyl-tRNA (a tRNA bound to an amino acid, the newly coming one); P site (peptide side) binds a peptidyl-tRNA (a tRNA bound to the peptide being synthesized). The E site (exit site) binds a free tRNA before it exits the ribosome (so, it is for releasing the uncharged tRNA after tRNA leaves the AA @ the P- site it is no longer needed exits the ribosome).



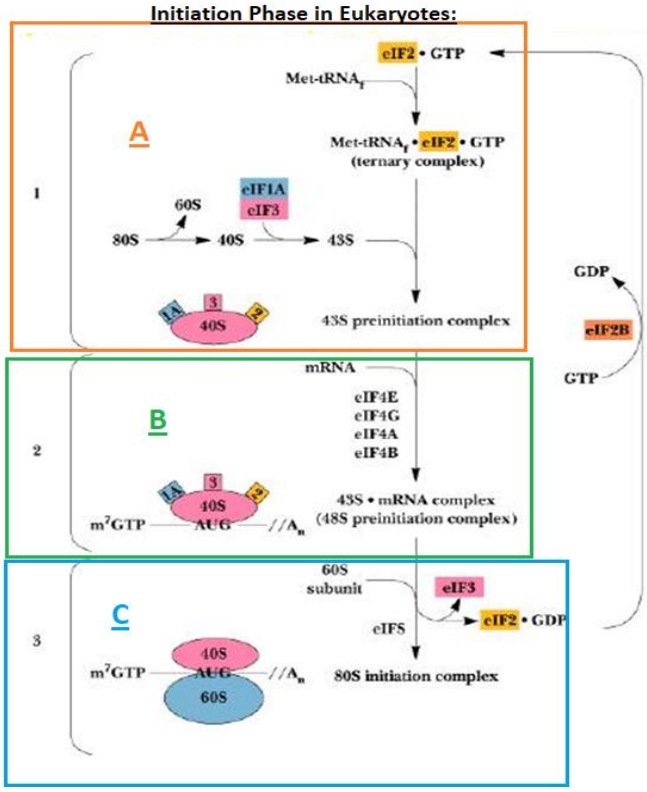
2. **In Eukaryotes: very important with its figures**

The process is a bit more complicated in here than in prokaryotes.

A- Eukaryotic initiation factors 1A and 3 (eIF1A and eIF3) will bind the small subunit in ribosomes (80S) to dissociate the small subunit (40S) from the large one (60S). eIF1A and eIF3 have sedimentation rates of their own→the small subunits' sedimentation rate will change from 40S to 43S. @ the same time, eIF2 with its GTP group will bind Met-tRNA$_f$ to form one ternary complex (composed of 3 units: eIF2.GTP, Methionine and tRNA). Now the 43S will bind the ternary complex to form a 45S preinitiation complex (the eIF2 increased the sedimentation rate from 43S to 45S).

B- mRNA will now bind the 45S preinitiation complex @ the mRNA binding site with the help of 4 initiation factors: eIF4E, eIF4G, eIF4A and eIF4B. The sedimentation rate will also change due to those factors from 45S to 48S, and the product will be called 45S.mRNA complex or 48S preinitiation complex. (So, the sedimentation rate changed from 40S to 45S and lastly to 48S).

C- Now, GTP molecule that is bound to eIF2 will go under hydrolysis to form: GDP+ inorganic phosphate+ energy. This energy will be used in: 1) uncoupling all the eIFs since they are no longer needed (the sedimentation rate will be restored to 40S consequently) and 2) re-associating the small and large subunits together in an 80S initiation complex.



Eukaryotes use a scanning mechanism to initiate translation (due to presence of 5' cap; the mRNA will move on the ribosome till it reaches the P- site. Moreover, this cap is the reason behind not having a formyl group in Met-tRNA$_f$ here).
Recognition of the AUG triggers GTP hydrolysis by eIF-2.
GTP hydrolysis by eIF2 is a signal for binding of the large subunit and beginning of translation

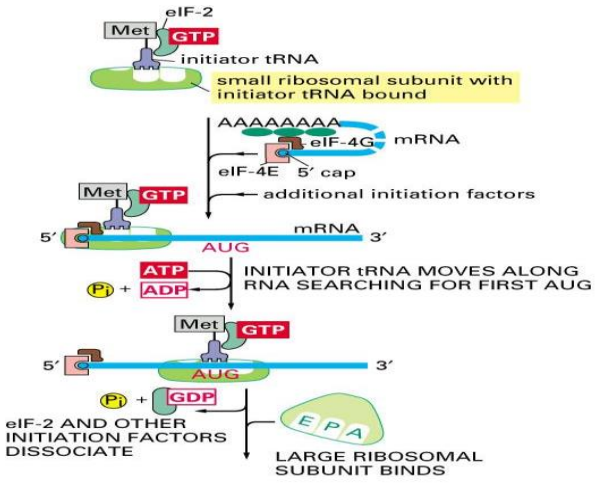
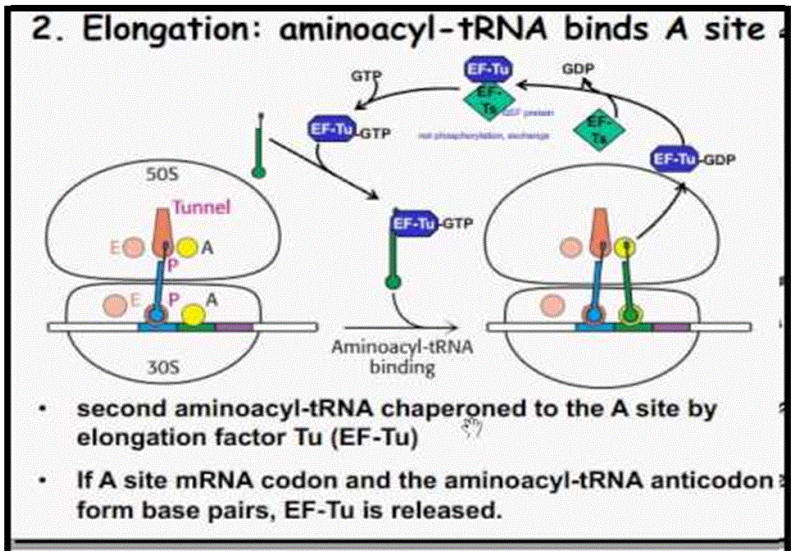


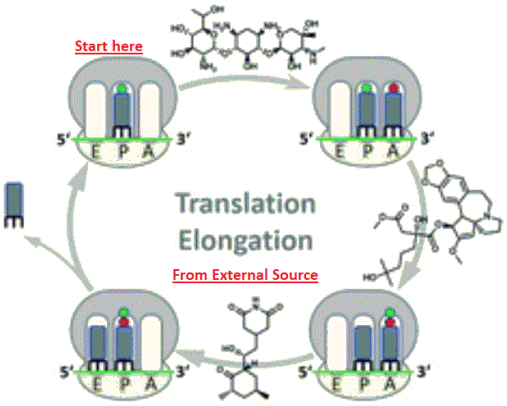
Figure 6–71 part 1 of 2. Molecular Biology of the Cell, 4th Edition.

b. **Elongation: (2nd step of translation)**
   - **Three special Elongation Factors are required for this phase of protein synthesis: EF-Tu (GTP) the most important one, EF-Ts (no GTP) and EF-G (GTP).**
   - **The Elongation phase of protein synthesis consists of a cyclic process whereby a new aminoacyl-tRNA (tRNA loaded with an amino acid) is positioned in the ribosome, the amino acid is transferred to the C-terminus of the growing polypeptide chain, and the whole assembly moves one position along the ribosome (towards the E- site).**
   - **A new codon is now positioned at the A site and awaits a new aminoacyl-tRNA.**
   - **Binding of a new aminoacyl-tRNA at the A site.**
   - **At the start of each elongation cycle:**
     - **The A (aminoacyl) site on the ribosome is empty, The P (peptidyl) site contains a peptidyl-tRNA (bound to a methionine AA @ the 1st cycle), and the E (exit) site contains an uncharged tRNA (empty @ the 1st cycle), and it leaves the E- site @ the beginning of any new cycle.**
     - **A specific genetic codon in the A- site with the help of the elongation factor, EF- Tu (GTP) binds with an aminoacyl-tRNA (the AA that will be able to recognize the genetic code) and brings it to the ribosome @ the A- site. (now I have 2 tRNAs: one with already- formed AA on the P- site and one with a newly- coming AA on A- site). Now, an enzyme called Peptidyl transferase enzyme remove the AA from p site and transfer it to A site to form peptide bond, here the p site has uncharged tRNA and the A site has charged tRNA by two amino acid linked to each other by peptide bond. Then each tRNA will take a step forward (as the mRNA will move along the ribosome in a process called translocation): making the uncharged tRNA in the E- site instead of the p- site and the charged tRNA in the P- site instead of the A- site. A new genetic codon will show in the A- site and the process will be repeated…**
     - **Once the correct aminoacyl-tRNA is positioned in the ribosome, GTP is hydrolyzed, EF- Tu (GDP) undergoes a conformational change and then dissociates away from**

the ribosome.

- There are two ways that EF- Tu functions to ensure that the correct aminoacyl-tRNA is in place (regulation of elongation for the 1st time): 1) EF- Tu prevents the aminoacyl end of the charged tRNA from entering the A site on the ribosome; This ensures that mRNA codon- tRNA anticodon pairing is checked first before the charged tRNA is irreversibly bound in the A site and a new, potentially incorrect, peptide bond is made. 2) GTP hydrolysis is SLOW and EF- Tu cannot dissociate from the ribosome until it occurs, more time for further checking.
- The amount of time prior to GTP hydrolysis allows the final fidelity check to take place.
- Hydrolysis is associated with a conformational change in EF-Tu.

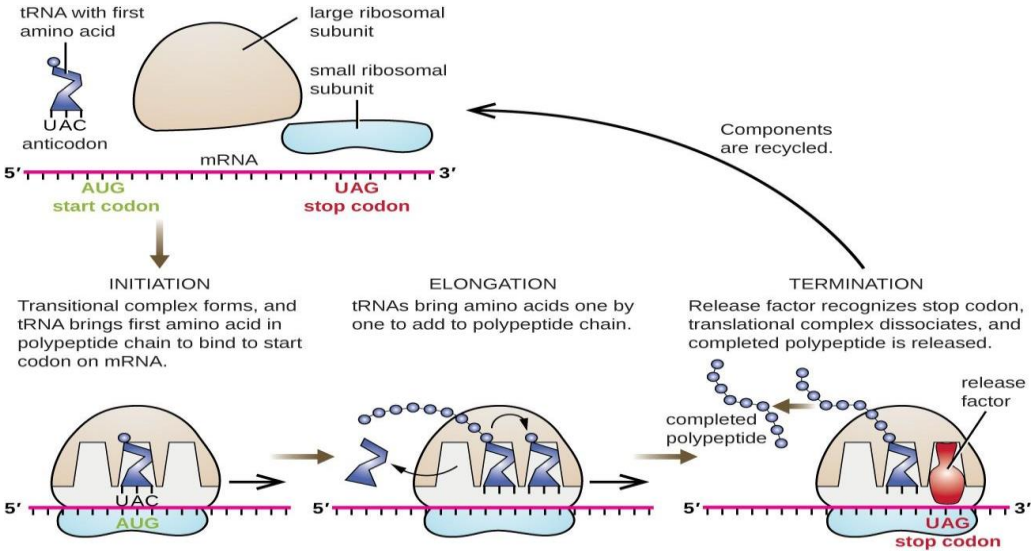




c. Termination: (3rd and last step in translation)

- The final phase of protein synthesis requires that the finished polypeptide chain be detached from a tRNA. This can only happen in response to the signal that a stop codon has been reached the A- site (since there is no tRNA anti- codons for mRNA stop codons, releasing (termination) factors are needed. Those factors differ in prokaryotes (the following 3) and eukaryotes.
- Binding of Release factors:
  - There are no tRNAs that recognize the stop codons.
  - Rather stop codons are recognized by release factor RF1 (which recognizes the UAA and UAG stop codons) or RF2 (which recognizes the UAA and UGA stop codons). Note that UAA code is recognized by both
  - These release factors act at the A site of the ribosome. -
  - A third release factor, RF3 (GTP), stimulates the binding of RF1 and RF2 in the A- site.
- Hydrolysis of the peptidyl-tRNA (polypeptide chain linked to tRNA)

- **Binding of the release factors alters the peptidyltransferase activity with a nucleophilic effect.**
- **The result is hydrolysis of the peptidyl-tRNA and release of the completed synthesized polypeptide chain. by adding H2O molecule rather than amino acid.**
- **The uncharged tRNA in the E site can dissociate as can the release factors.**
- **GTP is hydrolyzed without dissociating tRNA in the P site.**
- **Then the process of initiation, elongation and termination will be repeated as needed.**



- **Antibiotic (AB) Inhibitors of Protein Synthesis:** very important

TABLE 29.4 Antibiotic inhibitors of protein synthesis

Antibiotic	Action
Streptomycin and other aminoglycosides	Inhibit initiation and cause misreading of mRNA (prokaryotes) makes ribosomes read mRNA from 3' instead of 5'
Tetracycline	Binds to the 30S subunit and inhibits binding of The sec aminoacyl-tRNAs (prokaryotes) specified in (A)site
Chloramphenicol	Inhibits the peptidyl transferase activity of the 50S ribosomal subunit (prokaryotes) no peptide bonds formation so no elongation
Cycloheximide First in eukaryotes	Inhibits the peptidyl transferase activity of the 60S ribosomal subunit (eukaryotes)
Erythromycin	Binds to the 50S subunit and inhibits translocation (prokaryotes) mRNA can't move tRNA won't move no more codes @ A- site
Puromycin common in both	it resembles the tyrosine ,No peptide bonds amino acid Causes premature chain termination by acting as an analog of aminoacyl-tRNA (prokaryotes and eukaryotes)

Diphtheria toxin it's not antibiotic — Inhibits eEF-2 by ADP-ribosylation of modified histidine in the factor

Adds a ribose to the histidine used in forming eukaryotics elongation factor- 2

Antibiotics Inhibiting Translation: also, very important

- The bacterial ribosomal structure and the accessory functions differ in many respects from its eukaryotic equivalent. The translation reaction itself can be subdivided into three parts:

1. Formation of the initiation complex, blocked by Streptomycin (inhibits initiation and cause misreading of mRNA; stop codons will be recognized initially and no protein synthesis will happen) and Tetracyclins; inhibiting binding of aminoacyl-tRNA to the ribosomal A- site at the 30S ribosomal subunit, which means no elongation.
2. Introduction of aminoacyl-tRNA and synthesis of a peptide bond, inhibited by puromycin (it has a tyrosine- like structure→ no peptide bonds are formed between AAs→[termination of protein synthesis (premature termination)] and chloramphenicol (probably inhibiting the peptidyltransferase). Both puromycin and chloramphenicol inhibit the elongation step.
3. Translocation of the mRNA relative to the ribosome (no mRNA translocation→no introduction of the newly- coming AA on tRNA in A- site). Blocked by erythromycin and fusidic acid (the latter preventing release of EF- G/GDP, generally after GTP hydrolysis to GDP, inorganic phosphate and energy, some conformational changes happen to elongation factors (EF) to be released, fusidic acid prevents these conformational changes from happening→ EF with GDP will constantly be bound to tRNA and won't be released. Both erythromycin and fusidic acid block the elongation step.

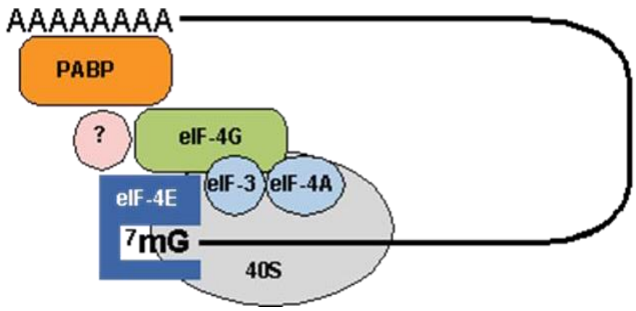
Protein Synthesis in Eukaryotes:

- A major difference between eukaryotes and prokaryotes is that, in a typical eukaryotic cell, protein synthesis takes place in the cytoplasm while transcription and RNA processing take place in the nucleus. In bacteria, these two processes can be coupled so that protein synthesis can start even before transcription has finished (no nuclear envelope to separate them so both happens in cytoplasm.
- The steps of protein synthesis are basically the same in eukaryotic cells as in prokaryotes. The ingredients, however, can be different:
  1- Ribosomes are larger. 60s and 40s subunits combine to give 80s ribosomes which contain 4 RNAs: 28S,5.8S and 5S in the 60S subunit; 18S in the 40S subunit. (in prokaryotes: 70S= 50S+ 30S, with 3 rRNAs of 23S, 16S and 5S).
  2- While the initiating amino acid in eukaryotic protein synthesis is still methionine, it is not formylated (in prokaryotes it is formylated methionine).
  3- Eukaryotic mRNA is capped. This is used as the recognition feature for ribosome binding (ribosomal scanning)—not the 18S rRNA. (in prokaryotes using Shine-Delgarno sequence).
  4- The initiation phase of protein synthesis requires over 10 eukaryotic initiation factors (eIFs) one of which is the cap binding protein (we have studied 7 eIFs: [eIF1A and eIF3: work on rRNA], [eIF2: works on tRNA] and [eIF4E, eIF4G, eIF4A and eIF4B: work on mRNA through its binding groove]. In prokaryotes we only have 3 IFs: IF1, IF2 and IF3.
  5- The eukaryotic elongation phase closely resembles that in prokaryotes (both are

3). The corresponding elongation factors are eEF-1a (EF-Tu.GTP), eEF- 1bg (EF-Ts NO GTP) and eEF-2 (EF-G.GTP).

6- Eukaryotes require just a single release factor, eRF. In prokaryotes there are 3: RF1 recognizes UAA and UAG stop codons, RF2 recognizes UAA and UGA stop codons and RF3(GTP) that stimulates RF1 and RF2 biding on the A- site of ribosome.

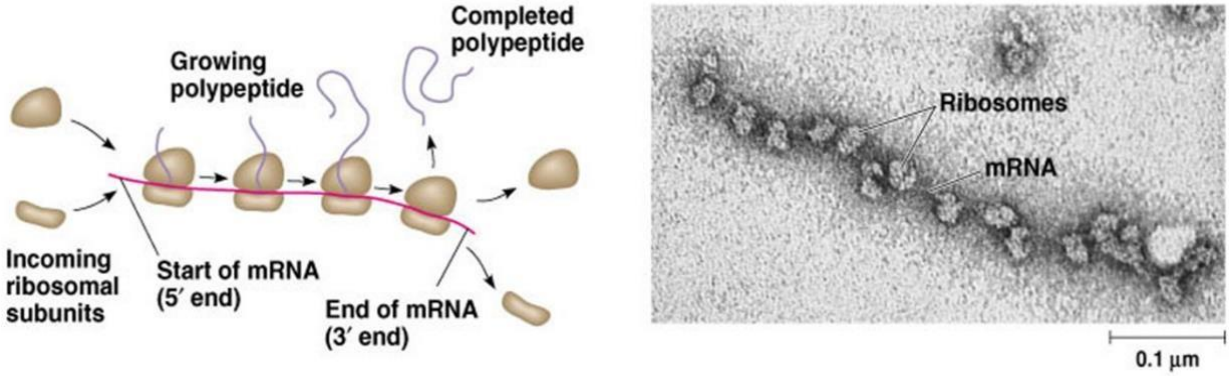
- Coordinating protein synthesis with mRNA synthesis:
- It has recently been found that the eukaryotic initiation factor eEF- 4G (also called cap binding protein→ recognizes 5' of mRNA) binds not only with other factors in the initiation complex but also with PABP (poly A binding protein) which binds to the poly A tail of mRNA after splicing is done→recognizes 3' end of mRNA.
- It is thought that the binding of eEF-4G to PABP serves as a critical recruitment step for driving downstream translation (from 5' end to 3' end).
- In another sense, however, the binding of eEF-4G to PABP represents a mechanism to ensure that only mature intact mRNAs are translated.



- Most mRNA are translated by more than one ribosome at a time; the result, a structure in which many ribosomes translate a mRNA in tandem, is called a polysomes (means the mRNA can be translated by more than one ribosome @ the same time as each ribosome will translate a portion, this process is done for long mRNAs for reducing the required time).

- **Post-translational modifications**
- **They are the chemical modifications of a protein after its translation.**
- **Characterized by:**

1. **Being numerous and diverse.**
2. **Able to change the charge (add acidic or basic amino acid), conformation(outside) or size**



(a) An mRNA molecule is generally translated simultaneously by several ribosomes in clusters called polyribosomes.

(b) This micrograph shows a large polyribosome in a prokaryotic cell (TEM).

(number of amino acid) of protein molecule.

- **Effects:**

1. **Stability of protein.**
2. **Biochemical activity (activity regulation).**
3. **Protein targeting (protein localization).**
4. **Protein signaling (protein - protein interaction).**

- **Insulin for example: important**

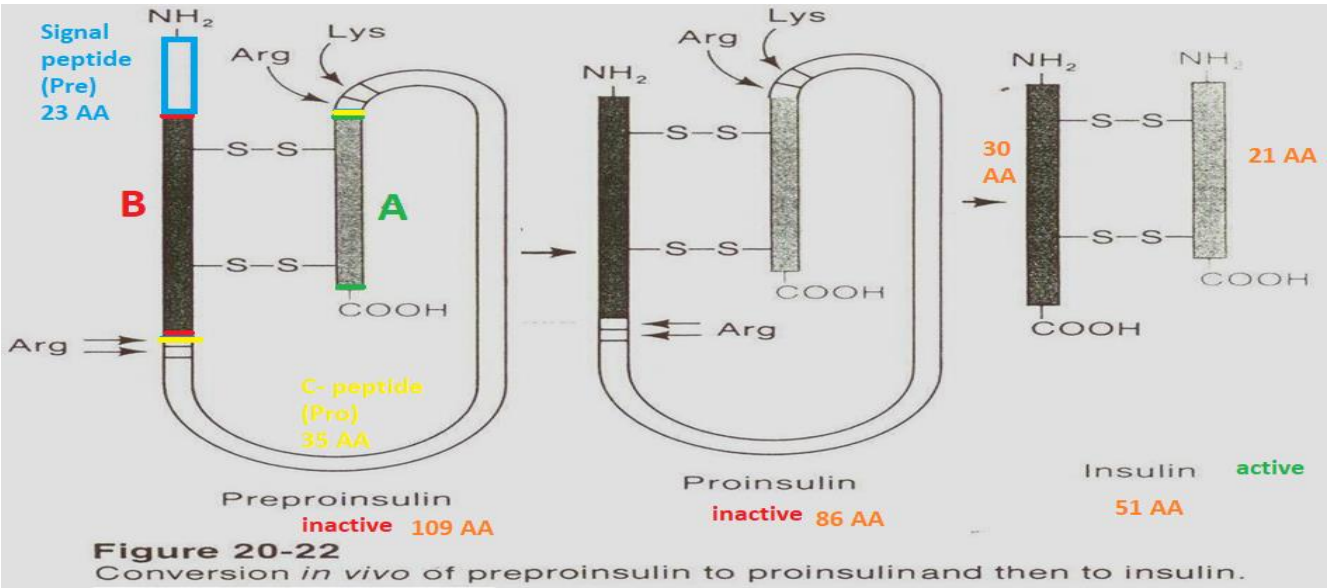


Figure 20-22
Conversion *in vivo* of preproinsulin to proinsulin and then to insulin.

- **insulin, which is a low molecular weight protein having two polypeptide chains (A and B) which fold to allow interchain and intrachain disulfide bridges. A specific protease (the process is enzymatically driven with also proteinases and endonucleases) then clips out the segment that connects the two chains which form the functional insulin molecule (A and B). 1st we have the preproinsulin where a signal peptide is removed in the endoplasmic reticulum (ER), the we have a proinsulin in which a propeptide is removed from the middle of the chain (this removal happens after it's secreted from Golgi Apparatus to the secretory vesicles); the resulting protein consists of two polypeptide chains connected by disulfide bonds. Trimming modification occurs here. (the AAs #s are from the doctor).**

- **Other modifications examples:**
  - **Also, most nascent polypeptides, the initial methionine {the first amino acid in protein synthesis} usually taken off during post-translational modification by specific aminopeptidases, not well- understood yet.**
  - **Other modifications, like phosphorylation, are part of common mechanisms for controlling the behavior of a protein, for instance activating or inactivating an enzyme.**
  - **Some animal viruses, as poliovirus (responsible for infantile paralysis) and hepatitis A virus, synthesize long polycistronic proteins {the protein that is synthesized by: more than one gene with related functions are transcribed into one mRNA, the mRNA is then translated into one polypeptide chain, which will be separated into more than one protein (polycistronic proteins)} from one long mRNA molecule, these proteins must be cleaved at specific sites to provide the several specific proteins required for viral function.**
  - **Collagen, an abundant protein in the extracellular spaces of higher eukaryotes, is synthesized as procollagen three polypeptide molecules (to be crosslinked with desmosomes), that align themselves in a particular way that is dependent upon the existence of specific amino terminal peptides. Specific enzymes then carry out hydroxylation and oxidation of specific amino acid residues within the procollagen molecules to provide cross-links for greater stability with cleavage of the NH2 terminal end to form a strong, insoluble collagen molecule.**

- **Types of posttranslational modifications:**

A- Trimming: removal of inhibitory sequence

- **Many proteins secretion from the cell are made as:**
  1. **large precursor molecules but functionally inactive.**
  2. **Change of protein from non-active for active molecule by removing portions of the protein chain by specialized endoproteases.**
- **Sites of the cleavage reaction:**
  1. **Endoplasmic reticulum.**
  2. **Golgi apparatus.**
  3. **Secretory vesicles.**

- **N.B. zymogens which are inactive enzymes (such as chymotrypsinogen), become activated (to chymotrypsin) through cleavage (-2 dipeptides or 4 AAs) when they reach their proper sites of action.**

B- Covalent alterations:

- **Proteins may be activated or inactivated by the covalent attachment of a variety of chemical groups:**

1. **Phosphorylation:**
   - **By adding phosphate group to the hydroxyl groups of (serine, threonine, tyrosine residues in a protein) which is catalyzed by protein kinases (adds phosphate) and reversed by protein phosphatases (removes phosphate)**
   - **The phosphorylation (and de- phosphorylation) may ↓ or ↑ the functional activity of protein Phosphate group 1) some enzymes will be active by phosphorylation 2) some enzymes will be inactive by phosphorylation**
2. **Glycosylation: binding carbohydrate molecules to proteins, for e.g., glycoprotein**
   - **Many of proteins → become part of a plasma membrane, lysosomes or secreted from the cell have carbohydrate chains attached to serine or threonine hydroxyl groups (O-linked) or the amide nitrogen of asparagine (N-linked)**
   - **Occurs: in the endoplasmic reticulum and Golgi apparatus.**
   - **Used to: target protein to specific organelles**
3. **Hydroxylation: Adding OH group**
   - **As proline and lysine in endoplasmic reticulum by prolyl or lysyl hydroxylases (e.g. in collagen), in order to form cross linkages.**
   - **In scurvy (vitamin C deficiency), the collagen in patients' tissues is deficient in hydroxyproline since vitamin C, ascorbic acid, is required to hydroxylate proline residues in procollagen.**
4. **Other Covalent Modifications:**
   a. **Carboxylation:**
      - **Carboxyl groups can be added to glutamate residues by vitamin K**
      - **The resulting carboxyglutamate residues are essential for the activity of several of the blood-clotting proteins**
   b. **Biotinylated enzyme:**
      - **Biotin (vitamin B7) is covalently bound to the amino groups of lysine residues of biotin-dependent enzymes that catalyze carboxylation reactions.**
      - **Such as: pyruvate carboxylase requires biotin catalyst.**
   c. **Farnesylated Protein:**
      - **Help anchor proteins in membranes.**
      - **Note: many proteins are acetylated.**

- **farnesyl group is lipid in nature (intermediate of cholesterol synthesis) so it helps in binding specific proteins with cell membrane.**