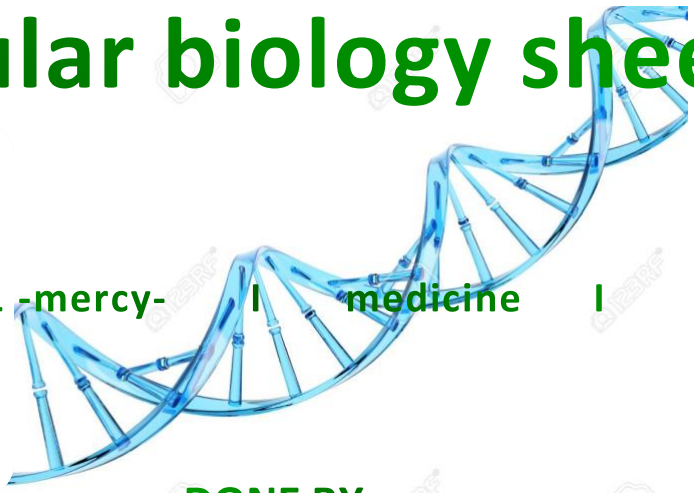
3

Molecular biology sheet



Doctor 2021 -mercy- | medicine | MU

DONE BY:

Shahd Alayouben

Shadeen Btoush

Aya Abu Samra

Baraa Ahmad

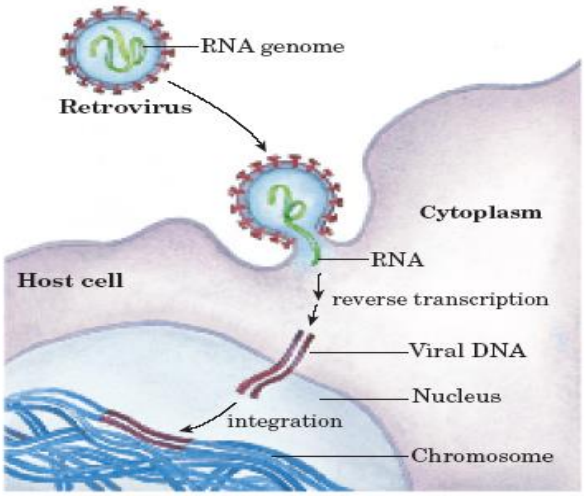
CORRECTED BY:

Emran Younis

DOCTOR

Dr. Sameer Mahjoub

Reverse transcription



- **The genetic information carrier of some biological systems is ssRNA. instead of dsDNA (such as ssRNA viruses).**

Retrovirus :Viruses that contain the genetic material RNA instead of DNA and it is ssRNA

- **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 dsDNA and the virus genetic material consists from ssRNA so can't occur any integration so, the virus genetic material 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 the following activities:**

1- RNA-dependent DNA polymerase 2- RNase
3- DNA-dependent DNA polymerase

The reverse transcription enzyme :

It use RNA template used for synthesis the first strand of the DNA called it : RNA dependent DNA polymares , and then the same template that comes from RNA templet will be used as a tamplet for synthesis the second strand DNA(dsDNA) .

بالمرحلة الاولى عند تعريف

RNA -dependent- DNA polymares is needed to synthesis the first strand of DNA : using RNA tamplet

المرحلة الثانيه :

DNA- dependent- DNA polymares because it will use this activity for synthesizing the second DNA strand of dsDNA from the first DNA molecule

It occurs this mechanism in the viruses that have the activity RTase but , the RTase doesn't have proof-reading that causes the difficulty to find the treatment for viral pathology like : HIV

- **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.**
- **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.**
- **This repeated sequence of DNA is important because, every time a linear chromosome is duplicated, it is shortened in length.**

- **Telomerase is often activated in cancer cells to enable cancer cells to duplicate their genomes indefinitely 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.**

Alternative splicing (eukaryotes only)

The number of gene ≠ the number of protein , and the number of proteins is double of gene because the alternative splicing .

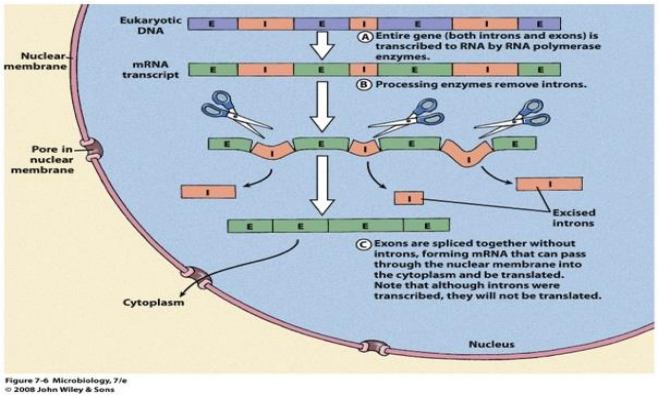


Figure 7-6 Microbiology, 7/e
© 2008 John Wiley & Sons

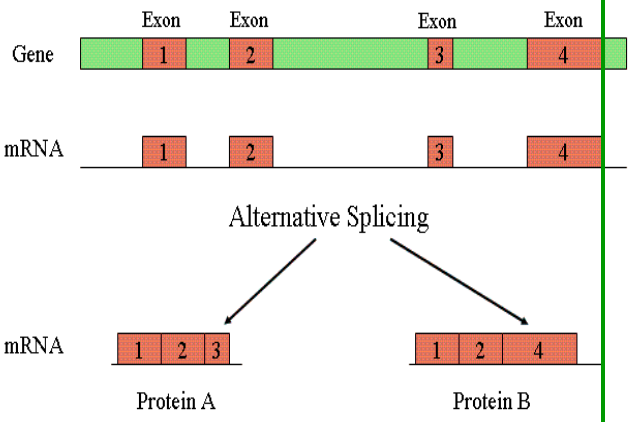
- Exons are “coding” regions

- Introns are removed

If the introns doesn’t remove :

-when the strand goes to ribosome , it won’t read the introns and read just the exons which form disconnected protein .

- the strand will be longer which causes to be degraded easily by enzymes that found in cytoplasm so that why the life span of long RNA molecule is shorter than other RNA .



- Different combinations of exons form different mRNA resulting in multiple proteins from the same gene
  - - Humans have 30 to 50 thousand genes but are capable of producing 100,000 proteins.

Genetic code

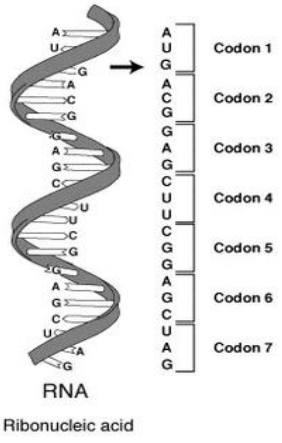
It is the set of rules by which information encoded in genetic material (DNA or mRNA sequences) is translated into proteins (amino acid sequences) by living cells.

- With some exceptions, a triplet codon in a nucleic acid sequence specifies a single amino acid.

- 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.

First Position	Seconed Position U		Seconed Position C		Seconed Position A		Seconed Position G		Third Position
	code	Amino Acid	code	Amino Acid	code	Amino Acid	code	Amino Acid	
U	UUU	phe	UCU	ser	UAU	tyr	UGU	cys	U
	UUC		UCC		UAC		UGC		C
	UUA	leu	UCA		UAA	STOP	UGA	STOP	A
	UUG		UCG		UAG	STOP	UGG	trp	G
C	CUU	leu	CCU	pro	CAU	his	CGU	arg	U
	CUC		CCC		CAC		CGC		C
	CUA		CCA		CAA	gln	CGA		A
	CUG		CCG		CAG		CGG		G
A	AUU	ile	ACU	thr	AAU	asn	AGU	ser	U
	AUC		ACC		AAC		AGC		C
	AUA		ACA		AAA	lys	AGA	arg	A
	AUG	met	ACG		AAG		AGG		G
G	GUU	val	GCU	ala	GAU	asp	GGU	gly	U
	GUC		GCC		GAC		GGC		C
	GUA		GCA		GAA	glu	GGA		A
	GUG		GCG		GAG		GGG		G



Characters of genetic code:g

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.**

There are 20 amino acids that enter in the synthesising of one protein, also the number of genetic codes are 64 so that explains that there some amino acids have more than one genetic codon .

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.

- **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...

Because each molecule is in one genetic codon, therefore a ribonucleotide can't be in more than one genetic codon.

- **Rather, the code would be read GAC, UGA, CUG, ACU...**

3- The genetic code is degenerate. 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 codon: 3 of them are stop (non-sense) codons and one initiation codon.

- 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.

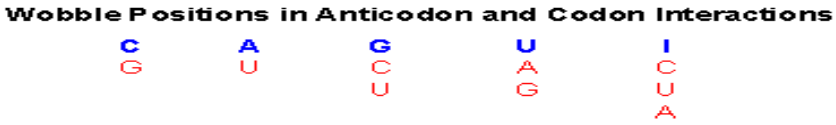
- In some cases, the redundant codons are related to each other by sequence; e.g., leucine is specified by CUU, CUA, CUC, and CUG.

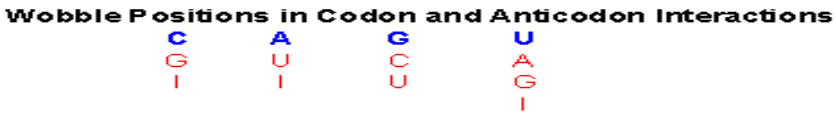
- The codons are the same except for the 3rd nucleotide position.

- 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.

- This is because in a number of cases, the identity of the base at the third position can wobble, and the same amino acid will still be specified.





I: inosine

wobble hypothesis states that some ribonucleotides can form base-pairing even if they are not complementary (in third position)

THE AMINO ACID is specified by more than one genetic codon so the change in the 3rd position will form a genetic codon for the same amino acid

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.

5- The code is nearly universal. Almost all organisms in nature (from bacteria to humans) use exactly the same genetic code. The rare exceptions include some changes in the code in mitochondria, and in a few protozoan species.

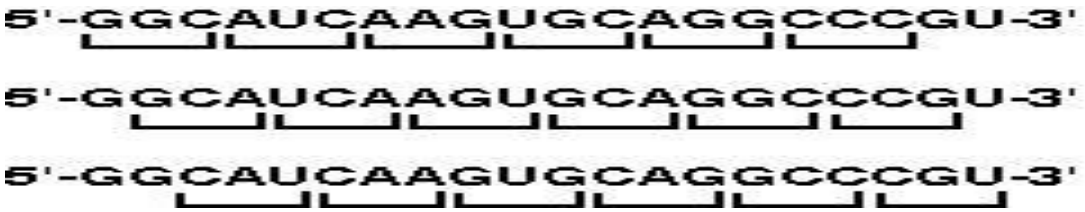
one genetic codon will be specified for one amino acid ,
We can't say that one genetic codon can indicate for more than one amino acid because the protein will has a lot of mistakes

The tRNA is different from other RNA :
-it's the only type is similar to DNA because it has a thymine instead of uracil and it has a base pairing consisting of the shank
- is the only type of RNA that contains unusual ribonucleotide .
- these ribonucleotides don't exist in other types of RNA, they are three types :
1- Pseudouridine
2-Dihydrouridine
3-Inosine

Reading frames

The frame that will be reading by ribosome to translate segment of RNA into protein .

- **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, 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 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) and at the 3' end is called the 3' untranslated region (3'UTR).

- **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.**
- **A position of a codon is said to be a fourfold degenerate site 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 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.

- **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 degenerate site.**

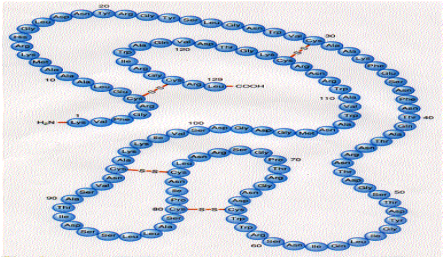
There is 50% possible for the mutation causing by changing 3rd position

- **A position of a codon is said to be a non-degenerate site if any mutation at this position results in amino acid substitution.**
- **There is only one threefold degenerate 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 isoleucine codon: AUU, AUC, or AUA all encode isoleucine, but AUG encodes methionine.**

Protein synthesis

- **It 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).**

Preproprotein: immature protein , Pre is represented by what we called signal sequence ,and pro is represented by what we called inhibitory sequence like proenzyme



- **When a protein must be available on short notice or in large quantities, a protein precursor is produced (proprotein).**
- **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 A preproprotein is a form that contains a signal sequence (an N-terminal signal peptide) that specifies its insertion into or through membranes, i.e., targets them for secretion.**

It's very important for the secretion of protein, an enzyme called signal peptidase removes the signal peptide (sequence) in endoplasmic reticulum and converts it to proprotein and the inhibitory sequence remain with it

- **The signal peptide is cleaved off in the endoplasmic reticulum.**
  - **Before going to golgi apparatus to be stored in vesicle.**

- **Preproproteins have both sequences (inhibitory and signal) still present. -That will be removed for the activation**

- **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 complex of processes is carried out by the ribosome, formed of two main chains of rRNA, and more than 50 different proteins.**

Gene expression

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.

Gene encoding region (ORF)

↓transcription

mRNA

↓translation

Precursor protein

↓post - translational

↓modifications

Mature protein

↓folding

Biologically active protein

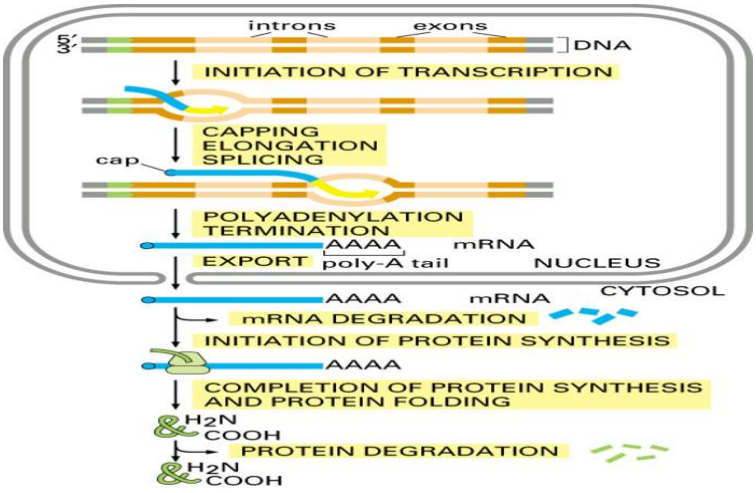


Figure 6–90. Molecular Biology of the Cell, 4th Edition.

➔ **Precursor protein : is inactive protein then it should be activated by the post translation modifications to give the mature protein then folding to give the biologically active protein**

Charging the tRNA

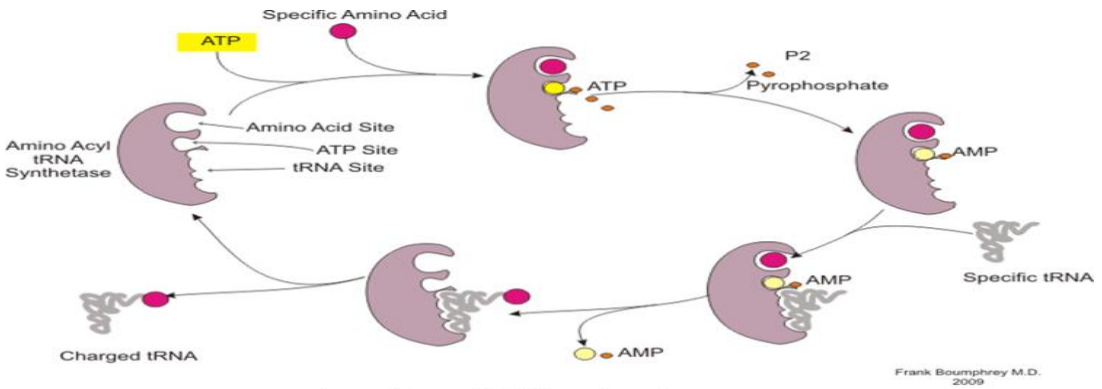
-tRNA acts as a translator between mRNA and protein

- Each tRNA has a specific anticodon and an amino acid acceptor site.

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 AT**

tRNA is regulating the process in protein synthesis because it checks the amino acid which will bind to the acceptor arm(site) not any amino acid can go and bind at the acceptor arm of any tRNA



Loading a tRNA molecule

Charging the tRNA : the first step in protein synthesis and it is the only step to be placed in the cytosol. it is done by binding of tRNA and the proper amino acid on amino acyl tRNA synthetase, which ensures that this suitable amino acid is bounded to the acceptor site of tRNA after the hydrolysis of ATP molecule. 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.

Aminoacyl tRNA synthetase checks for the complementarity between the anticodon of tRNA and the amino acid on the acceptor site.

Aminoacyl tRNA synthetases:

- **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**
- **Two classes of synthetases, differ in the 3-dimensional structures, which side of the tRNA they recognize and how they bind ATP**
- **Class I - monomeric, acylates the 2'OH on the terminal ribose**
  Arg, Cys , Gln, Glu, Ile, Leu, Met, Trp Tyr, Val
- **Class II - dimeric, acylates the 3'OH on the terminal ribose**
  Ala, Asn, Asp, Gly, His, Lys, Phe, Ser, Pro, Thr

Wobble

- **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**

- **Example: U in tRNA can pair with either A or G in the third position of an mRNA codon**
- **This flexibility is called wobble**

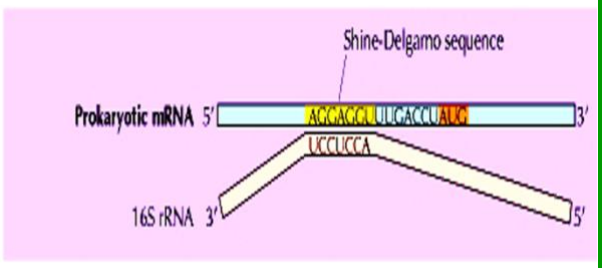
- **There are two levels of control to ensure that the proper amino acid is incorporated into protein through:**

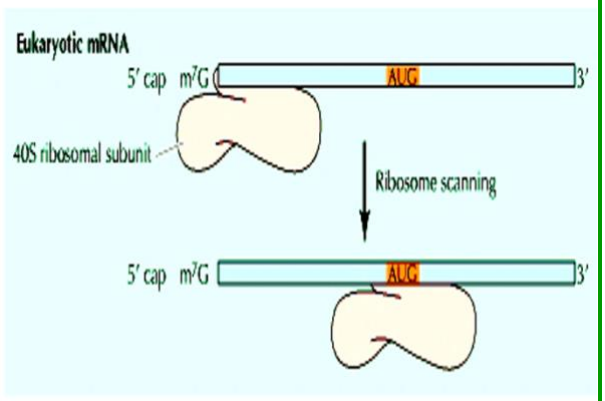
1- The reaction of amino acyl tRNA synthetase for charging the proper tRNA

2- Matching the specific tRNA to a particular codon of mRNA

Wobble: it is definition in the 3rd position which is decreasing the possibility of mutation if there is changing. The amino acid represented by four genetic codons what ever the chaging in the 3rd position it will bring up the genetic codon for the same amino acid

- **In prokaryotes, specific sequences in the mRNA around the AUG codon, called Shine – Delgarno sequences, are recognized by an initiation complex consisting of a Met amino-acyl tRNA, initiation factors (IFs) and the small ribosomal subunit.**
- **In eukaryotes, there is a process called ribosome scanning, where mRNA is moving along the small subunit of ribosome till finding the codon of initiation (AUG) of methionine to be located in the P site.**





In prokaryotes there is no capping or poly A or splicing ,it will translation

Shine –Delgarno sequences:

Recognized by complementary sequence at 3' end of the 16s rRNA

Initiation

- **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 **2- The mRNA**

3- Three Initiation Factors - IF1, IF2 (GTP) and IF3

4- The initiator fMet-tRNAfMet

The following steps take place:

A- Binding of the ribosome 30S subunit with initiation factor (IF3) promotes the dissociation of the ribosome into its two component subunits.

B- The presence of IF3 permits the assembly of the initiation complex and prevents binding of the 50S subunit prematurely, IF1 assists IF3 in some way, perhaps by increasing the dissociation rate of the 30S and 50S subunits of the ribosome.

IF2 : in prokaryotes

fMet-tRNAfMet : in the prokaryotic cells (formyl group is required to recognize the 5' end direction of mRNA) while in eukaryotes there is no need for formyl group due to the presence of the cap, so Met-tRNAMet is used.

C- Binding of the mRNA and the fMet-tRNAfMet

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 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.

IF2(GTP) assists the fMet-tRNAfMet to bind to the 30S subunit in the correct site - the P site.

30S initiation complex: it is small ribosomal subunit binding with mRNA on initiation codon, which binds with anti-codon of tRNA (bounded with IF2 with GTP)

GTP decomposed into GDP +inorganic phosphate

For the produced energy of the GTP binding of the small sub-unit with the large sub-unit because the end stage of protein snthesys of prokarute cell happen in the whole ribosome

-At this stage of assembly, the 30S initiation complex is complete and IF3 can dissociate.

The energy that produced from of GTP another 2 ribosomeal protein it is the cause of helping the IF2 between 50s+30s to give 70s

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

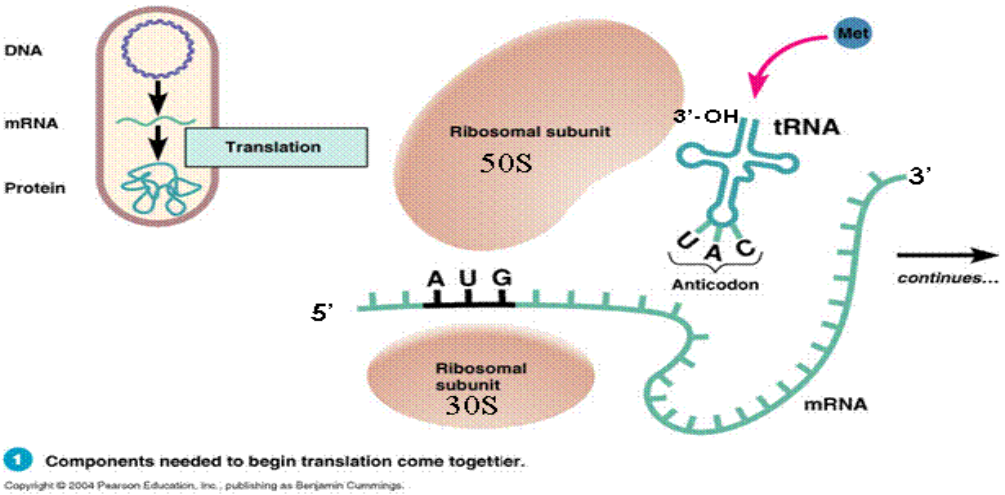
Three events now happen "simultaneously".

As the 50S subunit of the ribosome associates with the 30S initiation complex, GTP hydrolysis occurs on IF2.

This hydrolysis may be helped by the L7/L12 ribosomal proteins rather than by IF2 itself.

It is the only stage in protein senthysis in prokaryote cell that happen in the small sub-unit

Fig. 8.9



1 Components needed to begin translation come togettier.

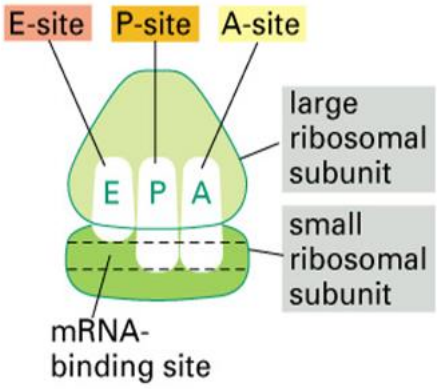
-

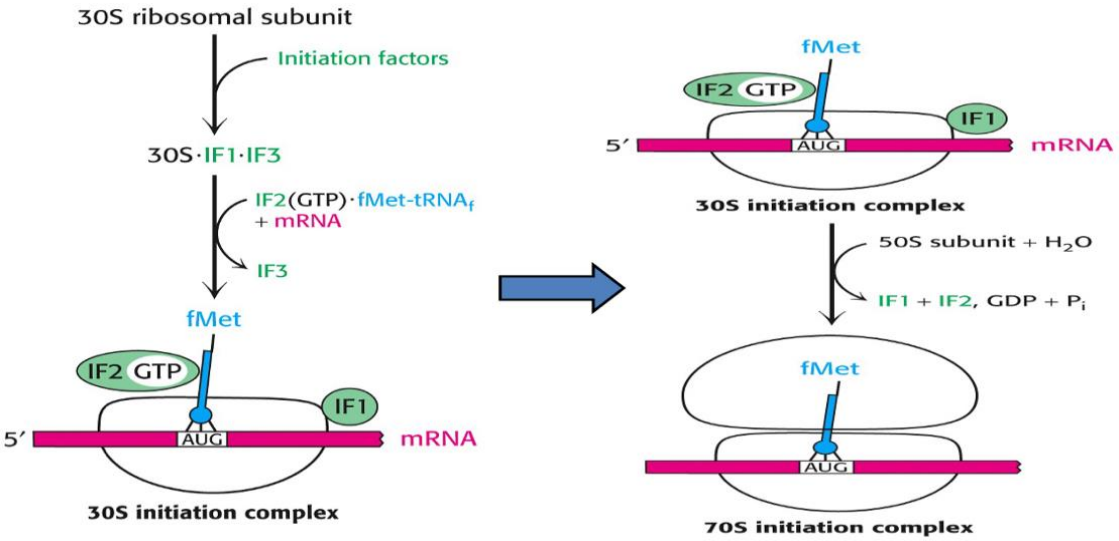
- **In addition to the APE sites there is an mRNA binding groove that holds onto the message being translated**
- **The A site binds an aminoacyl-tRNA**

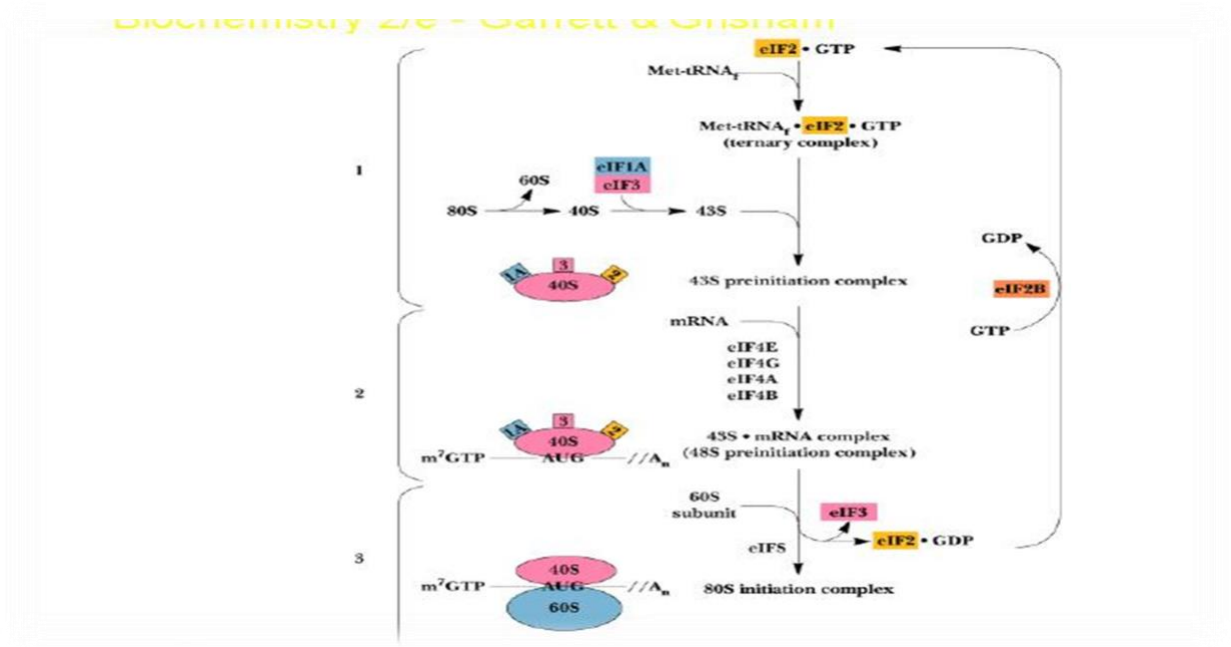
(a tRNA bound to an amino acid);

P site binds a peptidyl-tRNA (a tRNA bound to the peptide being synthesized).

- **The E site binds a free tRNA before it exits the ribosome**







.

There is type of eIF :

eIF1A & eIF3: bind to the small sub-unit to cause dissotiation between long and small sub-units

eIF2 : bind to GTP with methionine on tRNA

eIF4A ,eIF4E , eIF4B, eIF4G: These four bind with mRNA

These 7 eIF for the formation of initiation complex 80s

When the reassociation between small sub-unit ribosome and large sub-unit 80s ribosomal sub unital all the type of initiation factor will be relased and the

GTP that binding to tRNA break down to give energy for reassociation between the small subunit and large sub unit in eukaryotic cells

Initiation stage of the protein senthayses is also take place in small sub unit

The two intiation factors eIF3 and eIF1A they bind 40s (small substance) where the initionation stage will happen on

There is no formyl group because the 5' end has been recognition by the ribosomal camp

TERNARY COMPLEX : Complex is methanion tRNA with e1f2 and GTP

1.THERE IS SMALL SUB-unit BINDING TO e1F1A and eIF3 and tRNA with eIF2 and GTP and methionin that on the acceptor arm of tRNA

2. 43s preinitiation complex will bind with 4 eIF with mRNA and this will lead to increase the sidmentetion and give us preinatitation complex and bind to 60s to give 80s inatition complex

- Eukaryotes use a scanning mechanism to initiate translation.
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

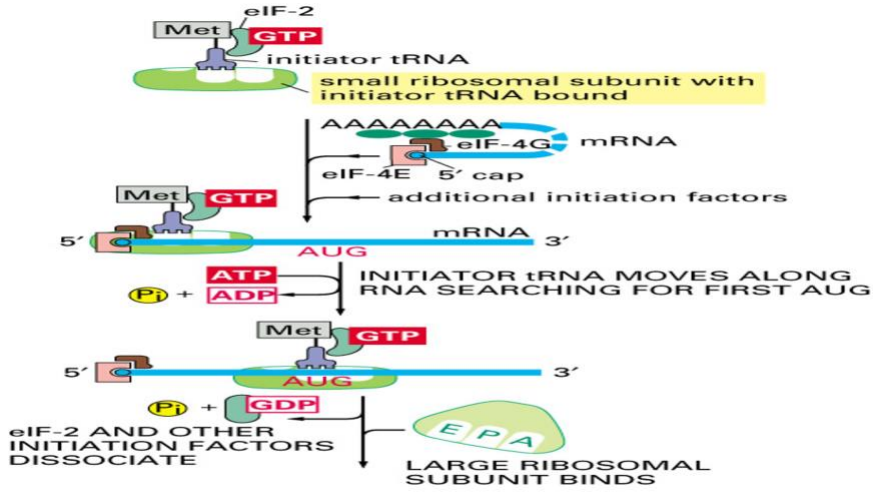


Figure 6–71 part 1 of 2. Molecular Biology of the Cell, 4th Edition.

Elongation

Three special Elongation Factors are required for this phase of protein synthesis: EF- Tu (GTP), EF-Ts and EF-G (GTP).

-The Elongation phase of protein synthesis consists of a cyclic process whereby a new aminoacyl-tRNA 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: 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