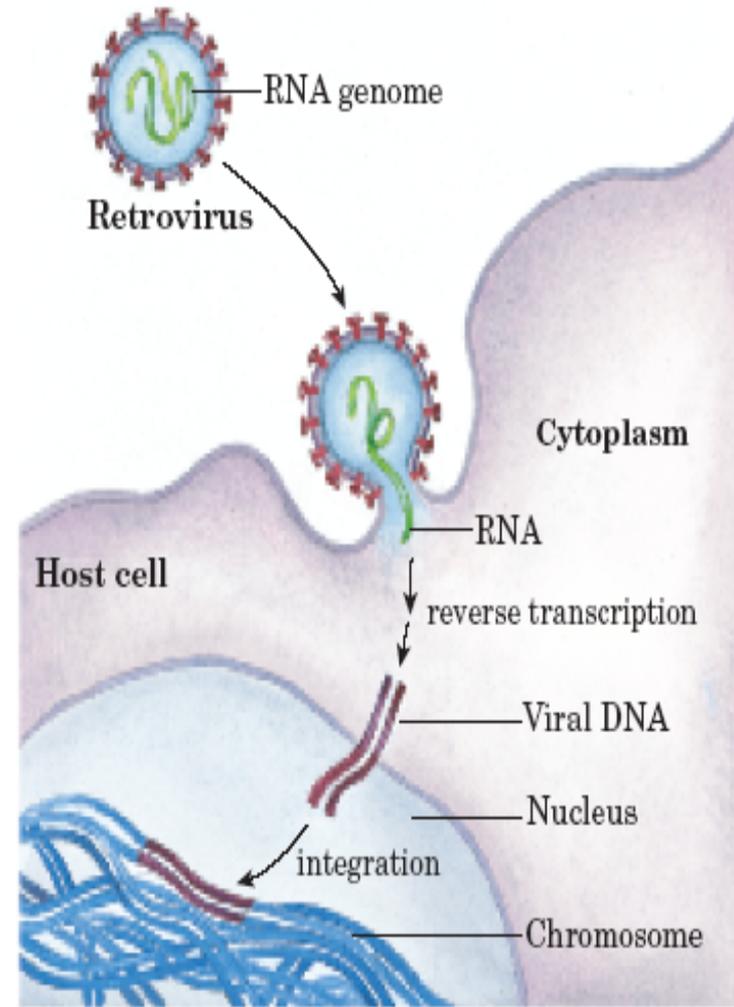


Reverse transcription

Reverse transcription

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

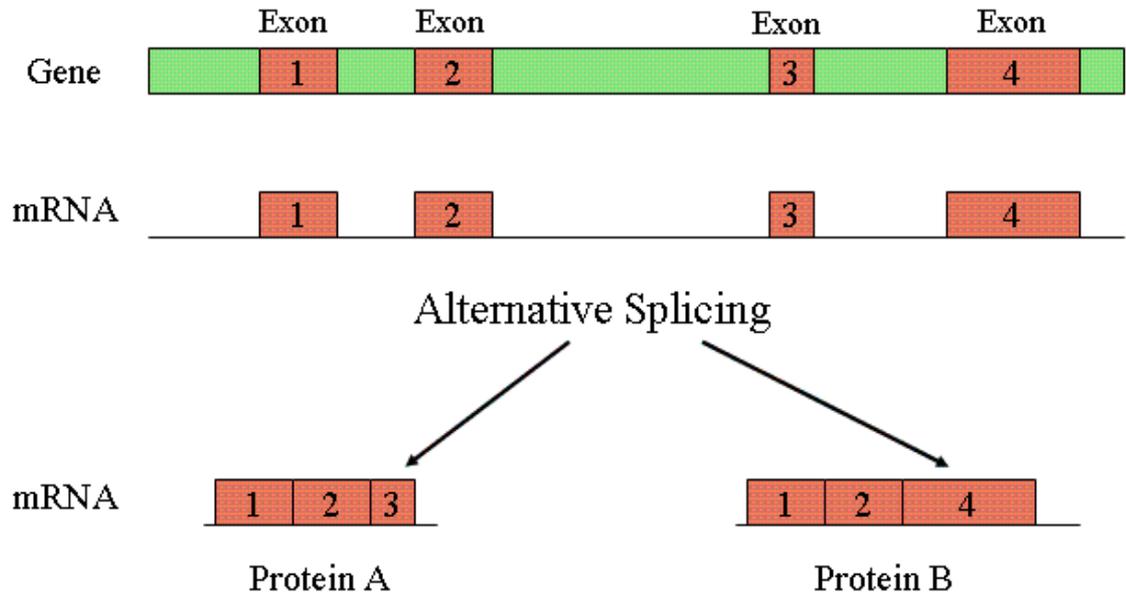
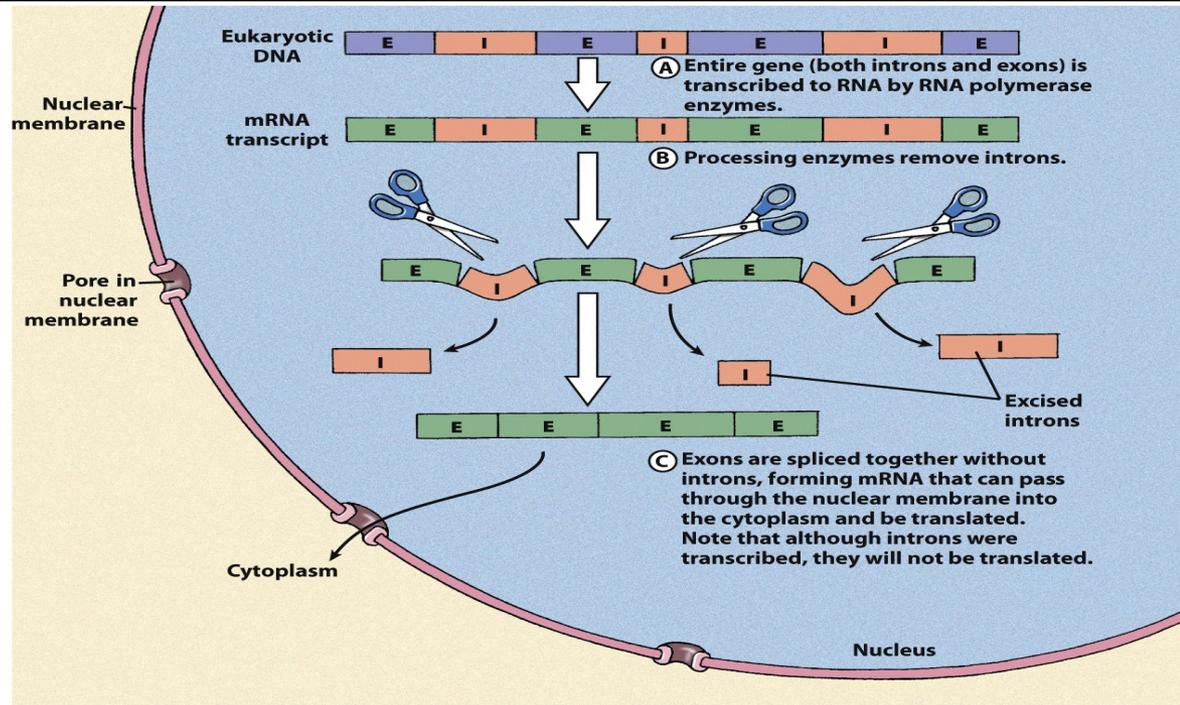


- In the case of HIV, **reverse transcriptase** is responsible for synthesizing a complementary DNA strand (cDNA) to the viral RNA genome.
- An associated enzyme, **ribonuclease H**, digests the RNA strand, and **reverse transcriptase** synthesizes DNA complementary strand to form a double helix DNA 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)

- Exons are “coding” regions
- Introns are removed
- 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 codes

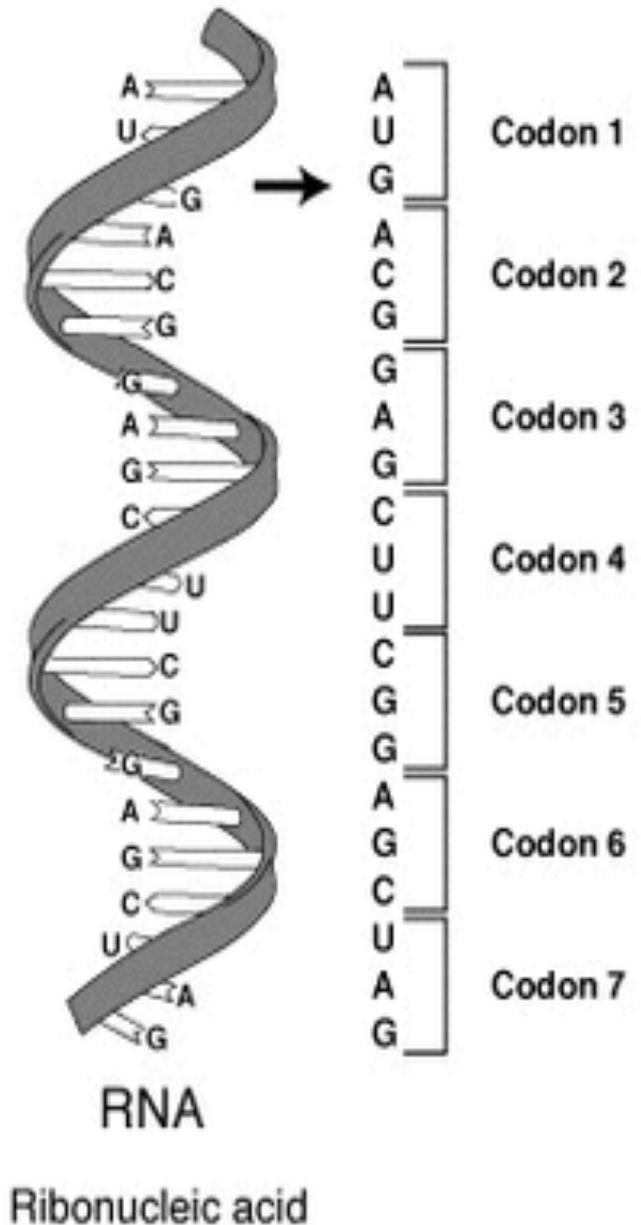
Seconded Position

		Seconded Position									
		U		C						A	
		code	Amino Acid	code	Amino Acid	code	Amino Acid	code	Amino Acid		
		First Position	U	UUU	phe	UCU	ser	UAU	tyr		
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	

Third Position

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.



Characters of genetic code:

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.

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...



- 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.
 - 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.
 - 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.

Wobble Positions in Anticodon and Codon Interactions



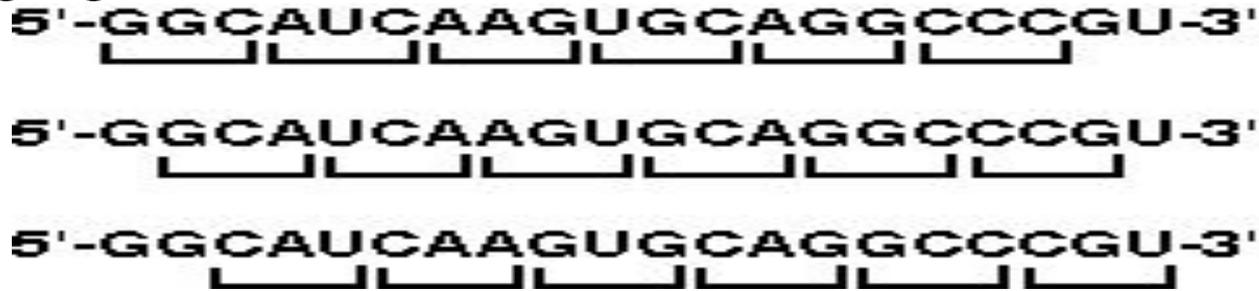
Wobble Positions in Codon and Anticodon Interactions



- 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.

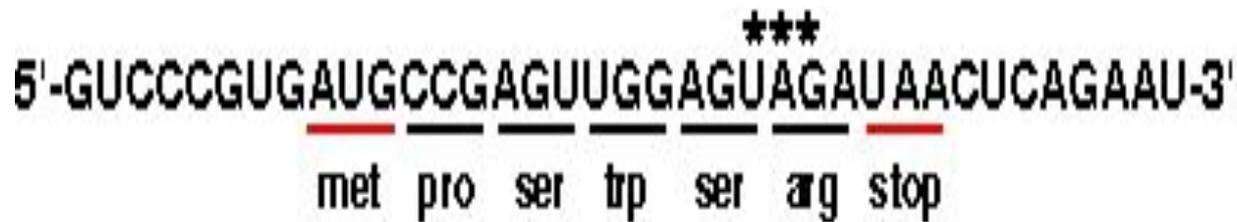
Reading frames

- 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.

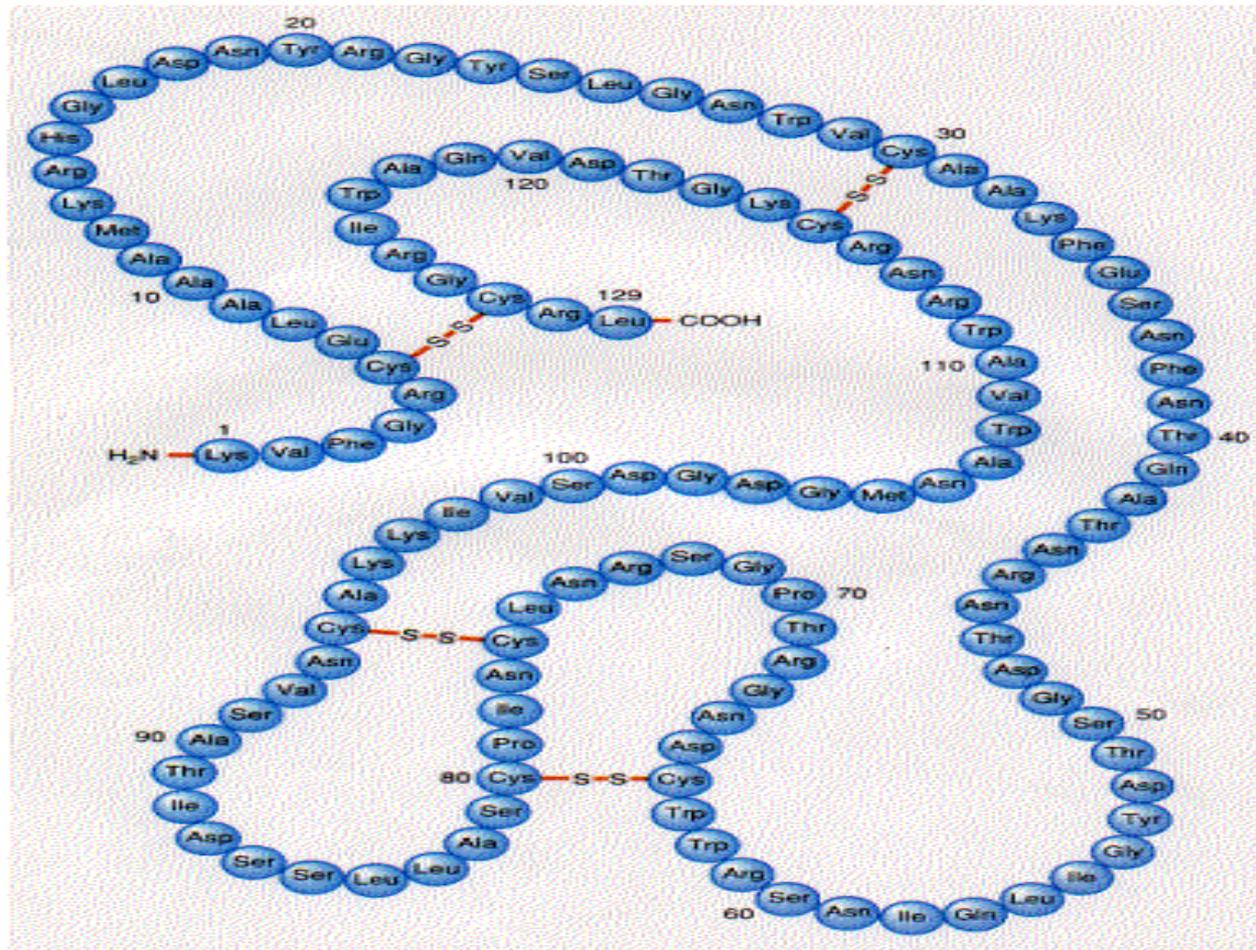


- In this sequence, there are nucleotides at either end that are outside of the open reading frame. Because they are outside of the open reading frame, these nucleotides are not used to code for amino acids.

- 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 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 (GAAA, GAGG) is a twofold degenerate site.
- 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).
- 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 modification (**trimming**).
- 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.

- The signal peptide is cleaved off in the endoplasmic reticulum.
- Preproteins have both sequences (**inhibitory and signal**) still present.
- 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)

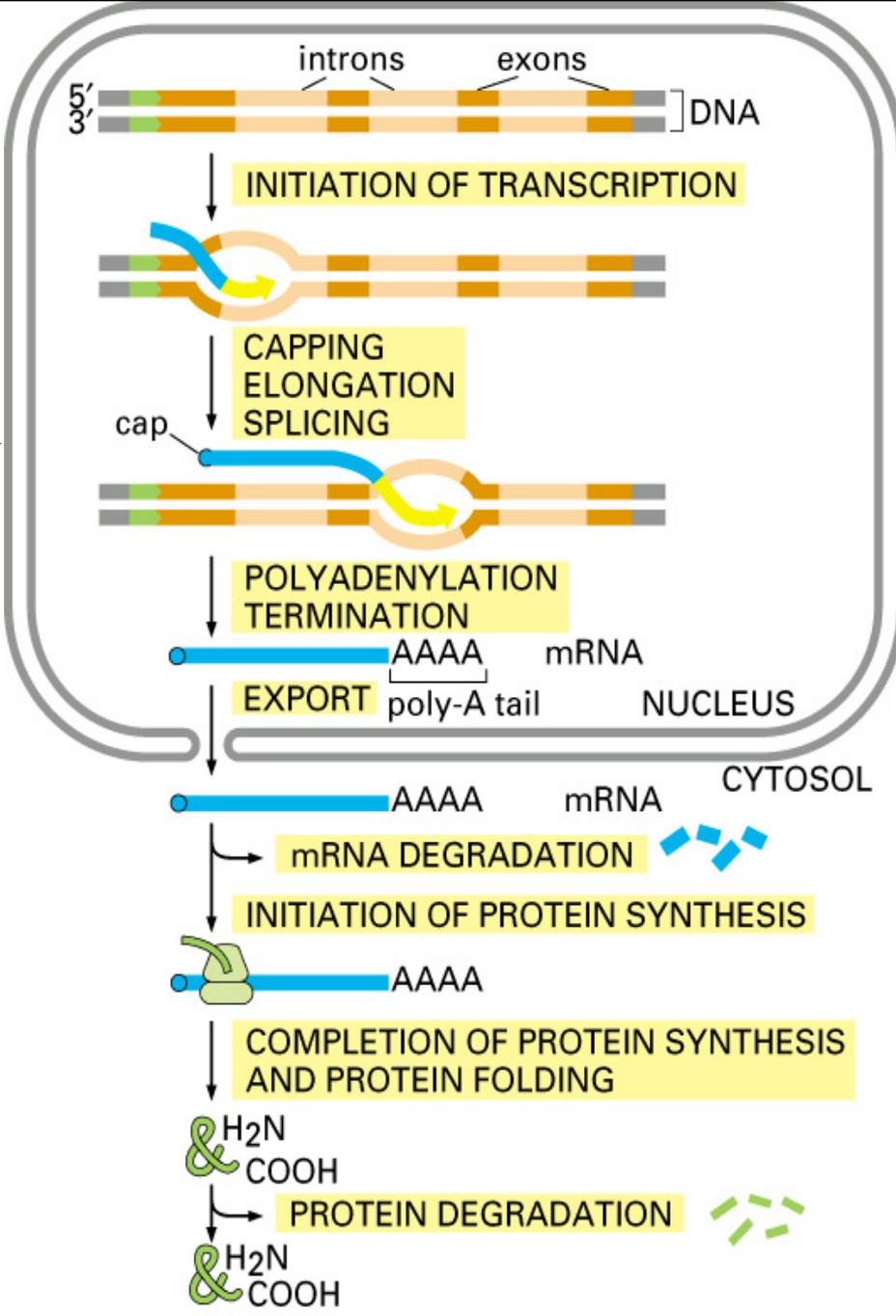
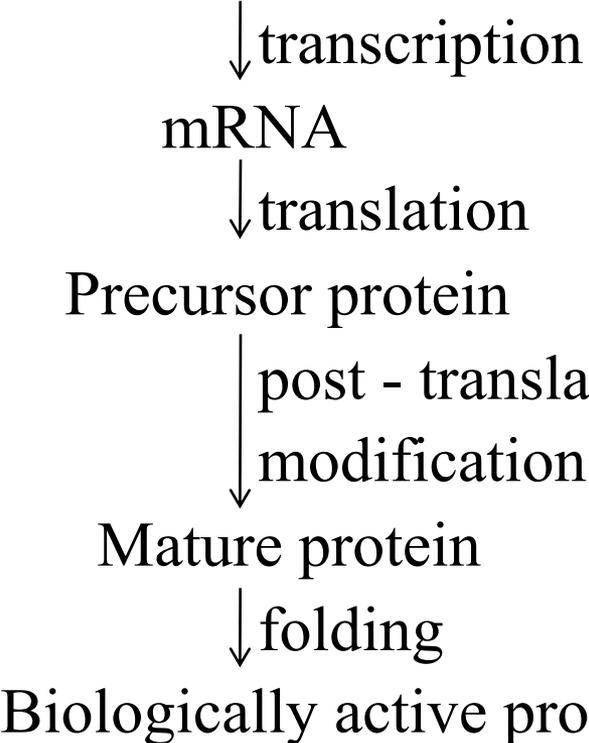
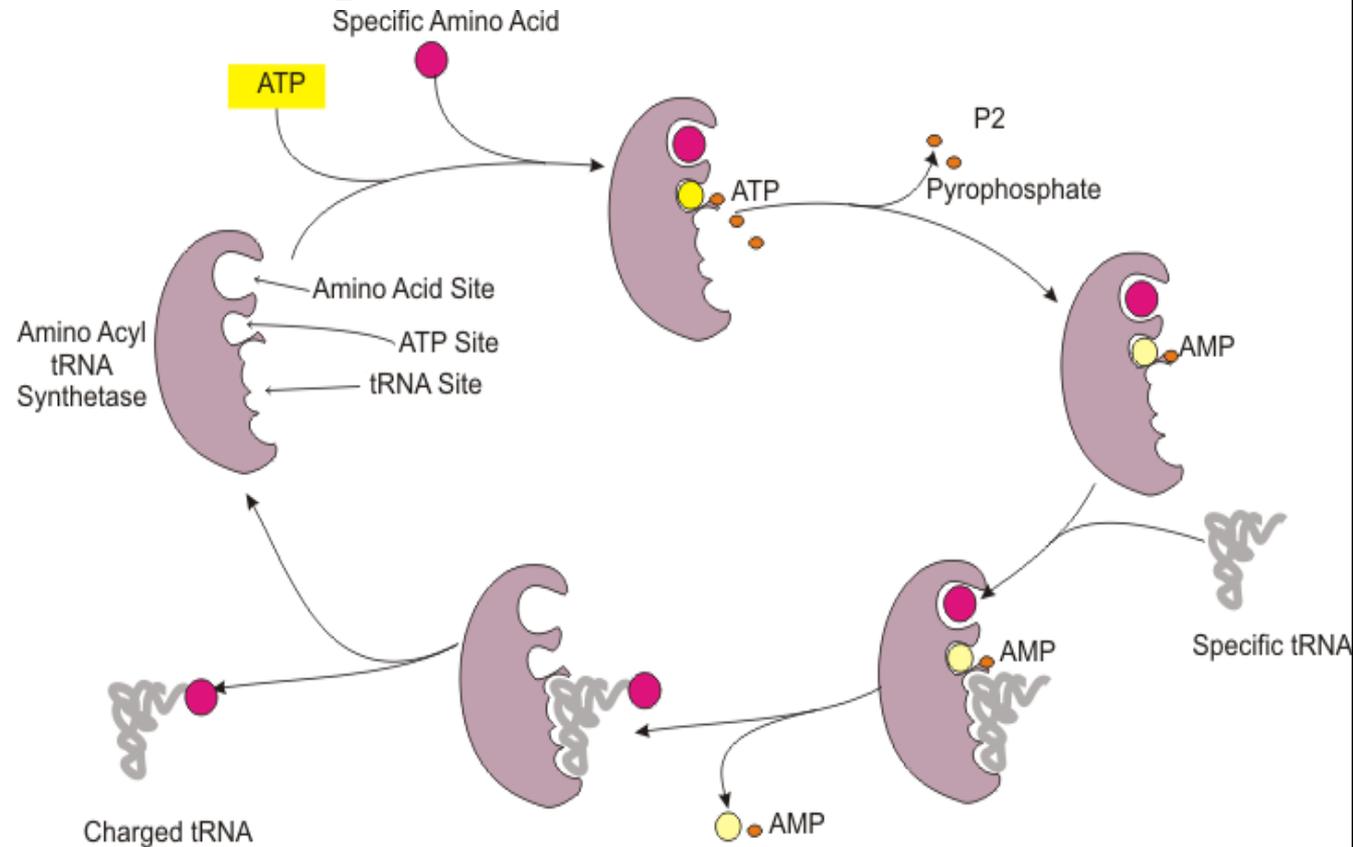


Figure 6-90. Molecular Biology of the Cell, 4th Edition.

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 ATP.



Loading a tRNA molecule

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

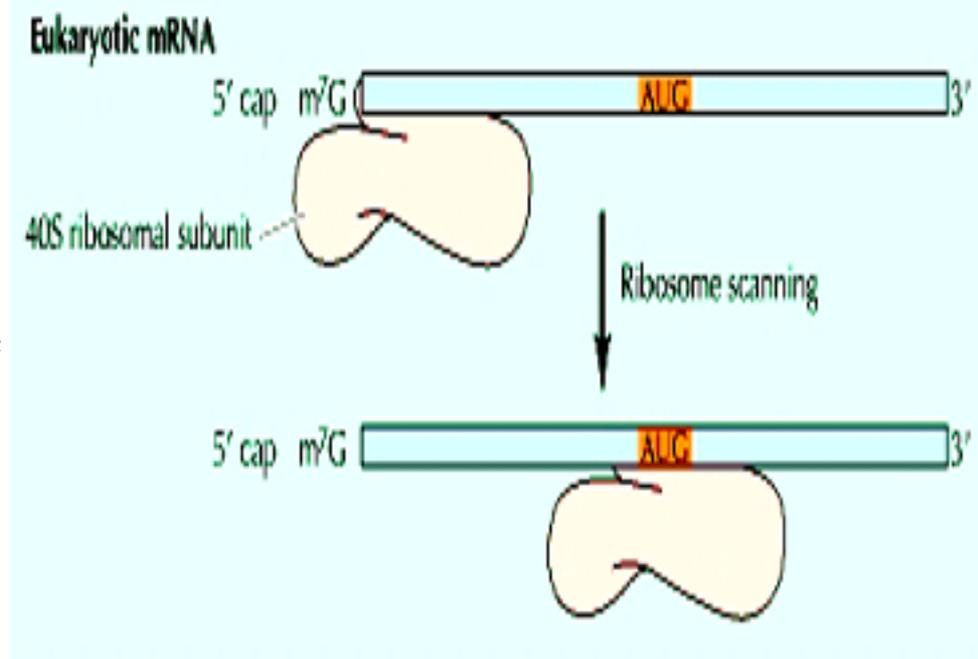
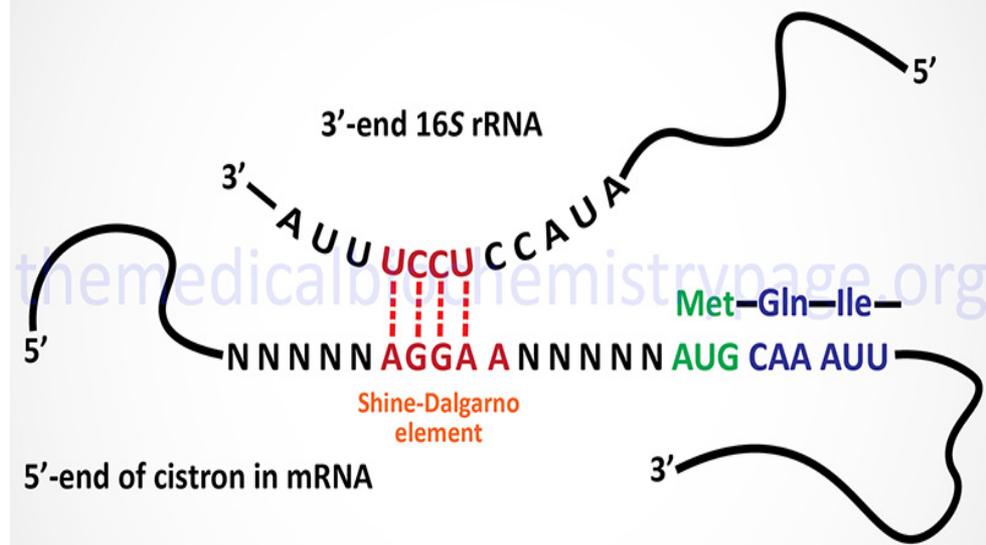
Wobble

- 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 as 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

- 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.



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-tRNA_f^{Met}

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.

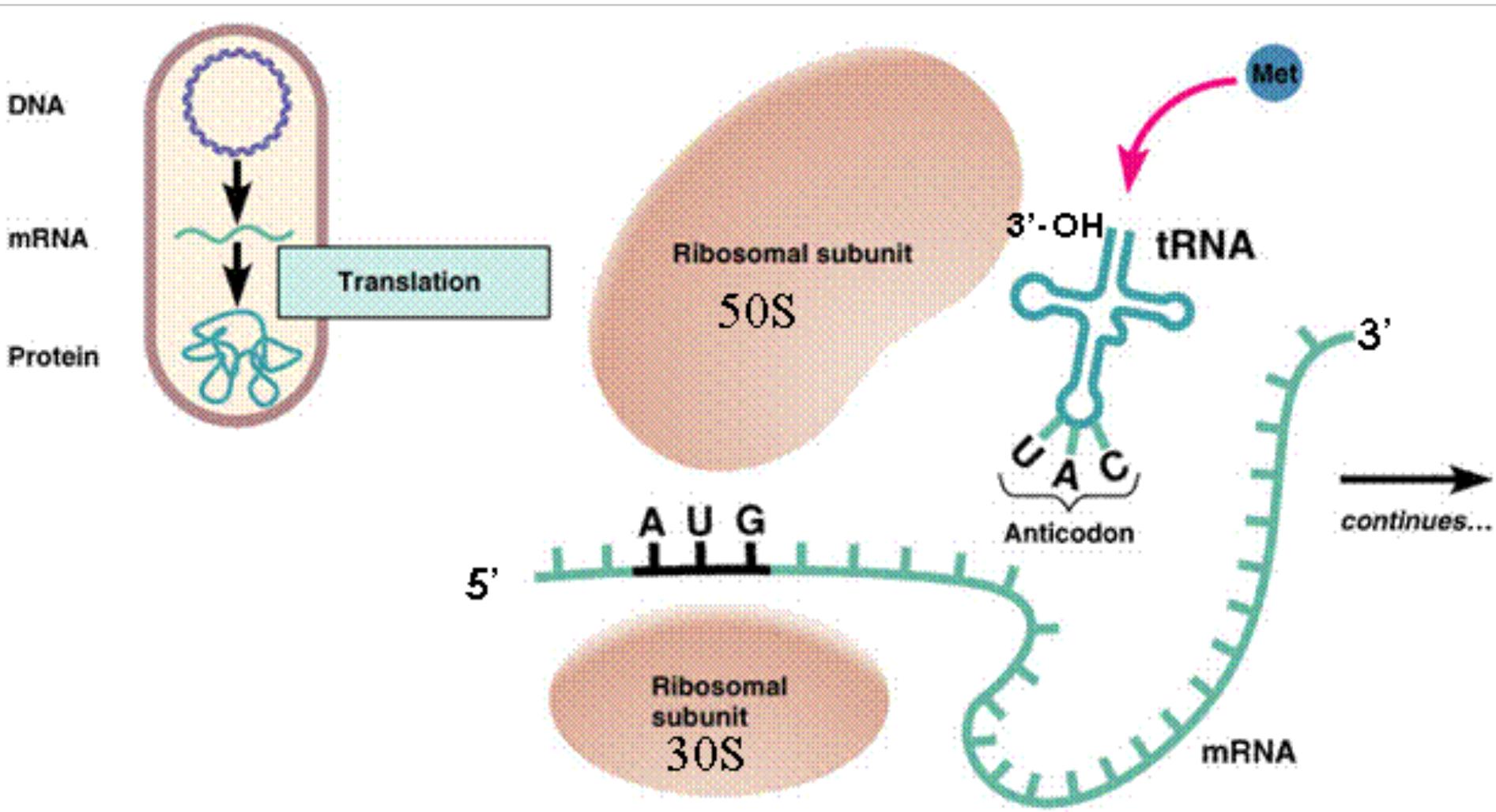
C- Binding of the mRNA and the fMet-tRNA_f^{Met}

- 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 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-tRNA_f^{Met} to bind to the 30S subunit in the correct site - the P site.
- 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

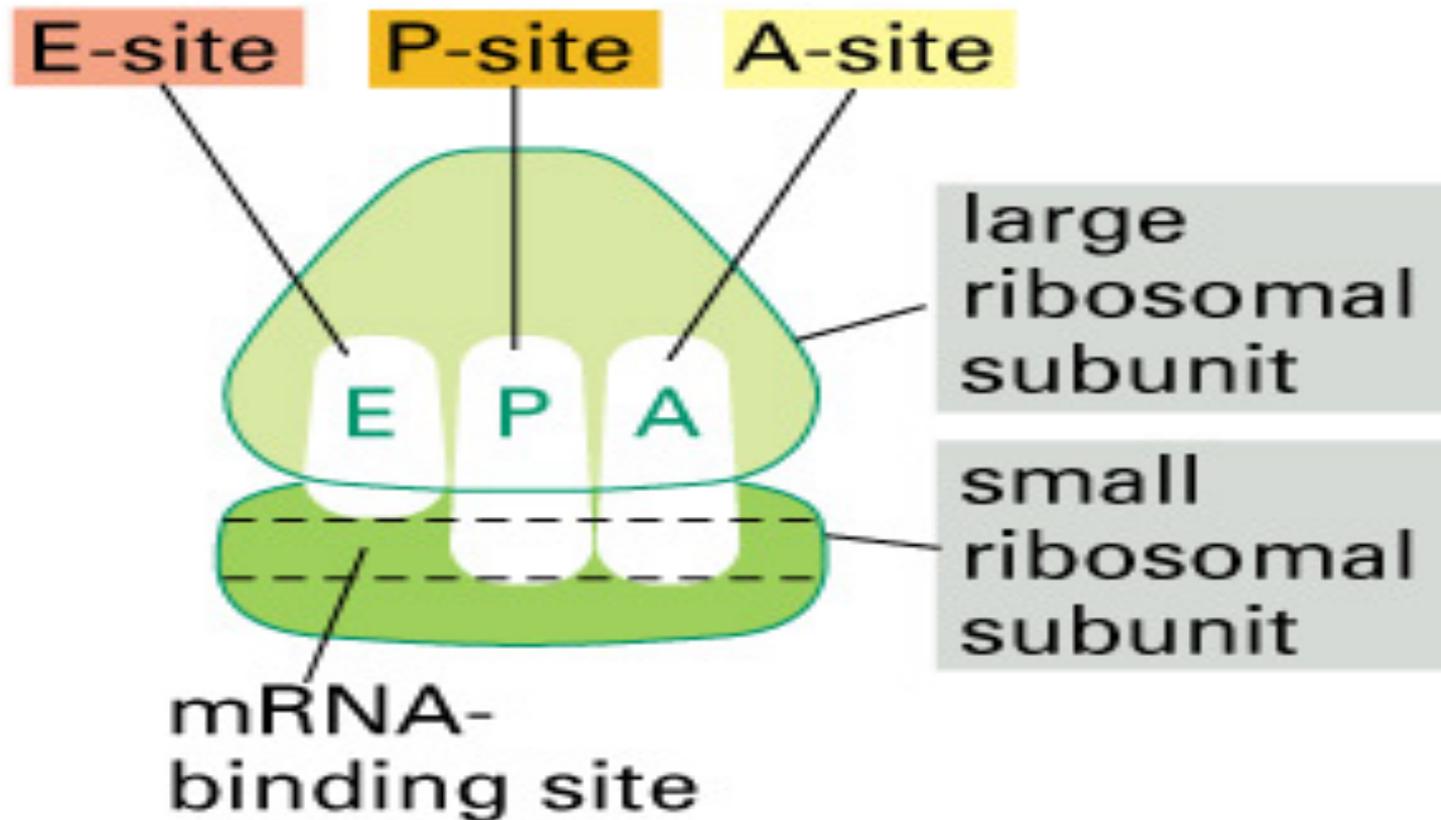
- Three events now happen "simultaneously".
- As the 50S subunit of the ribosome associates with the 30S initiation complex, GTP hydrolysis occurs on IF2.
- This hydrolysis may be helped by the L7/L12 ribosomal proteins rather than by IF2 itself.

Fig. 8.9



1 Components needed to begin translation come together.

- 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.



30S ribosomal subunit

Initiation factors

30S·IF1·IF3

IF2(GTP)·fMet-tRNA_f
+ mRNA

IF3

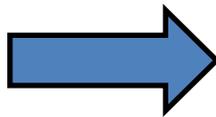
fMet

IF2 GTP

IF1

5' mRNA AUG

30S initiation complex



fMet

IF2 GTP

IF1

5' mRNA AUG

30S initiation complex

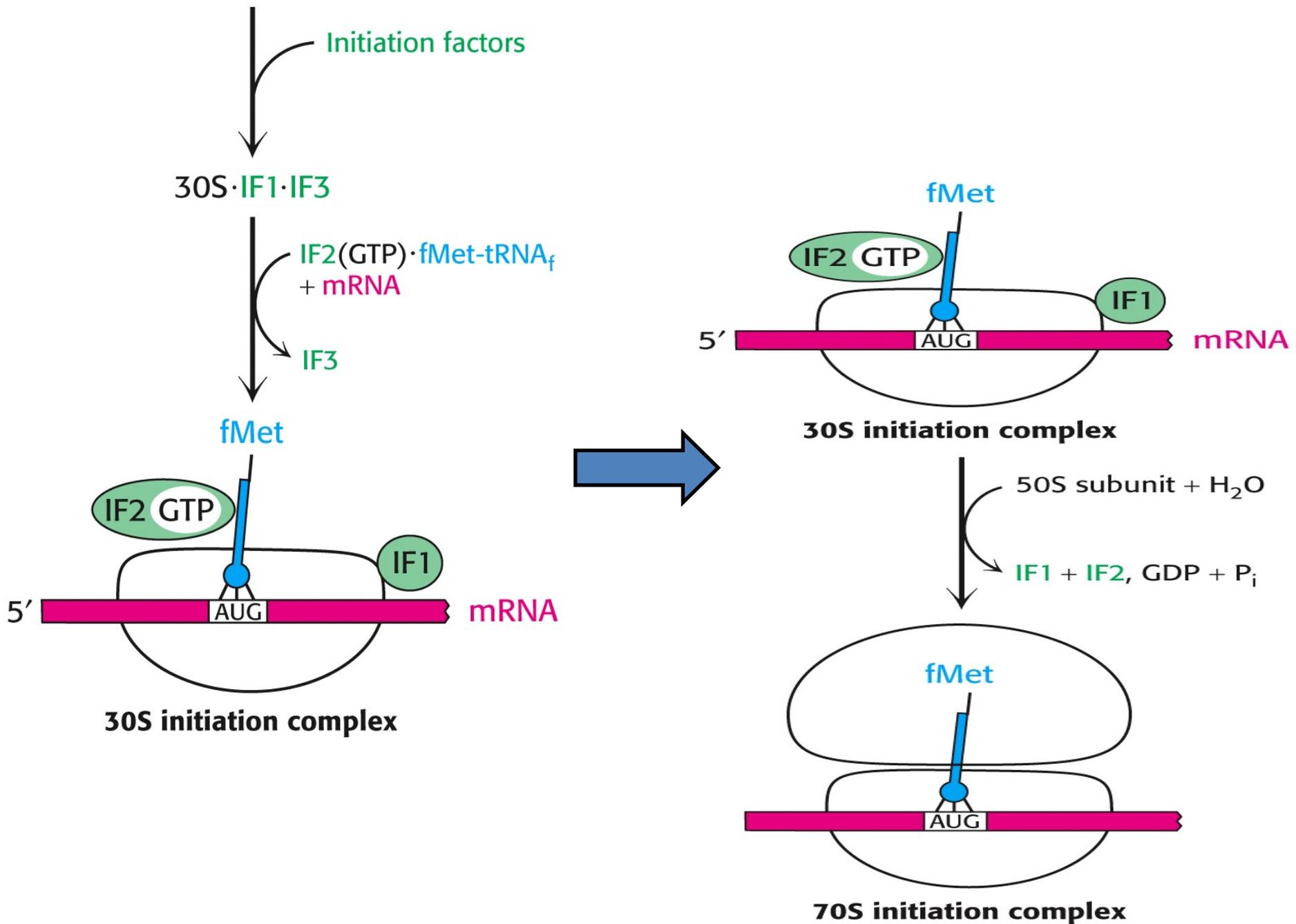
50S subunit + H₂O

IF1 + IF2, GDP + P_i

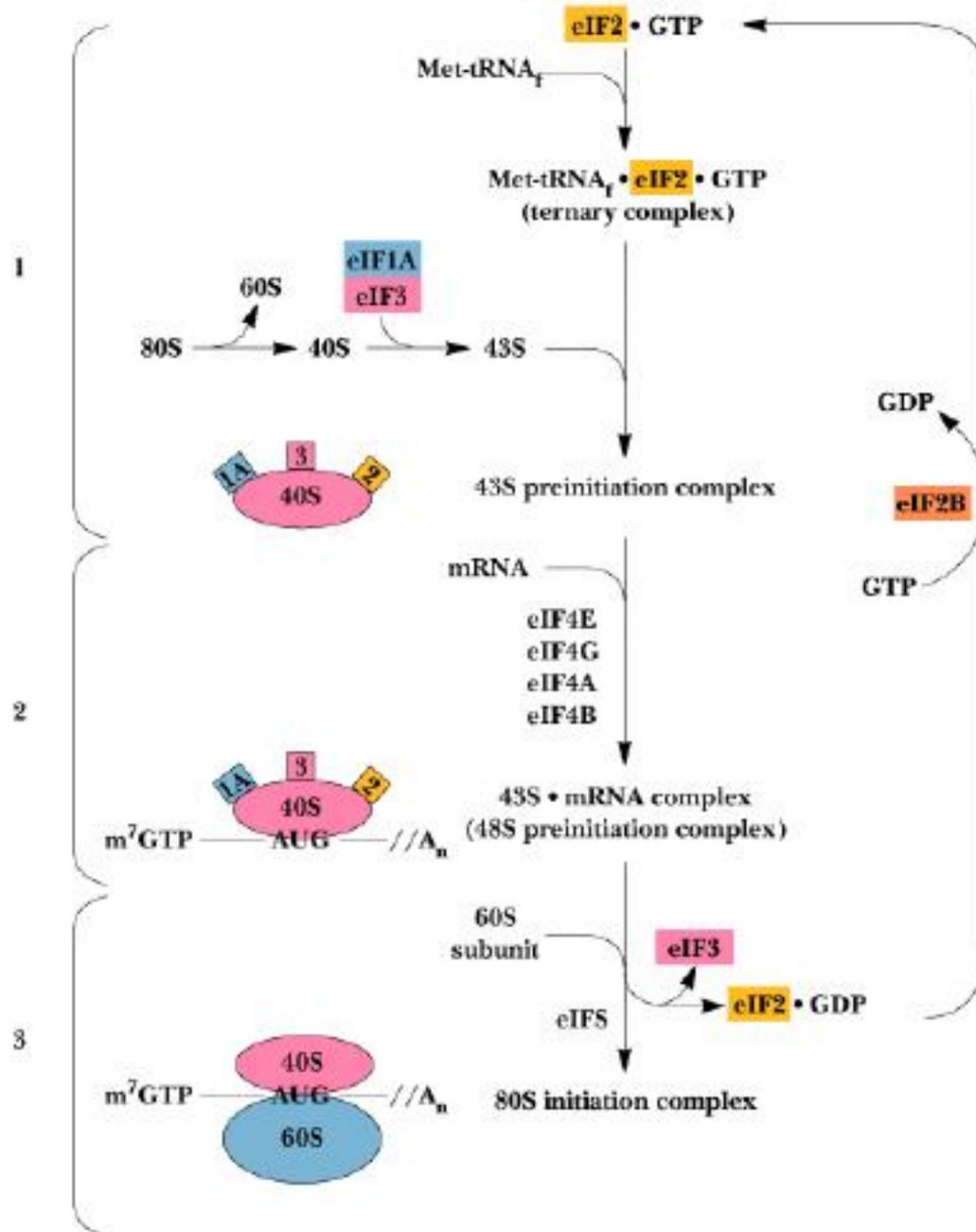
fMet

5' mRNA AUG

70S initiation complex



Biochemistry 2/e - Garrett & Grisham



- 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

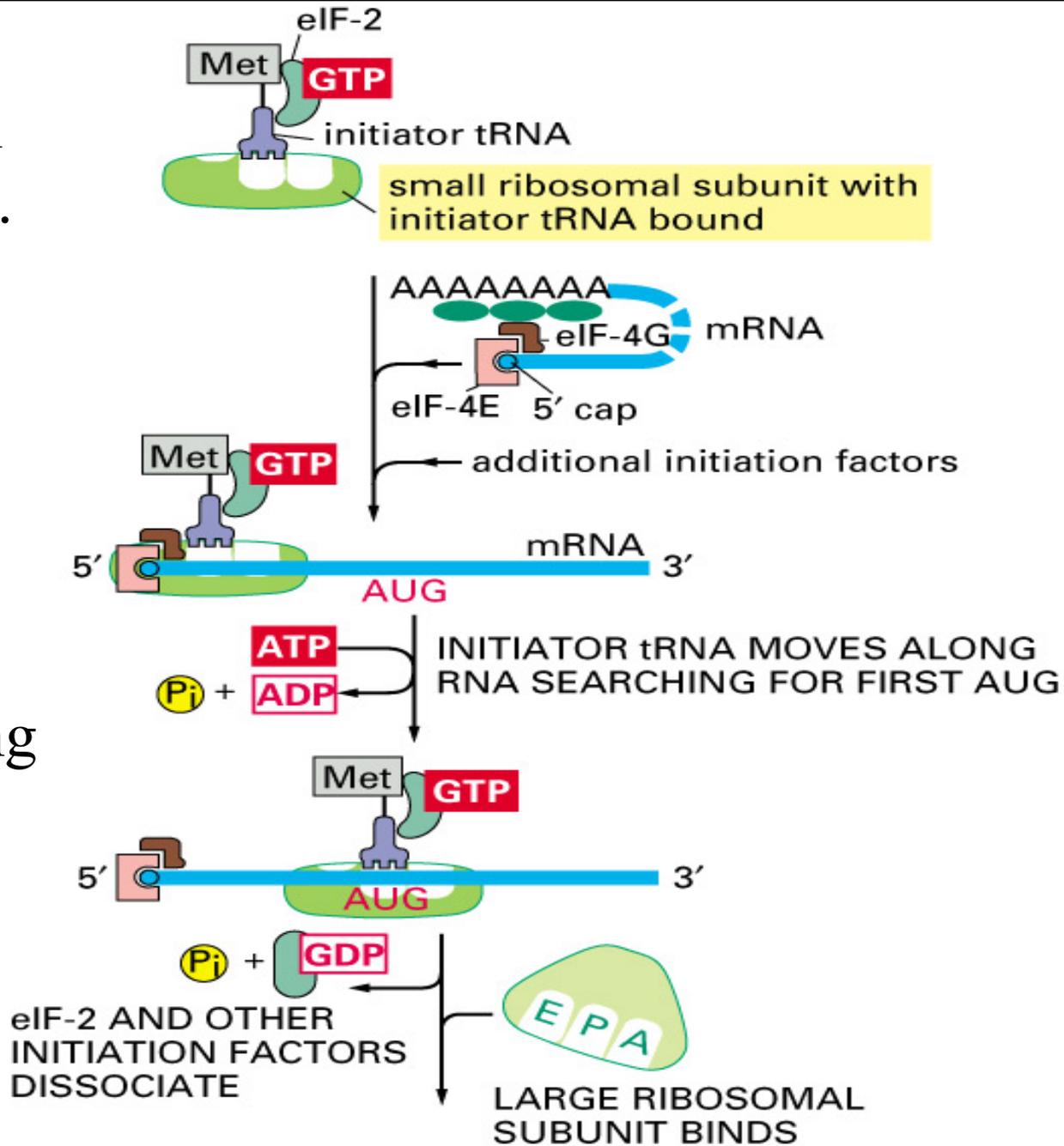


Figure 6-71 part 1 of 2. Molecular Biology of the Cell, 4th Edition.

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

At the start of each cycle:

- The A (aminoacyl) site on the ribosome is empty
- The P (peptidyl) site contains a peptidyl-tRNA, and the E (exit) site contains an uncharged tRNA.
- The elongation factor, EF- Tu (GTP) binds with an aminoacyl-tRNA and brings it to the ribosome.

- 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:
 - EF- Tu prevents the aminoacyl end of the charged tRNA from entering the A site on the ribosome.
 - This ensures that codon- 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.
 - GTP hydrolysis is SLOW and EF- Tu cannot dissociate from the ribosome until it occurs.
- 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.

Termination

- 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.

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).
- These release factors act at the A site of the ribosome.
- A third release factor, RF3 (GTP), stimulates the binding of RF1 and RF2.

Hydrolysis of the peptidyl-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 polypeptide chain.

-The uncharged tRNA in the E site can dissociate as can the release factors.

- GTP is hydrolyzed without dissociating tRNA in the P site.

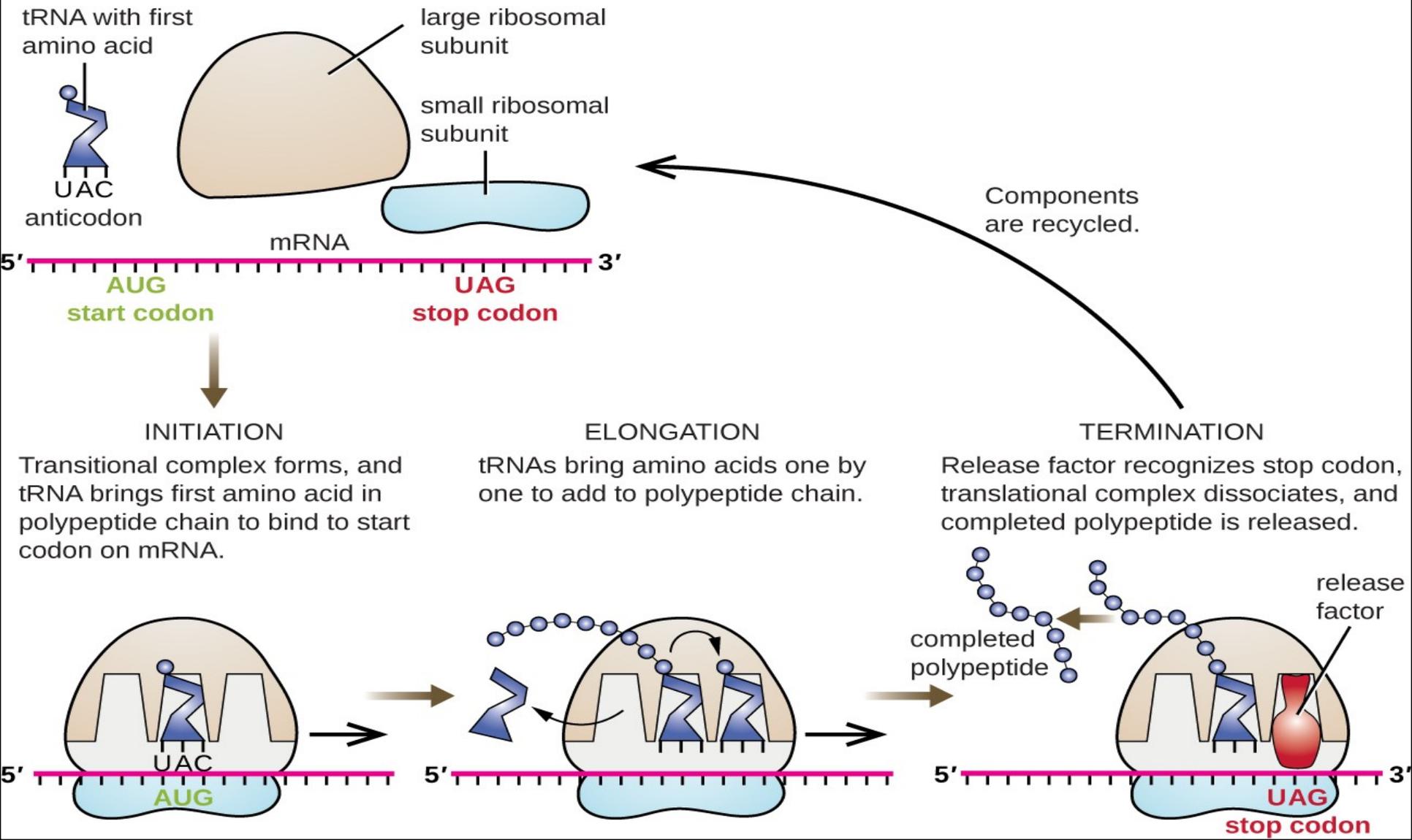


TABLE 29.4 Antibiotic inhibitors of protein synthesis

Antibiotic	Action
Streptomycin and other aminoglycosides	Inhibit initiation and cause misreading of mRNA (prokaryotes)
Tetracycline	Binds to the 30S subunit and inhibits binding of aminoacyl-tRNAs (prokaryotes)
Chloramphenicol	Inhibits the peptidyl transferase activity of the 50S ribosomal subunit (prokaryotes)
Cycloheximide	Inhibits the peptidyl transferase activity of the 60S ribosomal subunit (eukaryotes)
Erythromycin	Binds to the 50S subunit and inhibits translocation (prokaryotes)
Puromycin	Causes premature chain termination by acting as an analog of aminoacyl-tRNA (prokaryotes and eukaryotes)
Diphtheria toxin	Inhibits eEF-2 by ADP-ribosylation of modified histidine in the factor

Antibiotics inhibiting translation

-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 and Tetracyclins (the latter inhibiting binding of aminoacyl-tRNA to the ribosomal A- site at the 30S ribosomal subunit).
2. Introduction of aminoacyl-tRNA and synthesis of a peptide bond, inhibited by puromycin (leading to premature termination) and chloramphenicol (probably inhibiting the peptidyltransferase).
3. Translocation of the mRNA relative to the ribosome blocked by erythromycin and fusidic acid (the latter preventing release of EF-G/GDP).

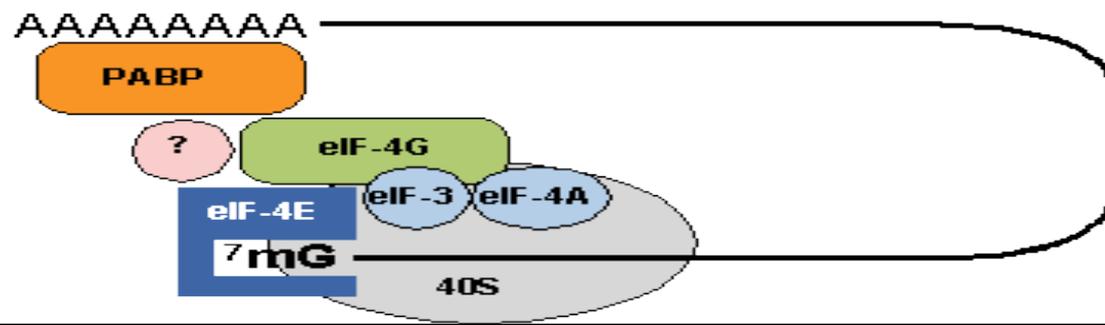
Protein synthesis in eukaryotes

- 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.
- 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.
 - 2- While the initiating amino acid in eukaryotic protein synthesis is still methionine, it is not formylated.
 - 3- Eukaryotic mRNA is capped. This is used as the recognition feature for ribosome binding -- not the 18S rRNA.
 - 4- The initiation phase of protein synthesis requires over 10 eukaryotic initiation factors (eIFs) one of which is the cap binding protein.

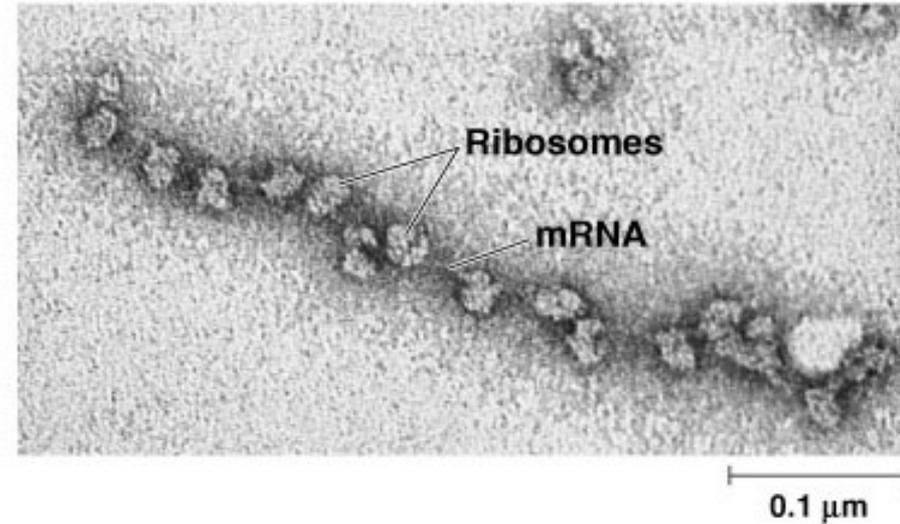
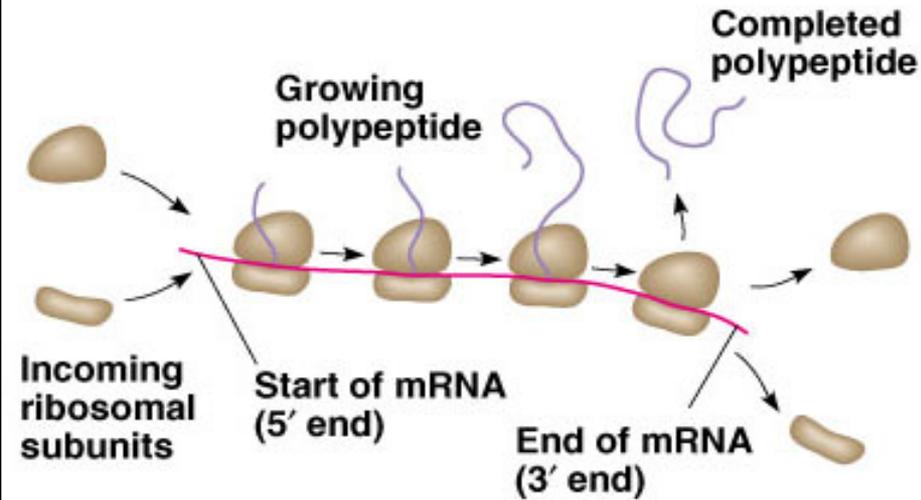
- 5- The eukaryotic elongation phase closely resembles that in prokaryotes.
- The corresponding elongation factors are eEF-1a (EF-Tu), eEF-1b γ (EF-Ts) and eEF-2 (EF-G).
- 6- Eukaryotes require just a single release factor, eRF.

Coordinating protein synthesis with mRNA synthesis

- It has recently been found that the eukaryotic initiation factor eEF-4G 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.
- It is thought that the binding of eEF-4G to PABP serves as a critical recruitment step for driving downstream translation.
- 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.



(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).

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, conformation or size of protein molecule

Effects:

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

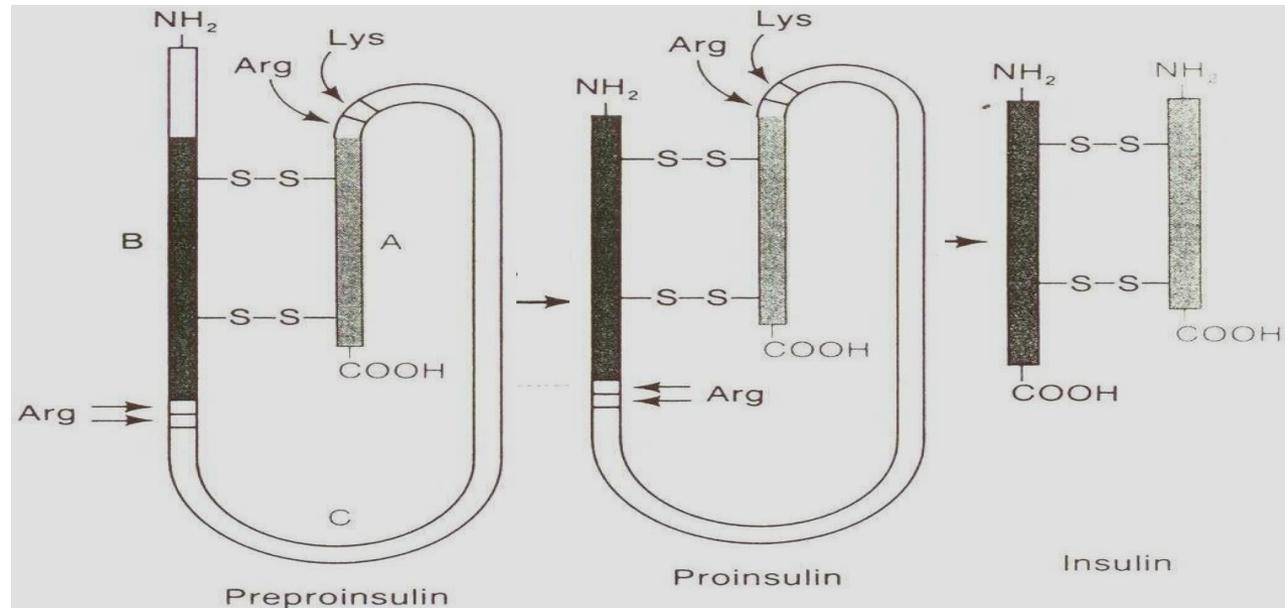


Figure 20-22

Conversion *in vivo* of preproinsulin to proinsulin and then to insulin.

- 1- Insulin, which is a low-molecular-weight protein having two polypeptide chains which fold to allow interchain and intrachain disulfide bridges.
 - A specific protease then clips out the segment that connects the two chains which form the functional insulin molecule, where a propeptide is removed from the middle of the chain; the resulting protein consists of two polypeptide chains connected by disulfide bonds.
- 2- Also, most nascent polypeptides, the initial methionine is usually taken off during post-translational modification by specific aminopeptidases.
- 3- Other modifications, like phosphorylation, are part of common mechanisms for controlling the behavior of a protein, for instance activating or inactivating an enzyme.

- 4- Some animal viruses, as poliovirus and hepatitis A virus, synthesize long 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.
- 5- Collagen, an abundant protein in the extracellular spaces of higher eukaryotes, is synthesized as procollagen three polypeptide molecules, that align themselves in a particular way that is dependent upon the existence of specific amino terminal peptides.
- 6- Specific enzymes then carry out hydroxylations and oxidations 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:

- Many proteins secretion from the cell are made as:
- large precursor molecules but functionally inactive
- Change of protein from non active for active molecule by removing portions of the protein chain by specialized endoproteases
- Sites of the cleavage reaction:
 - Endoplasmic reticulum
 - Golgi apparatus
 - Secretory vesicles
- N.B. zymogens which are inactive enzymes, become activated through cleavage 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 and reversed by protein phosphatases
- The phosphorylation may ↓ or ↑ the functional activity of protein

2- Glycosylation :

- 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 :

- As proline and lysine in endoplasmic reticulum by prolyl or lysyl hydroxylases (e.g. in collagen).

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 is covalently bound to the amino groups of lysine residues of biotin-dependent enzymes that catalyze carboxylation reactions
- Such as: pyruvate carboxylase

c- Farnesylated protein:

- Help anchor proteins in membranes.
- Note : many proteins are acetylated.