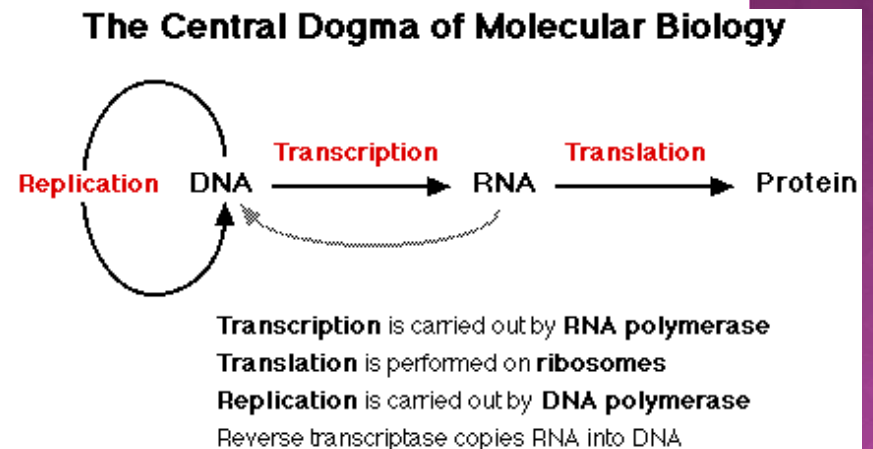


# TRANSCRIPTION & TRANSLATION

# CENTRAL DOGMA OF MOLECULAR BIOLOGY

- The flow of information in the cell starts at DNA, which replicates to form more DNA. Information is then ‘transcribed’ into RNA, and then it is “translated” into protein. The proteins do most of the work in the cell.
- Information does not flow in the other direction. This is a molecular version of the incorrectness of “inheritance of acquired characteristics”. Changes in proteins do not affect the DNA in a systematic manner (although they can cause random changes in DNA).

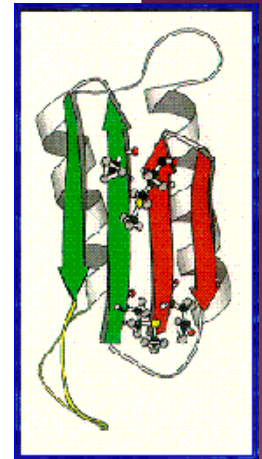


# REVERSE TRANSCRIPTION

- However, a few exceptions to the Central Dogma exist.
- Most importantly, some RNA viruses, called “retroviruses” make a DNA copy of themselves using the enzyme reverse transcriptase. The DNA copy incorporates into one of the chromosomes and becomes a permanent feature of the genome. The DNA copy inserted into the genome is called a “provirus”. This represents a flow of information from RNA to DNA.
- Closely related to retroviruses are “retrotransposons”, sequences of DNA that make RNA copies of themselves, which then get reverse-transcribed into DNA that inserts into new locations in the genome. Unlike retroviruses, retrotransposons always remain within the cell. They lack genes to make the protein coat that surrounds viruses.

# PRIONS

- A prion is an “infectious protein”.
- Prions are the agents that cause mad cow disease (bovine spongiform encephalopathy), chronic wasting disease in deer and elk, scrapie in sheep, and Creutzfeldt-Jakob syndrome in humans.
- These diseases cause neural degeneration. In humans, the symptoms are approximately those of Alzheimer’s syndrome accelerated to go from onset to death in about 1 year. Fortunately, the disease is very hard to catch and very rare, and they usually have a long incubation time. No cure is known, and not enough is known about how it is spread to do a thorough job of preventing it. Avoid eating brains is a good start though.
- The prion protein (PrP) is normally present in the body. Like all proteins, it is folded into a specific conformation, a state called PrP<sup>C</sup>. Prion diseases are caused by the same protein folding abnormally, a state called PrP<sup>Sc</sup>. A PrP<sup>Sc</sup> can bind to a normal protein and convert it to PrP<sup>Sc</sup>. This conversion spreads throughout the body, causing the disease to occur. It is also a form of inheritance that does not involve nucleic acids.



# RNA

- RNA plays a central role in the life of the cell. We are mostly going to look at its role in protein synthesis, but RNA does many other things as well.
- RNA can both store information (like DNA) and catalyze chemical reactions (like proteins).
- One theory for the origin of life has it starting out as RNA only, then adding DNA and proteins later. This theory is called the “RNA World”.
- RNA/protein hybrid structures are involved in protein synthesis (ribosome), splicing of messenger RNA, telomere maintenance, guiding ribosomes to the endoplasmic reticulum, and other tasks.
- Recently it has been found that very small RNA molecules are involved in gene regulation.

# RNA USED IN PROTEIN SYNTHESIS

- ◉ messenger RNA (mRNA). A copy of the gene that is being expressed. Groups of 3 bases in mRNA, called “codons” code for each individual amino acid in the protein made by that gene.
- ◉ ribosomal RNA (rRNA). Four different RNA molecules that make up part of the structure of the ribosome. They perform the actual catalysis of adding an amino acid to a growing peptide chain.
- ◉ transfer RNA (tRNA). Small RNA molecules that act as adapters between the codons of messenger RNA and the amino acids they code for.

# RNA VS. DNA

- RNA contains the sugar ribose; DNA contains deoxyribose.
- RNA contains the base uracil; DNA contains thymine instead.
- RNA is usually single stranded; DNA is usually double stranded.
- RNA is short: one gene long at most; DNA is long, containing many genes.

# TRANSCRIPTION

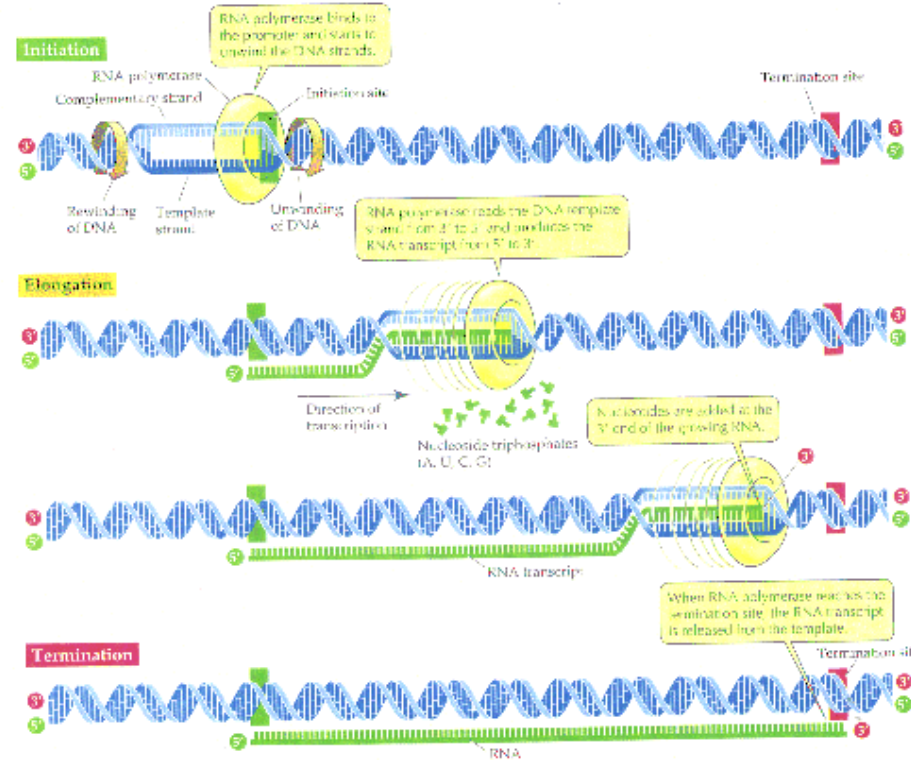
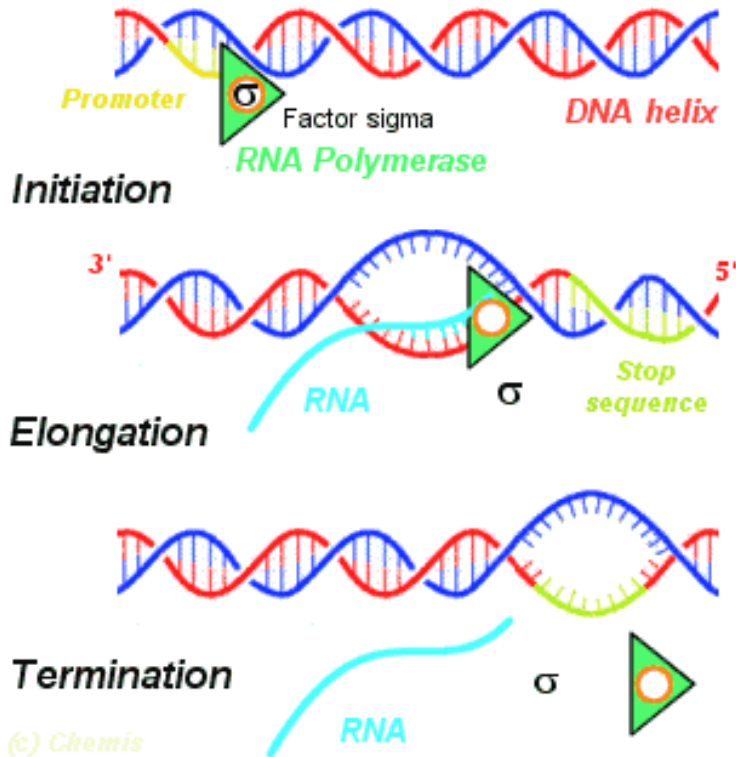
- ◉ Transcription is the process of making an RNA copy of a single gene. Genes are specific regions of the DNA of a chromosome.
- ◉ The enzyme used in transcription is “RNA polymerase”. There are several forms of RNA polymerase. In eukaryotes, most genes are transcribed by RNA polymerase 2.
- ◉ The raw materials for the new RNA are the 4 ribonucleoside triphosphates: ATP, CTP, GTP, and UTP. It’s the same ATP as is used for energy in the cell.
- ◉ As with DNA replication, transcription proceeds 5- to 3’: new bases are added to the free 3’ OH group.
- ◉ Unlike replication, transcription does not need to build on a primer. Instead, transcription starts at a region of DNA called a “promoter”. For protein-coding genes, the promoter is located a few bases 5’ to (upstream from) the first base that is transcribed into RNA.
- ◉ Promoter sequences are very similar to each other, but not identical. If many promoters are compared, a “consensus sequence” can be derived. All promoters would be similar to this consensus sequence, but not necessarily identical.



# PROCESS OF TRANSCRIPTION

- ◉ Transcription starts with RNA polymerase binding to the promoter.
- ◉ This binding only occurs under some conditions: when the gene is “on”. Various other proteins (transcription factors) help RNA polymerase bind to the promoter. Other DNA sequences further upstream from the promoter are also involved.
- ◉ Once it is bound to the promoter, RNA polymerase unwinds a small section of the DNA and uses it as a template to synthesize an exact RNA copy of the DNA strand.
- ◉ The DNA strand used as a template is the “coding strand”; the other strand is the “non-coding strand”. Notice that the RNA is made from 5’ end to 3’ end, so the coding strand is actually read from 3’ to 5’.
- ◉ RNA polymerase proceeds down the DNA, synthesizing the RNA copy.
- ◉ In prokaryotes, each RNA ends at a specific terminator sequence. In eukaryotes transcription doesn’t have a definite end point; the RNA is given a definitive termination point during RNA processing.

# TRANSCRIPTION GRAPHICS

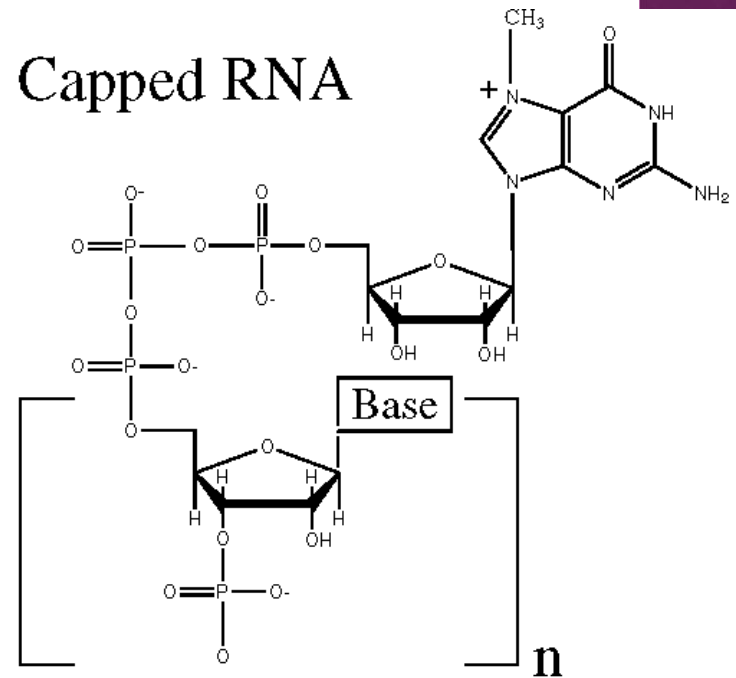


# AFTER TRANSCRIPTION

- In prokaryotes, the RNA copy of a gene is messenger RNA, ready to be translated into protein. In fact, translation starts even before transcription is finished.
- In eukaryotes, the primary RNA transcript of a gene needs further processing before it can be translated. This step is called “RNA processing”. Also, it needs to be transported out of the nucleus into the cytoplasm.
- Steps in RNA processing:
  - 1. Add a cap to the 5' end
  - 2. Add a poly-A tail to the 3' end
  - 3. splice out introns.

# CAPPING

- ◉ RNA is inherently unstable, especially at the ends. The ends are modified to protect it.
- ◉ At the 5' end, a slightly modified guanine (7-methyl G) is attached “backwards”, by a 5' to 5' linkage, to the triphosphates of the first transcribed base.
- ◉ At the 3' end, the primary transcript RNA is cut at a specific site and 100-200 adenine nucleotides are attached: the poly-A tail. Note that these A's are not coded in the DNA of the gene.

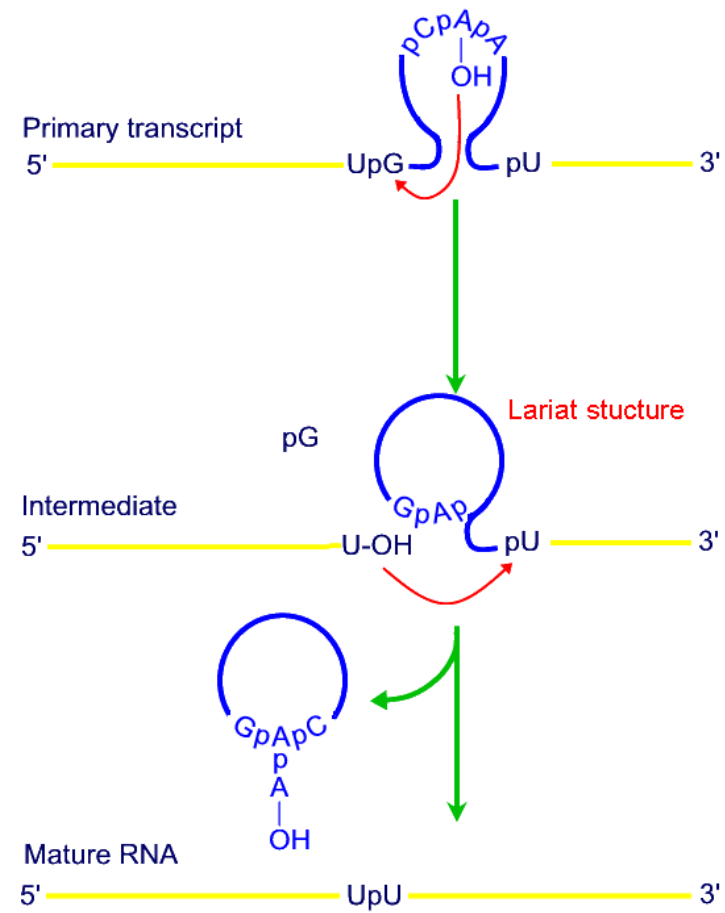


# INTRONS

- ◉ Introns are regions within a gene that don't code for protein and don't appear in the final mRNA molecule. Protein-coding sections of a gene (called exons) are interrupted by introns.
- ◉ The function of introns remains unclear. They may help in RNA transport or in control of gene expression in some cases, and they may make it easier for sections of genes to be shuffled in evolution. But, no generally accepted reason for the existence of introns exists.
- ◉ There are a few prokaryotic examples, but most introns are found in eukaryotes.
- ◉ Some genes have many long introns: the dystrophin gene (mutants cause muscular dystrophy) has more than 70 introns that make up more than 99% of the gene's sequence. However, not all eukaryotic genes have introns: histone genes, for example, lack introns.

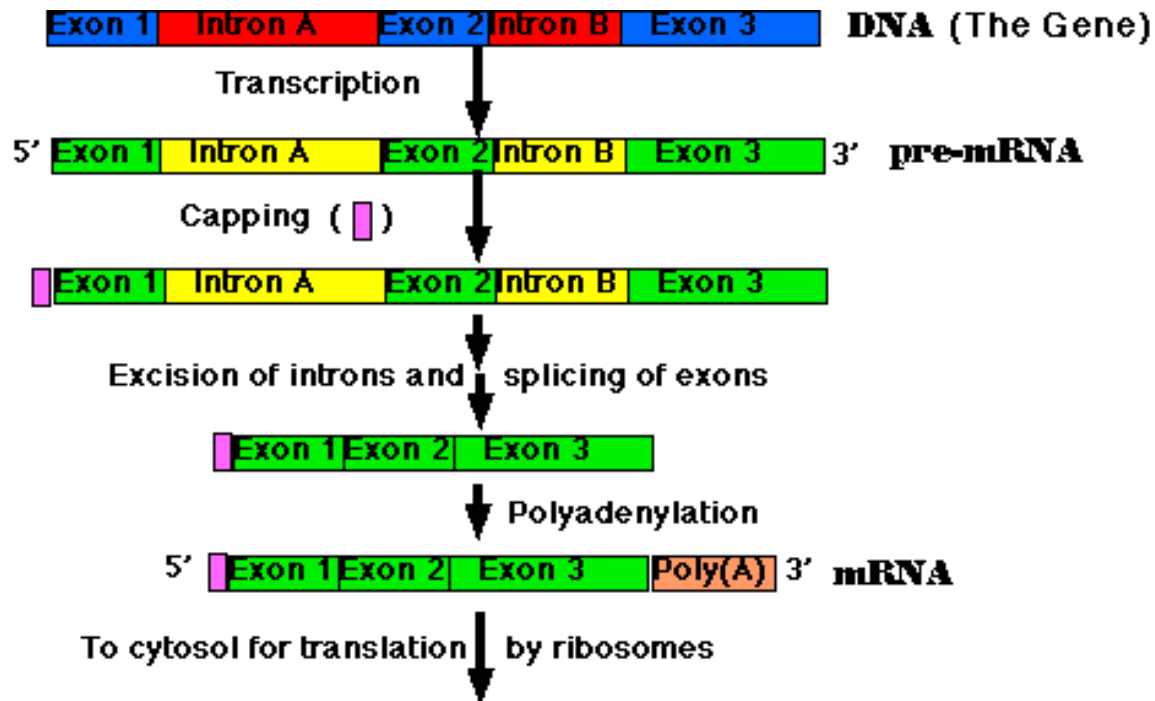
# INTRON SPLICING

- Introns are removed from the primary RNA transcript while it is still in the nucleus.
- Introns are “spliced out” by RNA/protein hybrids called “spliceosomes”. The intron sequences are removed, and the remaining ends are re-attached so the final RNA consists of exons only.



# SUMMARY OF RNA PROCESSING

- In eukaryotes, RNA polymerase produces a “primary transcript”, an exact RNA copy of the gene.
- A cap is put on the 5' end.
- The RNA is terminated and poly-A is added to the 3' end.
- All introns are spliced out.
- At this point, the RNA can be called messenger RNA. It is then transported out of the nucleus into the cytoplasm, where it is translated.



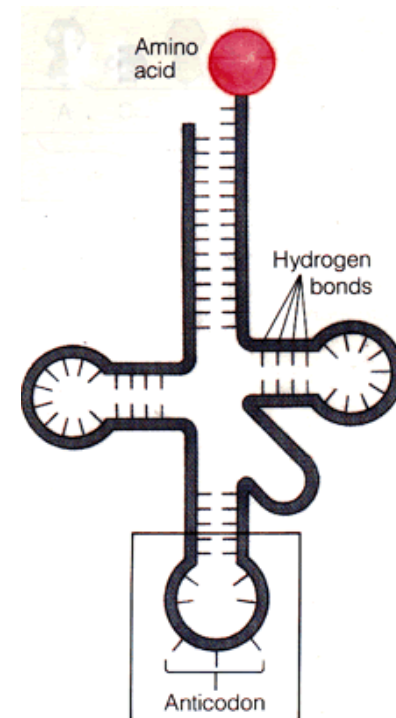
# TRANSLATION

- Translation of mRNA into protein is accomplished by the ribosome, an RNA/protein hybrid. Ribosomes are composed of 2 subunits, large and small.
- Ribosomes bind to the translation initiation sequence on the mRNA, then move down the RNA in a 5' to 3' direction, creating a new polypeptide. The first amino acid on the polypeptide has a free amino group, so it is called the “N-terminal”. The last amino acid in a polypeptide has a free acid group, so it is called the “C-terminal”.
- Each group of 3 nucleotides in the mRNA is a “codon”, which codes for 1 amino acids. Transfer RNA is the adapter between the 3 bases of the codon and the corresponding amino acid.



# TRANSFER RNA

- ◉ Transfer RNA molecules are short RNAs that fold into a characteristic cloverleaf pattern. Some of the nucleotides are modified to become things like pseudouridine and ribothymidine.
- ◉ Each tRNA has 3 bases that make up the anticodon. These bases pair with the 3 bases of the codon on mRNA during translation.
- ◉ Each tRNA has its corresponding amino acid attached to the 3' end. A set of enzymes, the “aminoacyl tRNA synthetases”, are used to “charge” the tRNA with the proper amino acid.
- ◉ Some tRNAs can pair with more than one codon. The third base of the anticodon is called the “wobble position”, and it can form base pairs with several different nucleotides.

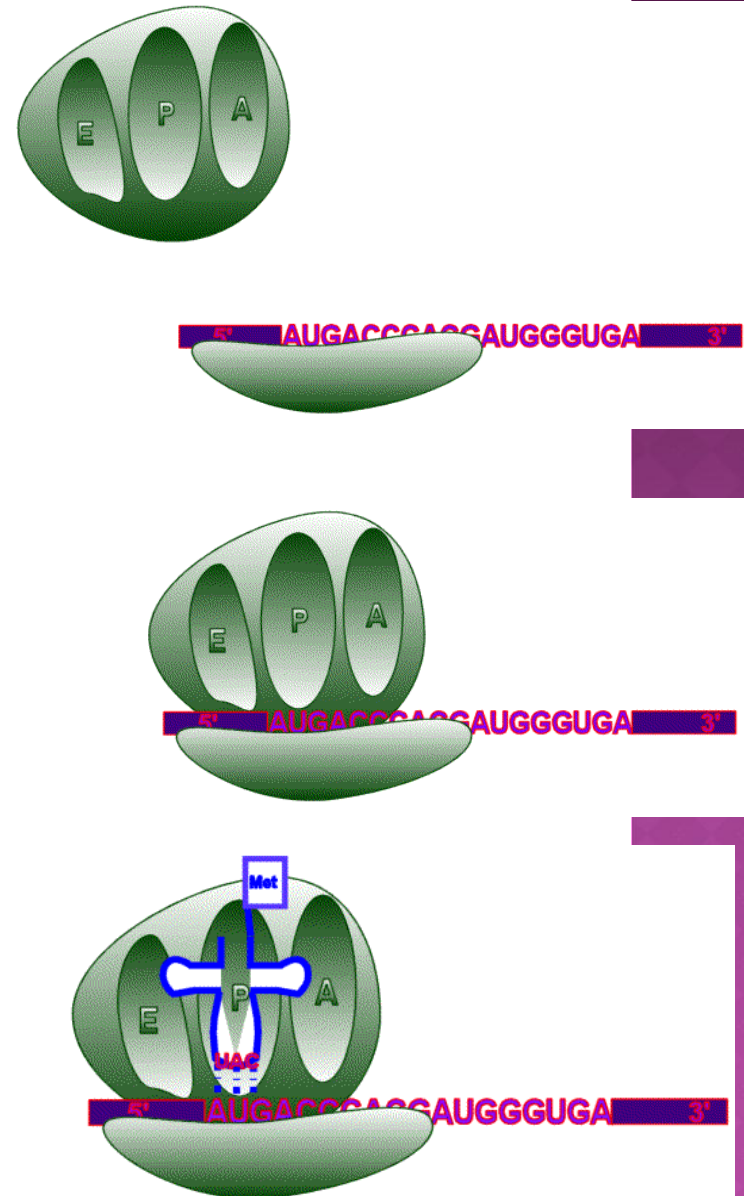


# INITIATION OF TRANSLATION

- In prokaryotes, ribosomes bind to specific translation initiation sites. There can be several different initiation sites on a messenger RNA: a prokaryotic mRNA can code for several different proteins. Translation begins at an AUG codon, or sometimes a GUG. The modified amino acid N-formyl methionine is always the first amino acid of the new polypeptide.
- In eukaryotes, ribosomes bind to the 5' cap, then move down the mRNA until they reach the first AUG, the codon for methionine. Translation starts from this point. Eukaryotic mRNAs code for only a single gene. (Although there are a few exceptions, mainly among the eukaryotic viruses).
- Note that translation does not start at the first base of the mRNA. There is an untranslated region at the beginning of the mRNA, the 5' untranslated region (5' UTR).

# MORE INITIATION

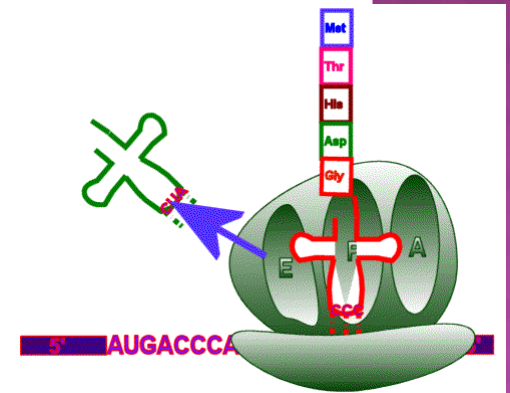
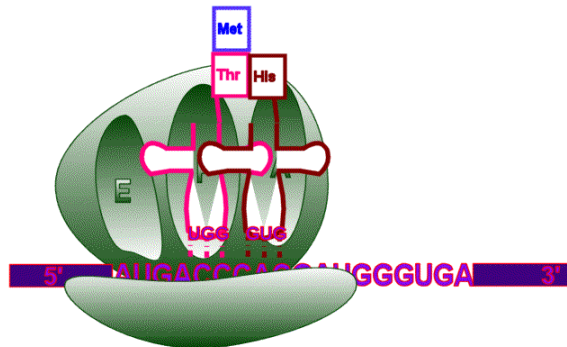
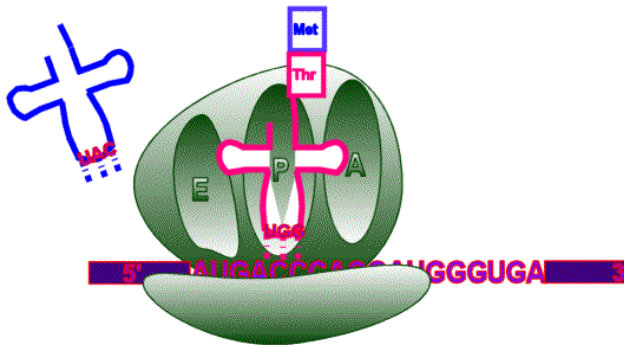
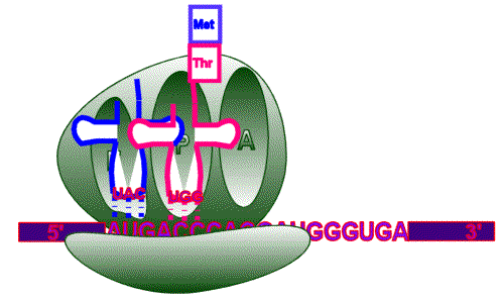
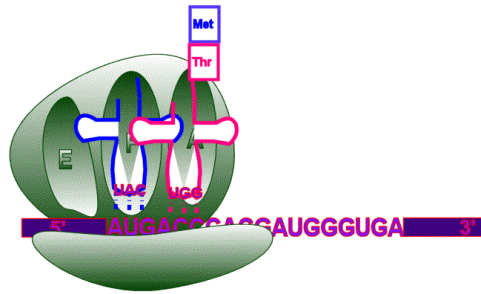
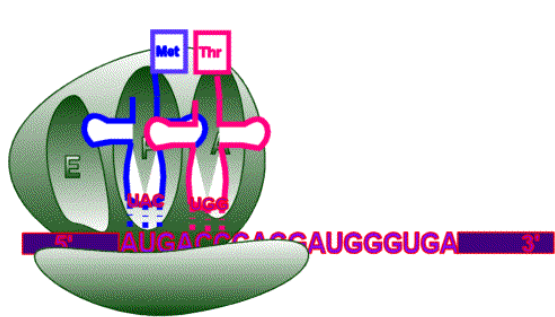
- The initiation process involves first joining the mRNA, the initiator methionine-tRNA, and the small ribosomal subunit. Several “initiation factors”--additional proteins--are also involved. The large ribosomal subunit then joins the complex.



# ELONGATION

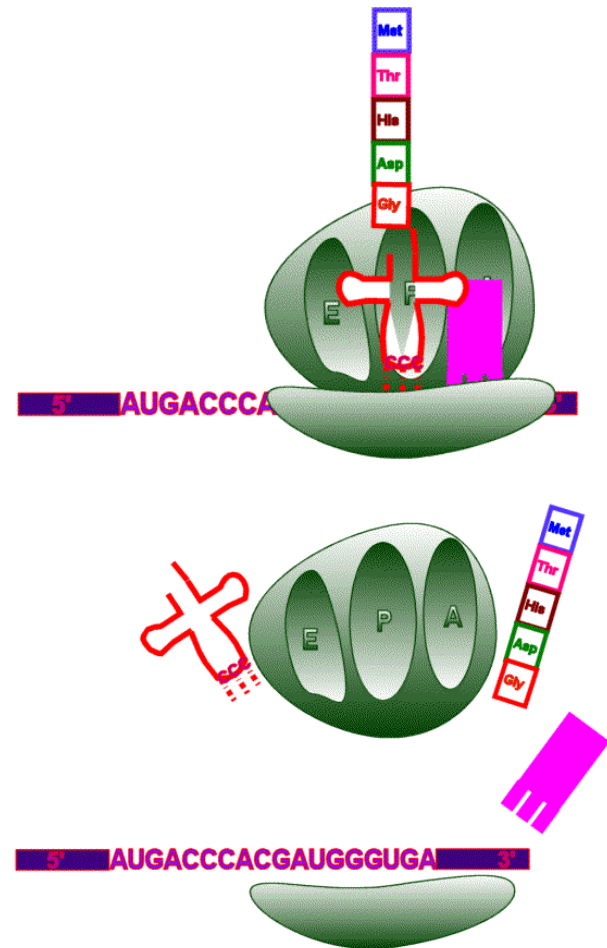
- ◉ The ribosome has 2 sites for tRNAs, called P and A. The initial tRNA with attached amino acid is in the P site. A new tRNA, corresponding to the next codon on the mRNA, binds to the A site. The ribosome catalyzes a transfer of the amino acid from the P site onto the amino acid at the A site, forming a new peptide bond.
- ◉ The ribosome then moves down one codon. The now-empty tRNA at the P site is displaced off the ribosome, and the tRNA that has the growing peptide chain on it is moved from the A site to the P site.
- ◉ The process is then repeated:
  - the tRNA at the P site holds the peptide chain, and a new tRNA binds to the A site.
  - the peptide chain is transferred onto the amino acid attached to the A site tRNA.
  - the ribosome moves down one codon, displacing the empty P site tRNA and moving the tRNA with the peptide chain from the A site to the P site.

# ELONGATION



# TERMINATION

- Three codons are called “stop codons”. They code for no amino acid, and all protein-coding regions end in a stop codon.
- When the ribosome reaches a stop codon, there is no tRNA that binds to it. Instead, proteins called “release factors” bind, and cause the ribosome, the mRNA, and the new polypeptide to separate. The new polypeptide is completed.
- Note that the mRNA continues on past the stop codon. The remaining portion is not translated: it is the 3’ untranslated region (3’ UTR).



# POST-TRANSLATIONAL MODIFICATION

- New polypeptides usually fold themselves spontaneously into their active conformation. However, some proteins are helped and guided in the folding process by chaperone proteins
- Many proteins have sugars, phosphate groups, fatty acids, and other molecules covalently attached to certain amino acids. Most of this is done in the endoplasmic reticulum.
- Many proteins are targeted to specific organelles within the cell. Targeting is accomplished through “signal sequences” on the polypeptide. In the case of proteins that go into the endoplasmic reticulum, the signal sequence is a group of amino acids at the N terminal of the polypeptide, which are removed from the final protein after translation.

# THE GENETIC CODE

- ◉ Each group of 3 nucleotides on the mRNA is a codon. Since there are 4 bases, there are  $4^3 = 64$  possible codons, which must code for 20 different amino acids.
- ◉ More than one codon is used for most amino acids: the genetic code is “degenerate”. This means that it is not possible to take a protein sequence and deduce exactly the base sequence of the gene it came from.
- ◉ In most cases, the third base of the codon (the wobble base) can be altered without changing the amino acid.
- ◉ AUG is used as the start codon. All proteins are initially translated with methionine in the first position, although it is often removed after translation. There are also internal methionines in most proteins, coded by the same AUG codon.
- ◉ There are 3 stop codons, also called “nonsense” codons. Proteins end in a stop codon, which codes for no amino acid.



Second letter

		Second letter									
		U	C	A	G						
First letter	U	UUU UUC	UCU UCC UCA UCG	UAU UAC	UGU UGC	U C A G					
		UUA UUG		UAA UAG			UGA UGG				
		CUU CUC CUA CUG		CCU CCC CCA CCG			CAU CAC	CGU CGC CGA CGG			
							CAA CAG				
A	A	AAU AUC AUA	ACU ACC ACA ACG	AAU AAC	AGU AGC	U C A G					
		AUG		AAA AAG			AGA AGG				
		G		G			GUU GUC GUA GUG	GCU GCC GCA GCG	GAU GAC	GGU GGC GGA GGG	U C A G
									GAA GAG		

Third letter

# The Genetic Code

	U	C	A	G	
U	<p>UUU <b>Ph</b>enyl alanine</p> <p>UUC</p> <p>UUG <b>L</b>eucine</p> <p>UUA</p>	<p>UCU</p> <p>UCC <b>S</b>erine</p> <p>UCA</p> <p>UCG</p>	<p>UAU <b>T</b>yr osine</p> <p>UAC</p> <p>UAA <b>S</b>top</p> <p>UAG</p>	<p>UGU <b>C</b>ys teine</p> <p>UGC</p> <p>UGA <b>S</b>top</p> <p>UGG <b>T</b>ryp tophan</p>	U C A G
C	<p>CUU</p> <p>CUC <b>L</b>eucine</p> <p>CUA</p> <p>CUG</p>	<p>CCU</p> <p>CCC <b>P</b>ro line</p> <p>CCA</p> <p>CCG</p>	<p>CAU <b>H</b>is tidine</p> <p>CAC</p> <p>CAA <b>G</b>lu tamine</p> <p>CAG</p>	<p>CGU</p> <p>CGC <b>A</b>rg inine</p> <p>CGA</p> <p>CGG</p>	U C A G
A	<p>AUU</p> <p>AUC <b>I</b>so leucine</p> <p>AUA</p> <p>AUG <b>M</b>eth ionine</p>	<p>ACU</p> <p>ACC <b>T</b>h reonine</p> <p>ACA</p> <p>ACG</p>	<p>AAU <b>A</b>s paragine</p> <p>AAC</p> <p>AAA <b>L</b>ys ine</p> <p>AAG</p>	<p>AGU <b>S</b>er ine</p> <p>AGC</p> <p>AGA <b>A</b>rg inine</p> <p>AGG</p>	U C A G
G	<p>GUU</p> <p>GUC <b>V</b>al ine</p> <p>GUA</p> <p>GUG</p>	<p>GCU</p> <p>GCC <b>A</b>la nine</p> <p>GCA</p> <p>GCG</p>	<p>GAU <b>A</b>s partic acid</p> <p>GAC</p> <p>GAA <b>G</b>lu tamic acid</p> <p>GAG</p>	<p>GGU</p> <p>GGC <b>G</b>ly cine</p> <p>GGA</p> <p>GGG</p>	U C A G

# MORE GENETIC CODE

- ◉ The genetic code is almost universal. It is used in both prokaryotes and eukaryotes.
- ◉ However, some variants exist, mostly in mitochondria which have very few genes.
- ◉ For instance, CUA codes for leucine in the universal code, but in yeast mitochondria it codes for threonine. Similarly, AGA codes for arginine in the universal code, but in human and *Drosophila* mitochondria it is a stop codon.
- ◉ There are also a few known variants in the code used in nuclei, mostly among the protists.