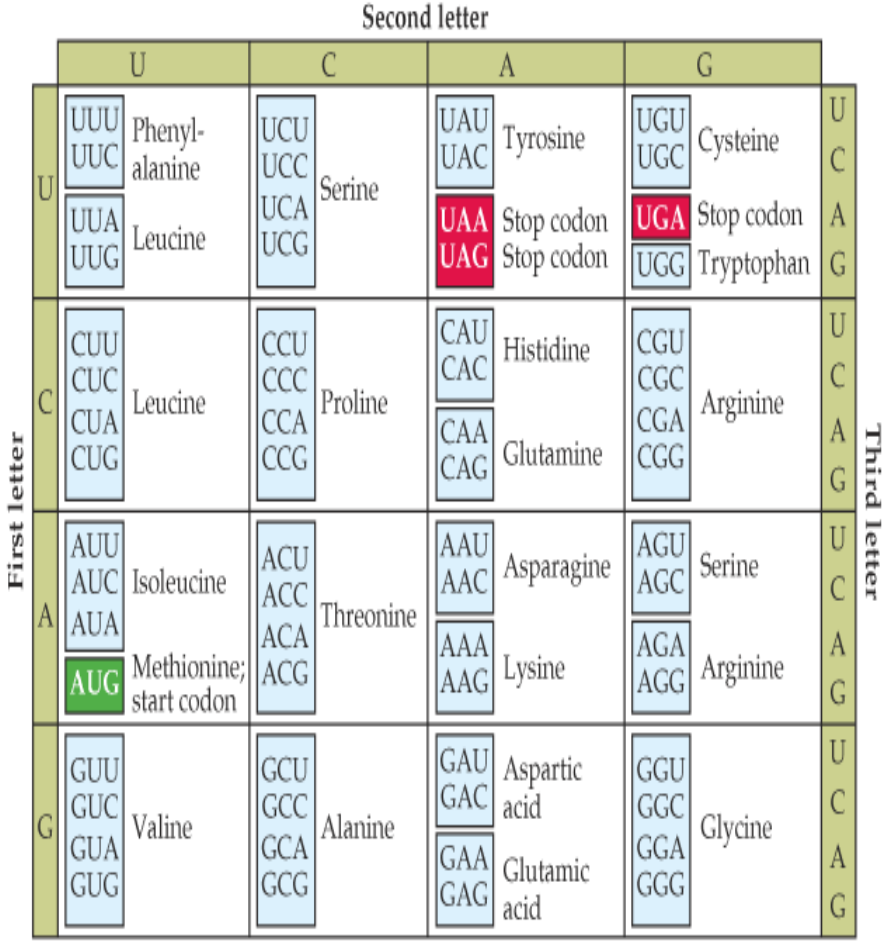
Molecular Biology
Part IV

Cell Information: Instruction book of Life

- DNA, RNA, and Proteins are examples of strings written in either the four-letter nucleotide of DNA and RNA (A C G T/U)
- or the twenty-letter amino acid of proteins. Each amino acid is coded by 3 nucleotides called codon. (Leu, Arg, Met, etc.)



The Genetic Code

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

The genetic code

GCA GCC GCG GCU	AGA AGG CGA CGC CGG CGU	GAC GAU	AAC AAU	UGC UGU	GAA GAG	CAA CAG	GGA GGC GGG GGU	CAC CAU	AUA AUC AUU	UUA UUG CUA CUC CUG CUU	AAA AAG	AUG	UUC UUU	CCA CCC CCG CCU	AGC AGU UCA UCC UCG UCU	ACA ACC ACG ACU	UGG	UAC UAU	GUA GUC GUG GUU	UAA UAG UGA
Ala	Arg	Asp	Asn	Cys	Glu	Gln	Gly	His	Ile	Leu	Lys	Met	Phe	Pro	Ser	Thr	Trp	Tyr	Val	stop
A	R	D	N	C	E	Q	G	H	I	L	K	M	F	P	S	T	W	Y	V	

Code Features

All possible 64 triplets have a meaning.

The codes are not unique for a particular amino acid except for *met (AUG)* and *trp (UGG)*.

If an amino acid has multiple codes, its first two letters are the same.

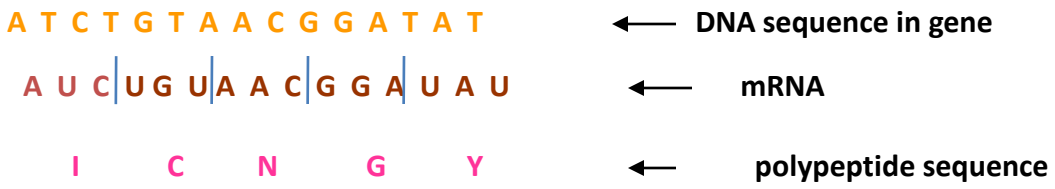
The codes UAA, UAG and UGA denote '***stop***' and do not represent any protein.

The code AUG appears at the beginning of every protein and it is therefore recognized as an **initiation code** but also has the significance that all proteins are synthesized with a methionine at the beginning (but later may be discarded within the cell). Methionine may not be always an initiation code and may also appear in the middle of a protein.

The genetic code is not universal in the sense that it applies only to nuclear genes. The mitochondrial genes use a slightly different code and also expressed differently using different ribosome and tRNAs.

The Genetic Code

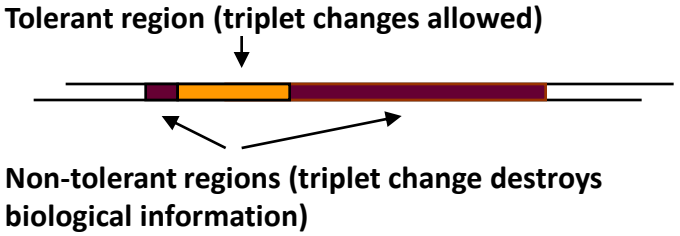
We have stated earlier: one gene, one protein. But, how a gene defines a protein uniquely? It took several years since the discovery of double helix by famous biologists (all Nobel Laureates: Severo Ochoa, Marshall Nirenberg and Gobind Khorana, Crick and Brenner) to decipher what is called the *genetic code*. The first observation was the order of nucleotides in the gene directly determine the order of amino acid in the polypeptide (protein).



The second important discovery was about the size of the code word. Since DNA alphabet has 4 symbols (A,C,T,G) [although U rather than T appears in mRNA, the code is stated in terms of DNA alphabet], two Symbols can code at most 4^2=16 polypeptide, so we need at least three letters since there are 20 proteins. Three it is – from elementary computer science principle – giving an excess number of 4^3=64 combinations! Nature has used all 64 combinations incorporating redundancy and robustness!! But, establishing this by biological experiments was an extremely challenging task.

Reading Frame

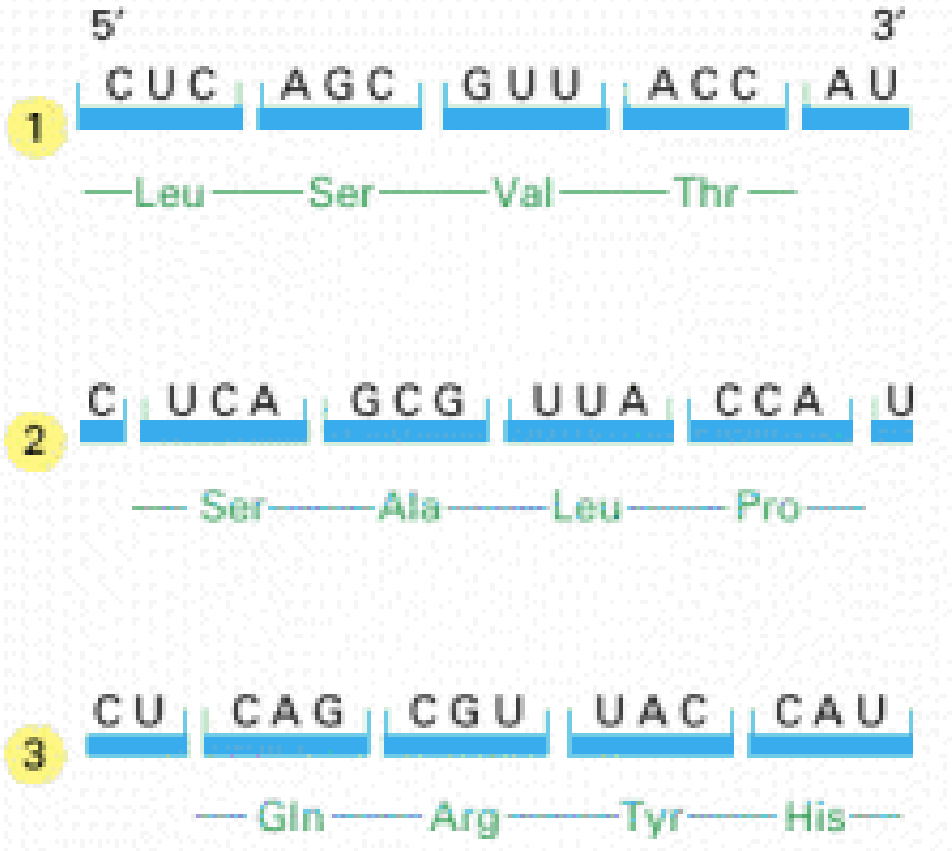
The linearity of gene and protein was established by a simple experiment. But, the experiment to proving the triplet nature of the code took some detailed Biology work. First there are these acridine dyes that can cause deletion or insertion of just one base pair in a double stranded DNA. It was also discovered that there are some proteins in which a segment of amino acids can be changed without altering its function. This portion of the protein is called ***tolerant*** *regions*.



Triplet Nature of Gene Coding

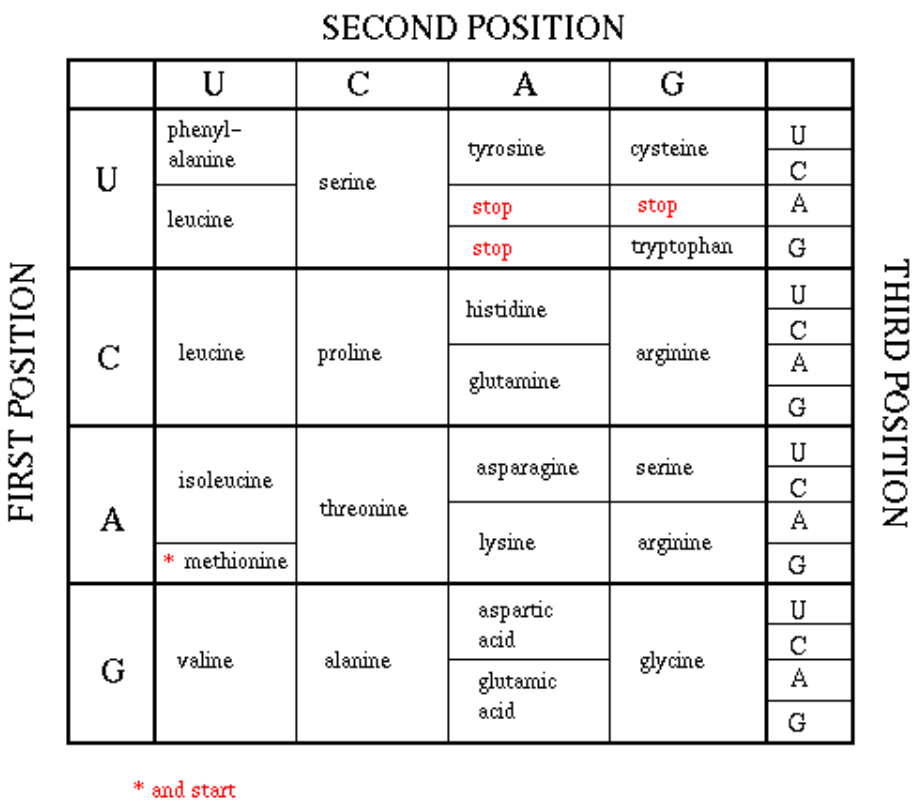
If regions of the gene corresponding to this tolerant region, is changed one at time what effect will have? If the code is a triplet and we delete or insert one bp in the gene, then all the amino acids downstream will be changed yielding a non-functional protein. If we delete or insert two bp in the gene, it will also have the same effect. But if we insert or delete three bp, then the correct **reading frame** in non-tolerant region will be restored and the protein will be functional again. Crick and Brenner performed an elegant experiment based on this principle and established the triplet nature of gene coding. Nirenberg, Holley and Govind Khorana actually later deciphered the genetic code table and shared Nobel prize in 1968**.**

Reading frames



Translation

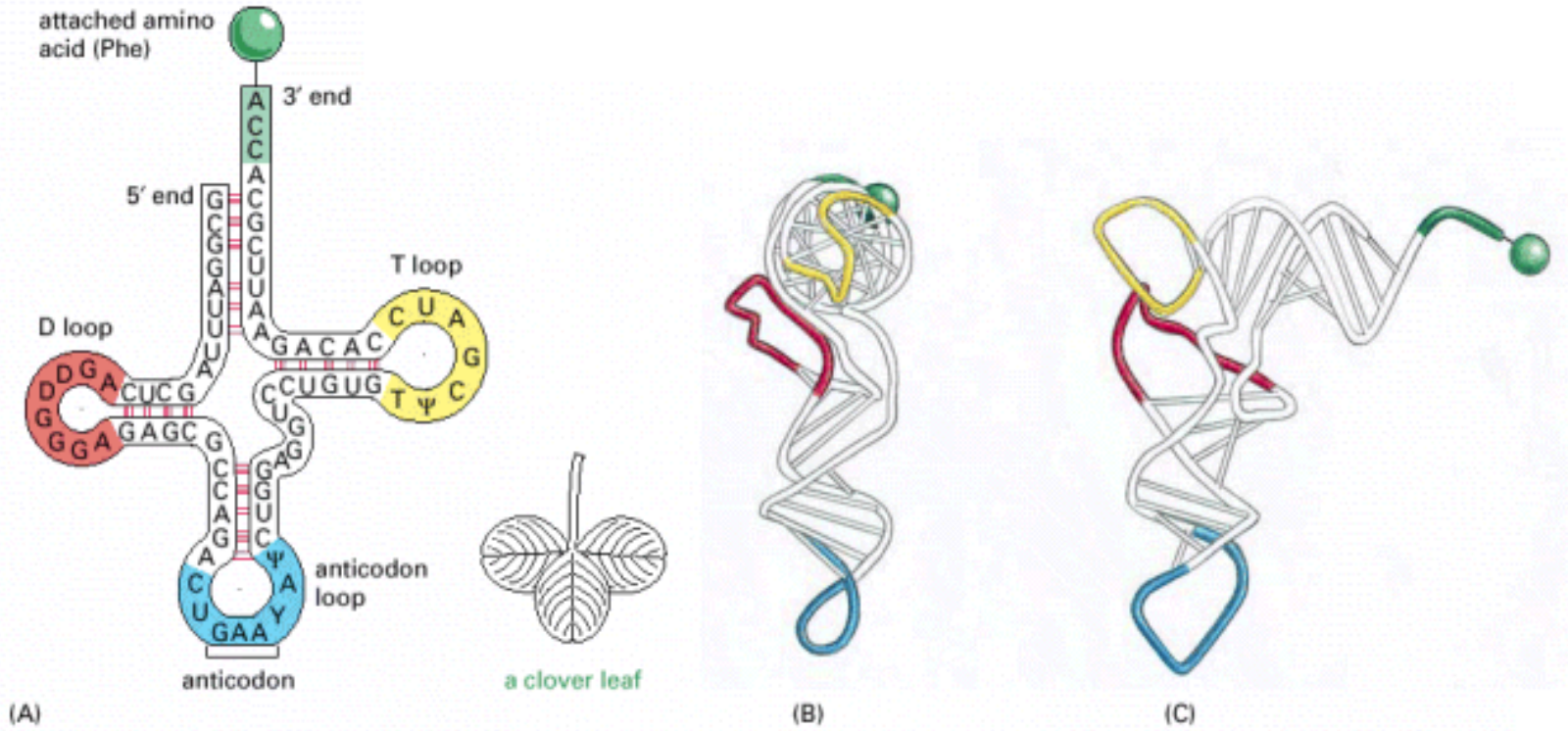
- The process of going from RNA to polypeptide.
- Three base pairs of RNA (called a codon) correspond to one amino acid based on a fixed table.
- Always starts with Methionine and ends with a stop codon



FIRST POSITION	SECOND POSITION: U	C	A	G	THIRD POSITION
U	phenyl-alanine	serine	tyrosine	cysteine	U
					C
	leucine		stop	stop	A
			stop	tryptophan	G
C	leucine	proline	histidine	arginine	U
					C
			glutamine		A
					G
A	isoleucine	threonine	asparagine	serine	U
					C
	* methionine		lysine	arginine	A
					G
G	valine	alanine	aspartic acid	glycine	U
					C
			glutamic acid		A
					G

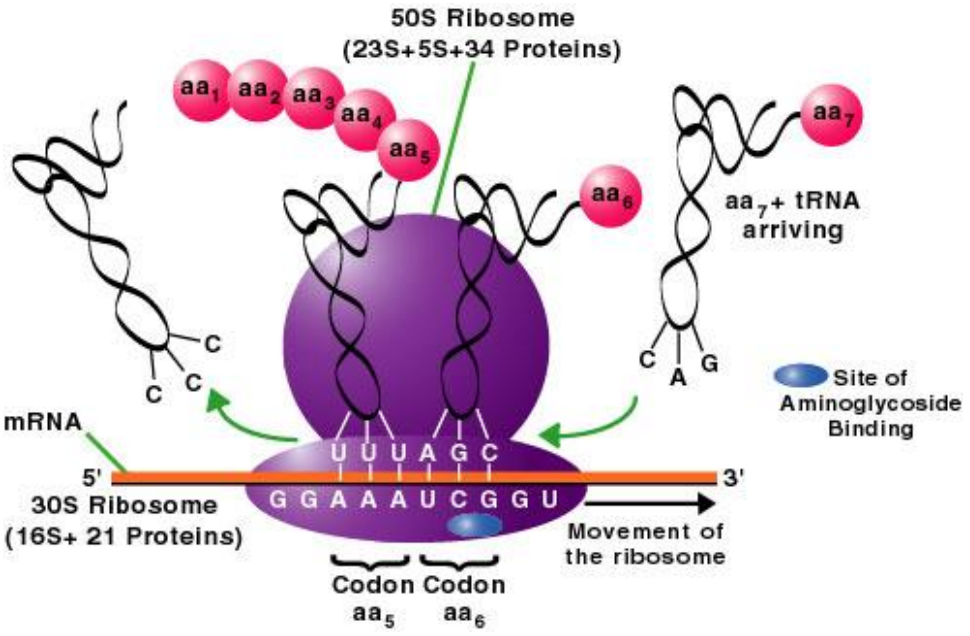
* and start

tRNA


attached amino acid (Phe)
3′ end
5′ end
T loop
D loop
anticodon loop
anticodon
a clover leaf
(A)
(B)
(C)
5′ GCGGAUUUAGCUCAGDDGGGAGAGCGCCAGACUGAAYAΨCUGGAGGUCCUGUGTΨCGAUCCACAGAAUUCGCACCA 3′
anticodon
(D)

Translation, continued

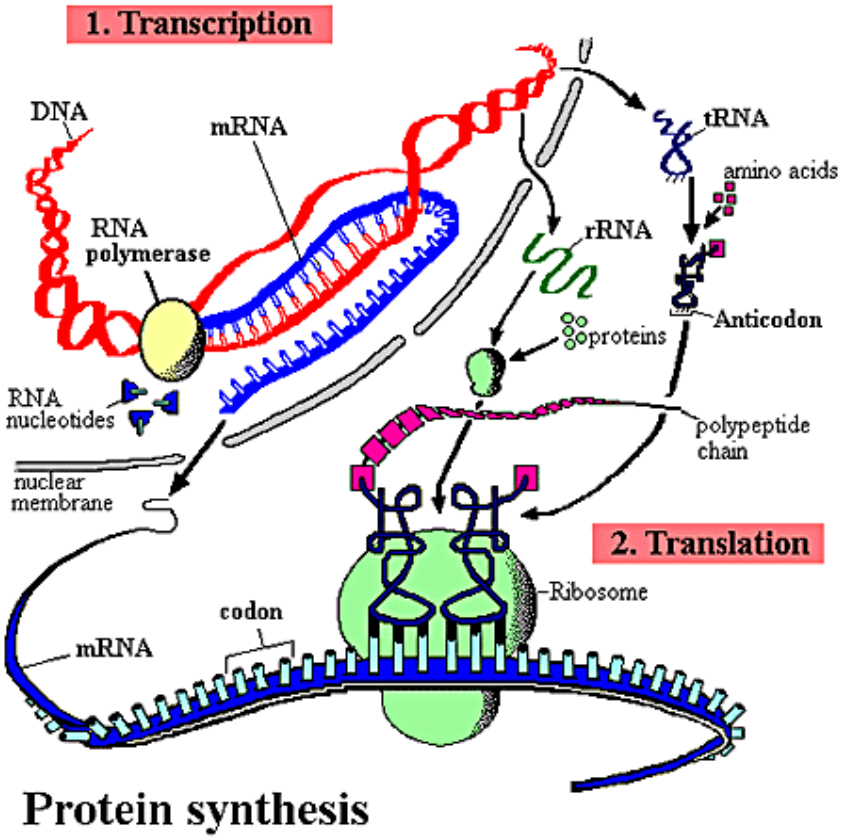
- Catalyzed by Ribosome
- Using two different sites, the Ribosome continually binds tRNA, joins the amino acids together and moves to the next location along the mRNA
- ~10 codons/second, but multiple translations can occur simultaneously



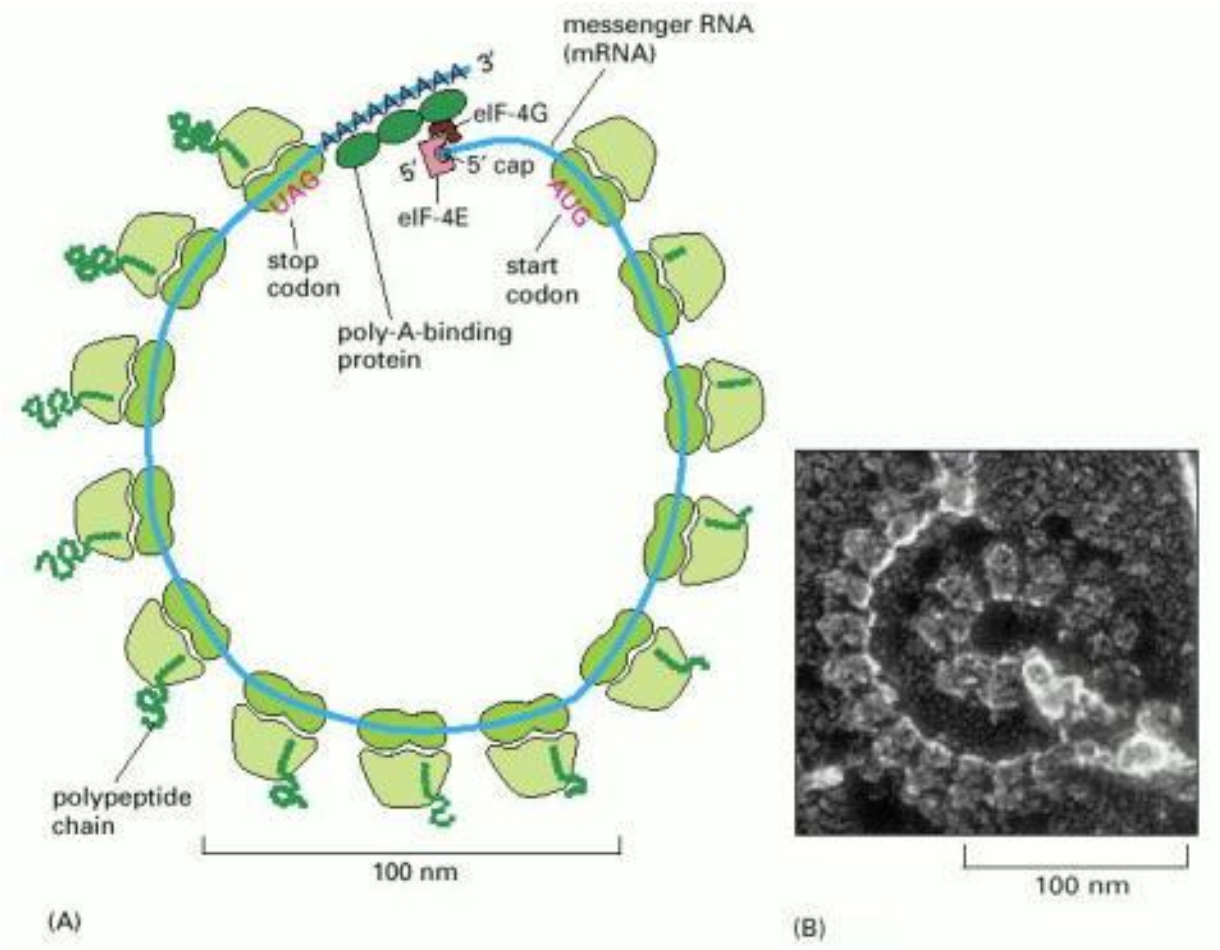
http://wong.scripps.edu/PIX/ribosome.jpg

Protein Synthesis: Summary

- There are twenty amino acids, each coded by three-base-sequences in DNA, called “codons”
  - This code is degenerate
- The **central dogma** describes how proteins derive from DNA
  - DNA → mRNA → (splicing?) → protein
- The protein adopts a 3D structure specific to it’s amino acid arrangement and function



Simultaneous translation



Translation
(This description is from the book “Cell”)

The information in the sequence of a messenger RNA molecule is read out in groups of three nucleotides at a time: each triplet of nucleotides, or *codon*, specifies (codes for) a single amino acid in a corresponding protein. Since there are 64 (= 4 × 4 × 4) possible codons, but only 20 amino acids, there are necessarily many cases in which several codons correspond to the same amino acid. The code is read out by a special class of small RNA molecules, the transfer RNAs (tRNAs). Each type of tRNA becomes attached at one end to a specific amino acid, and displays at its other end a specific sequence of three nucleotides—an *anticodon*—that enables it to recognize, through base-pairing, a particular codon or subset of codons in mRNA (Figure 1–9).

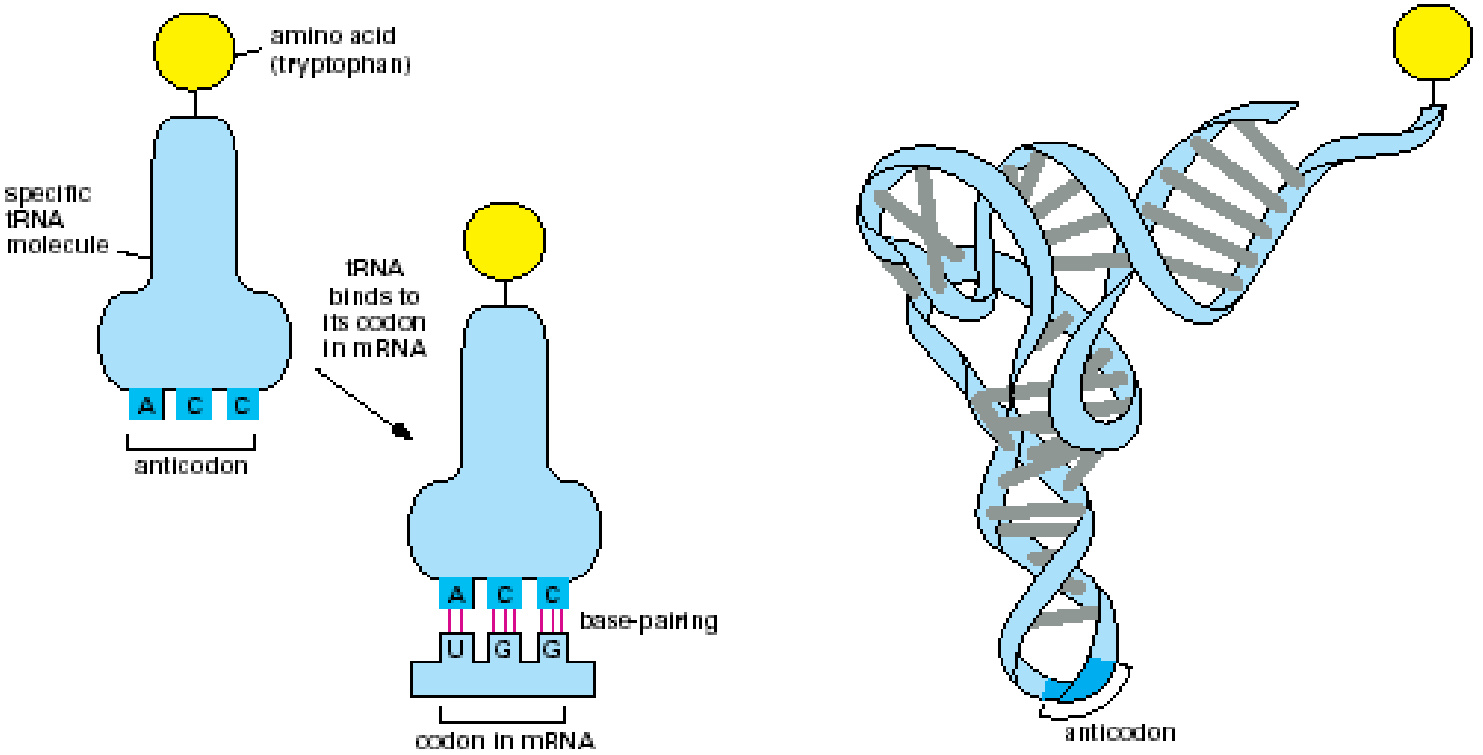
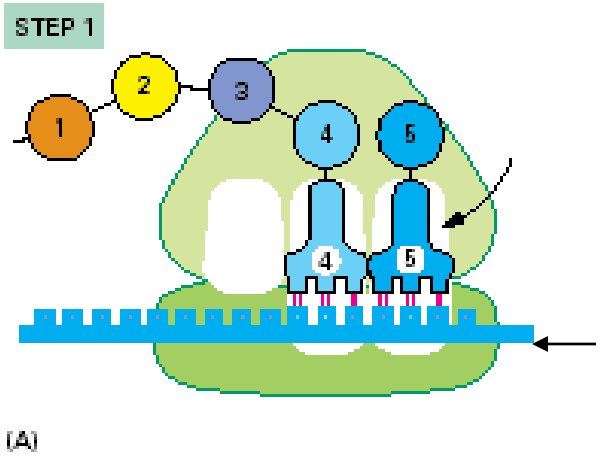


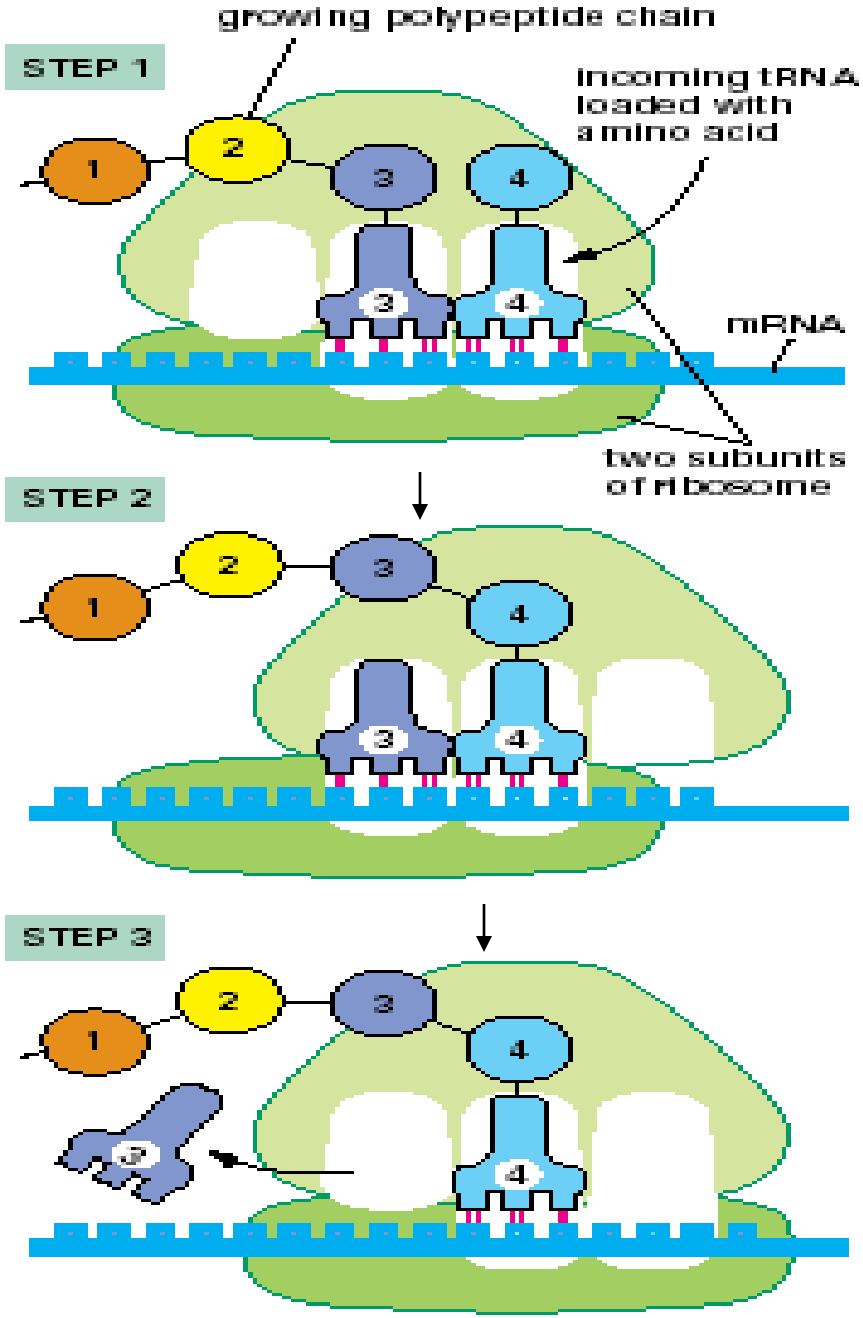
Figure 1–9 Transfer RNA. (A) A tRNA molecule specific for the amino acid tryptophan. One end of the tRNA molecule has tryptophan attached to it, while the other end displays the triplet nucleotide sequence CCA (its anticodon), which recognizes the tryptophan codon in messenger RNA molecules. (B) The three-dimensional structure of the tryptophan tRNA molecule. Note that the codon and the anticodon in (A) are in antiparallel orientations, like the two strands in a DNA double helix (see Figure 1–2), so that the sequence of the anticodon in the tRNA is read from right to left, while that of the codon in the mRNA is read from left to right.

For synthesis of protein, a succession of tRNA molecules charged with their appropriate amino acids have to be brought together with an mRNA molecule and matched up by base-pairing through their anticodons 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. This whole complex of processes is carried out by a giant multimolecular machine, the ribosome, formed of two main chains of RNA, called **ribosomal RNAs (rRNAs)**, and more than 50 different proteins. This evolutionarily ancient molecular juggernaut latches onto the end of an mRNA molecule and then trundles along it, capturing loaded tRNA molecules and stitching together the amino acids they carry to form a new protein chain (Figure 1–10).

Figure 1–10 A ribosome at work. (A) The diagram shows how a ribosome moves along an mRNA molecule, capturing tRNA molecules that match the codons in the mRNA and using them to join amino acids into a protein chain. The mRNA specifies the sequence of amino acids. (B) The

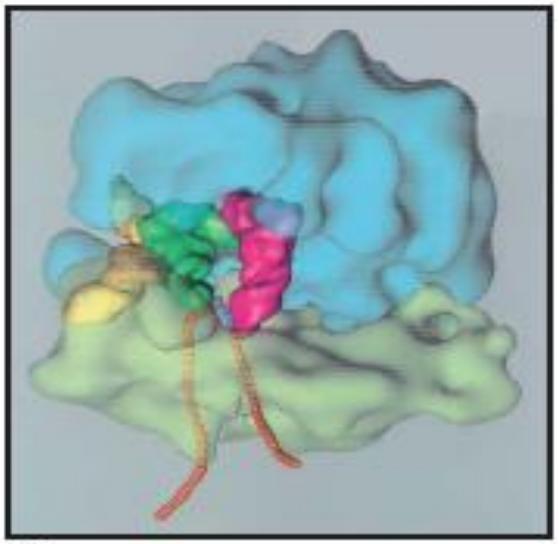


(A)



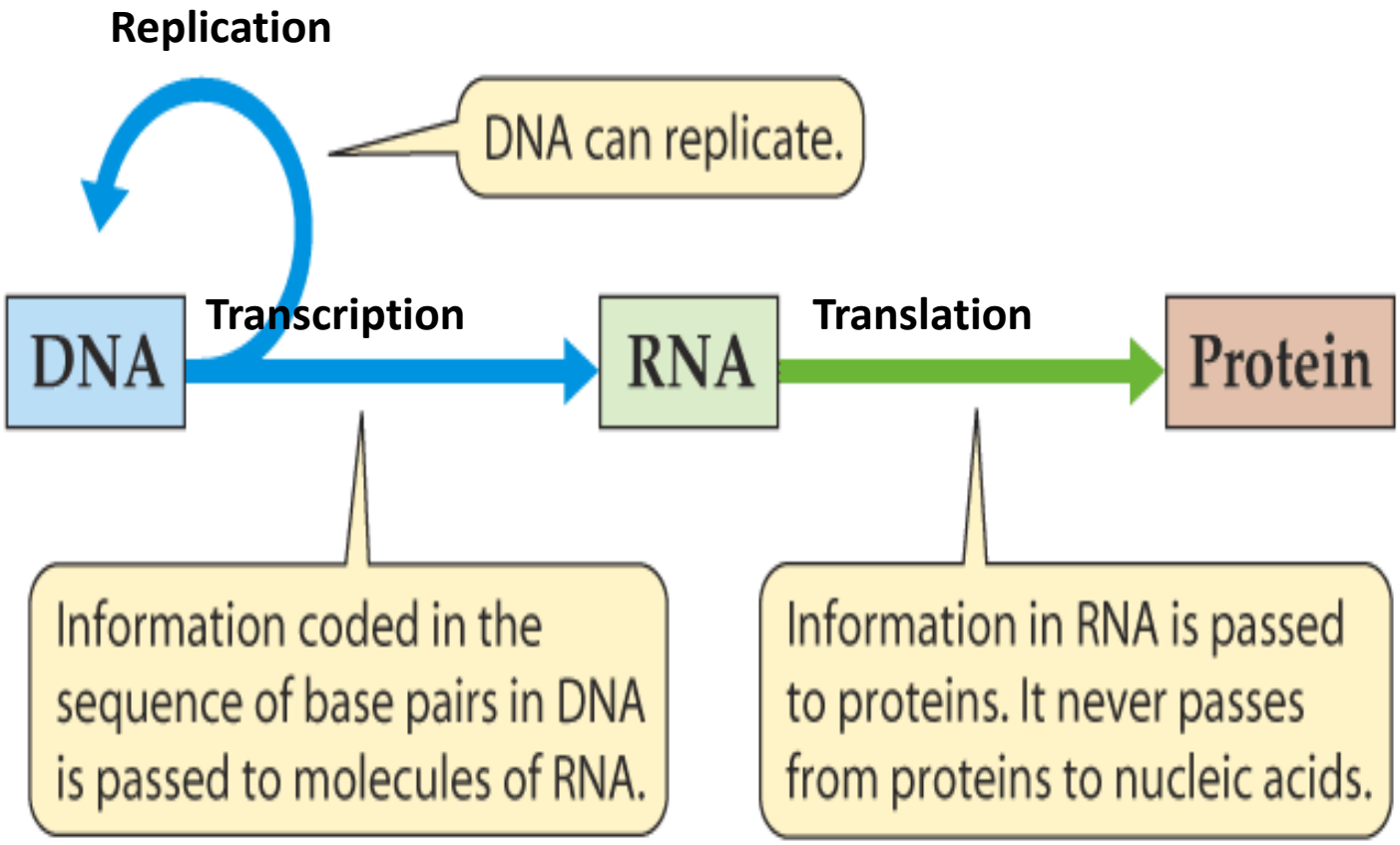
1.10 (B) The

three-dimensional structure of a bacterial ribosome (pale green and blue), moving along an mRNA molecule (orange beads), with three tRNA molecules (yellow, green, and pink) at different stages in their process of capture and release. The ribosome is a giant assembly of more than 50 individual protein and RNA molecules. (B, courtesy of Joachim Frank, Yanhong Li, and Rajendra Agarwal.)



(B)

DNA, RNA, and the Flow of Information



Protein Synthesis: Summary

- There are twenty amino acids, each coded by three-base-sequences in DNA, called “codons”
  - This code is degenerate
- The **central dogma** describes how proteins derive from DNA
  - DNA → mRNA → (splicing?) → protein
- The protein adopts a 3D structure specific to it’s amino acid arrangement and function