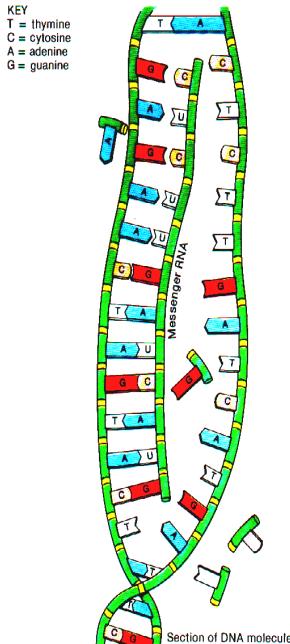
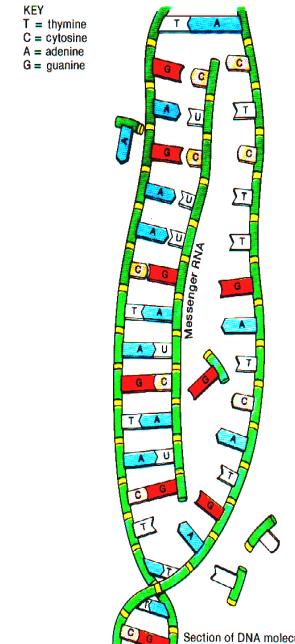


Bioinformatics

ONE Introduction to Biology



Wendy Lee
Dept of Computer Science
San José State University
Biology/CS/SE 123A
Fall 2014



bi·o·in·for·mat·ics

/bīō, īnfər'matiks/

noun

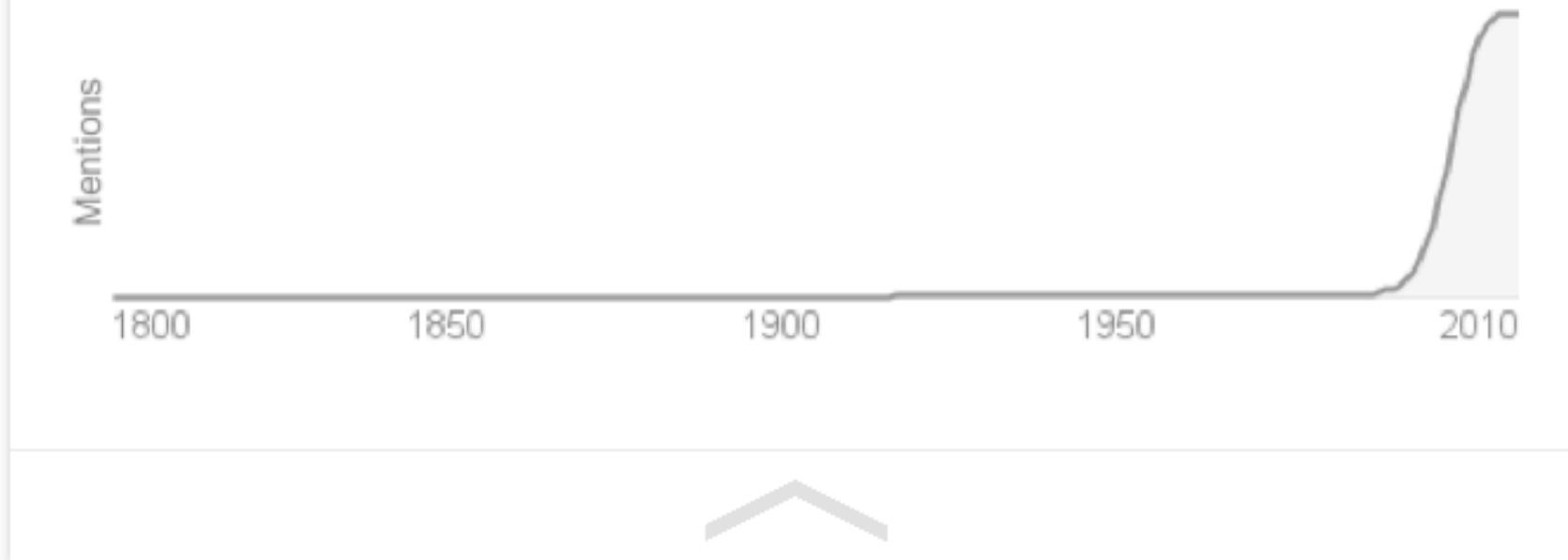
noun: bioinformatics; noun: bio-informatics

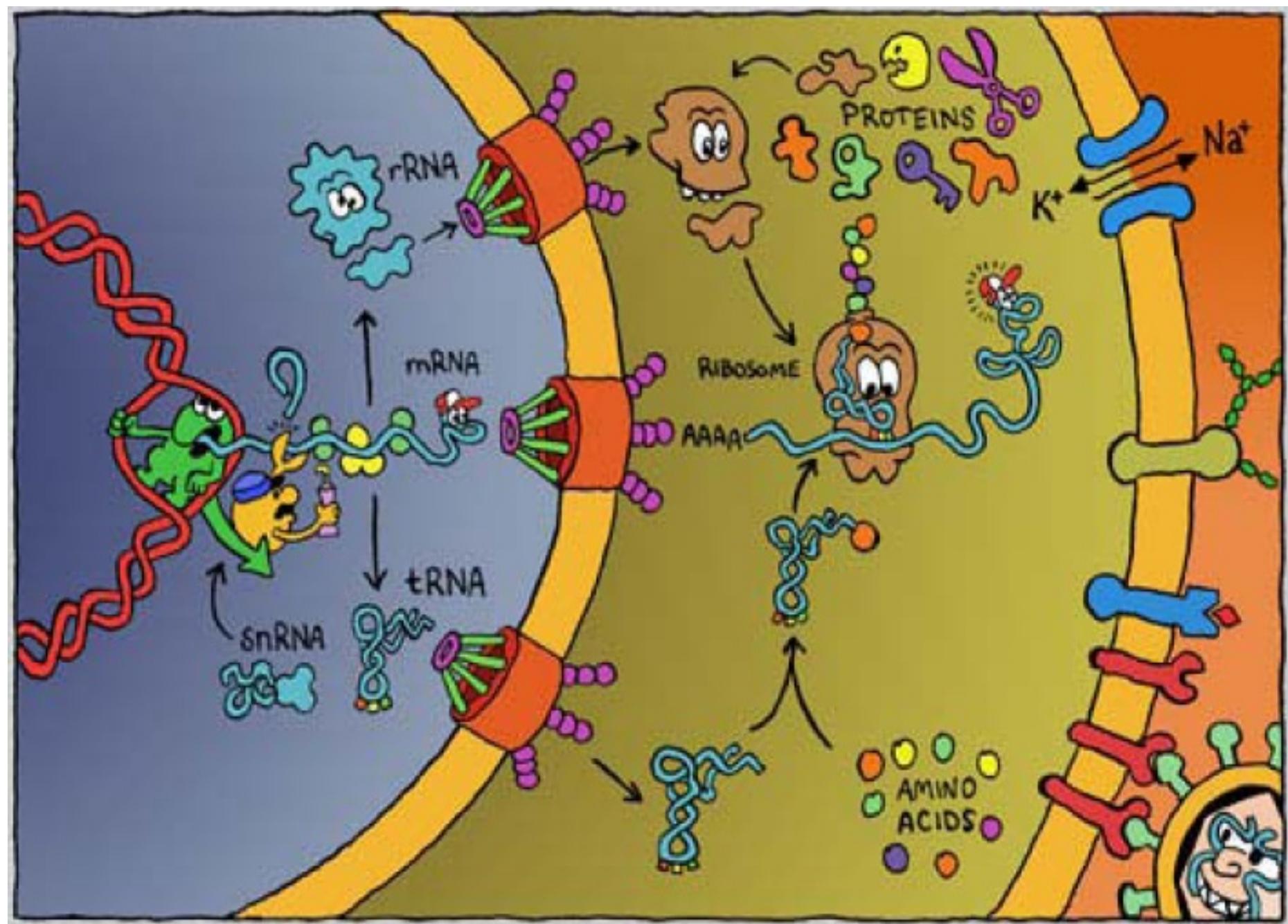
the science of collecting and analyzing complex biological data such as genetic codes.

Translate bioinformatics to

Choose language ▾

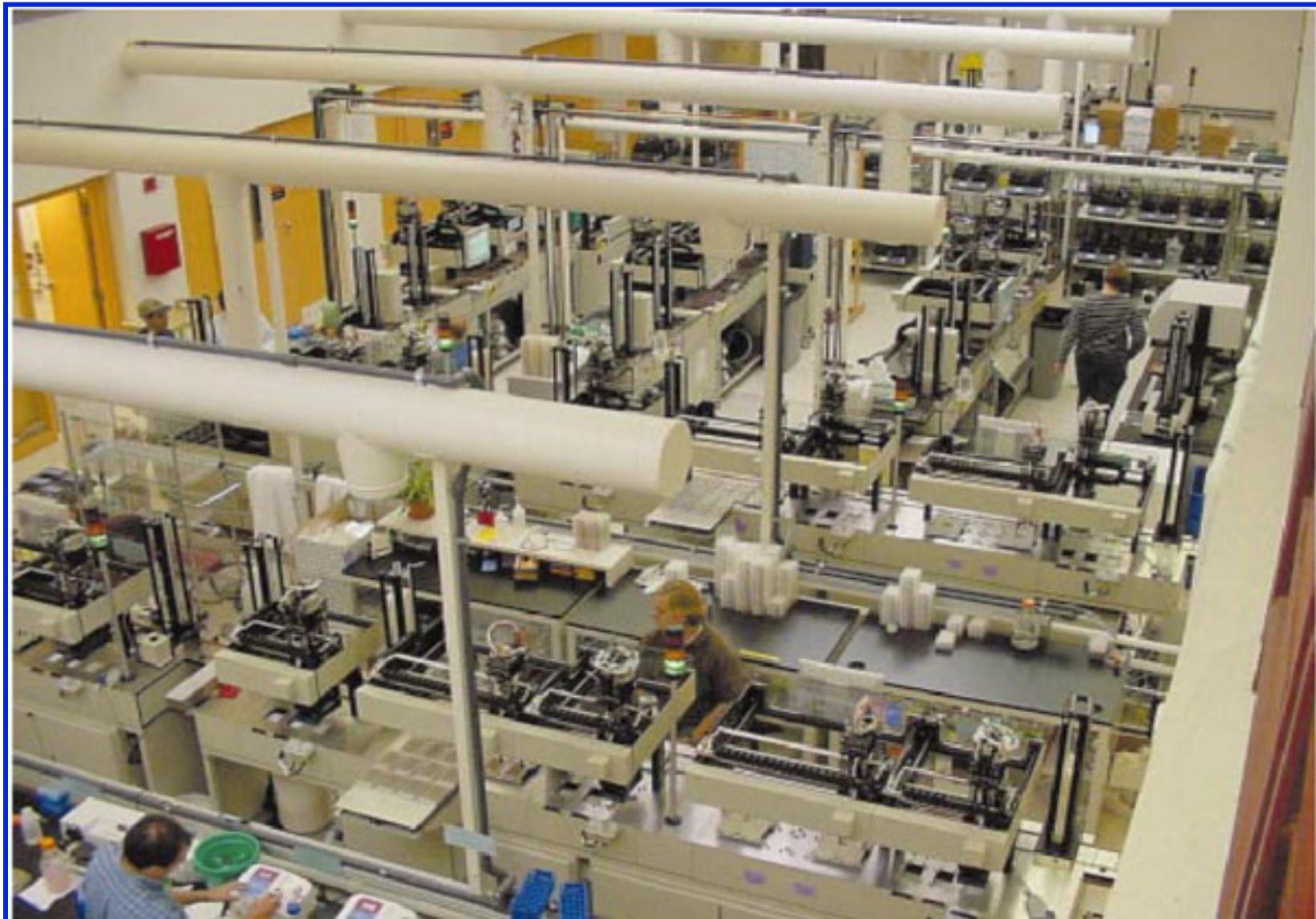
Use over time for: bioinformatics





© 2014 Wendy Lee

The New Face of Biology



© 2014 Wendy Lee

Sequencing SARS



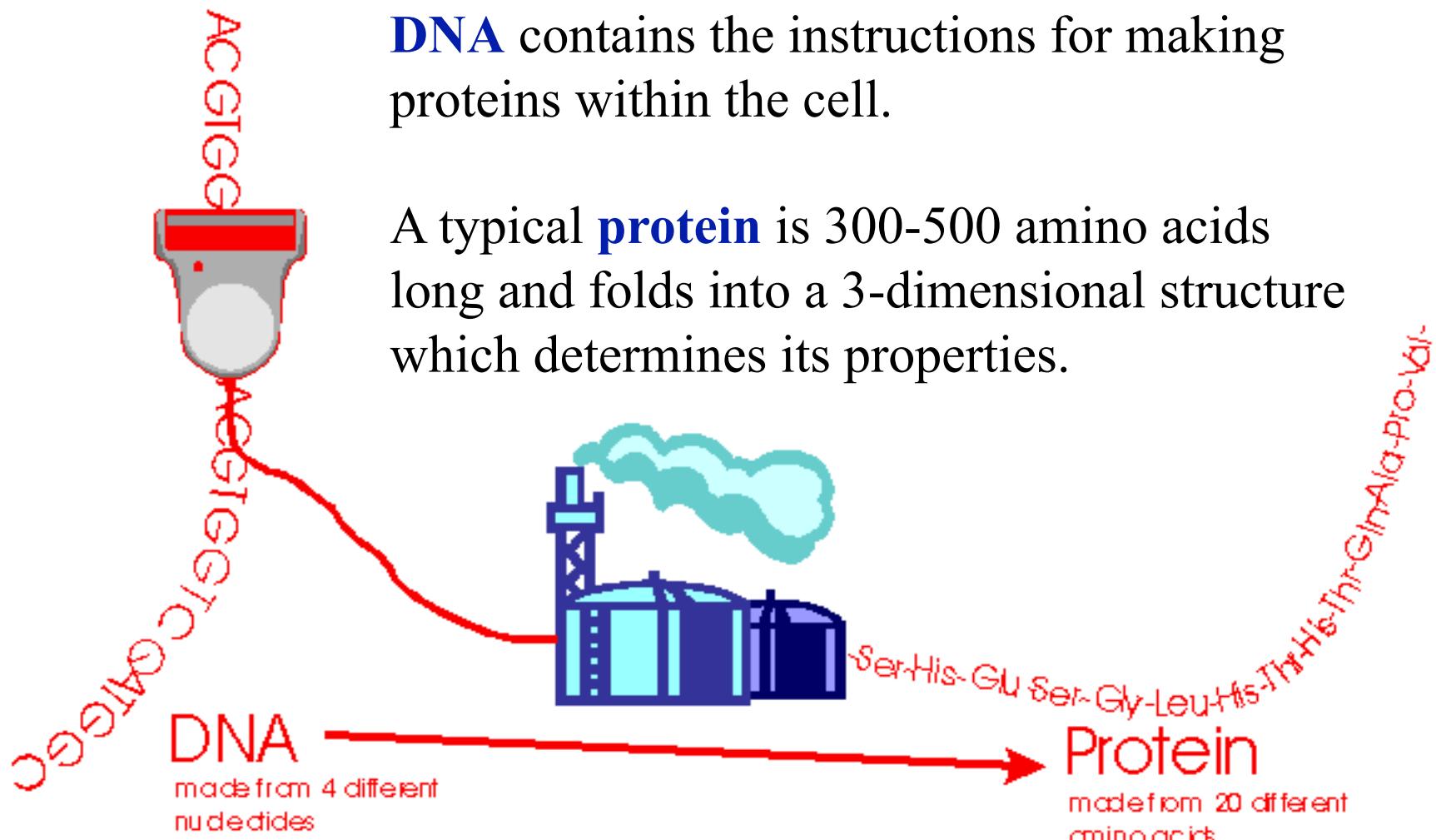
Some Biological Questions (I)

- Why don't all tumors of the same type respond to the same chemotherapies?
 - Could we individualize treatment to increase the chances of successful therapy?
- What are the causes of the hundreds of genetic diseases?
 - How do genes influence complex diseases such as diabetes, obesity, heart disease, cancer, and alcoholism?
 - What can we do about these diseases?

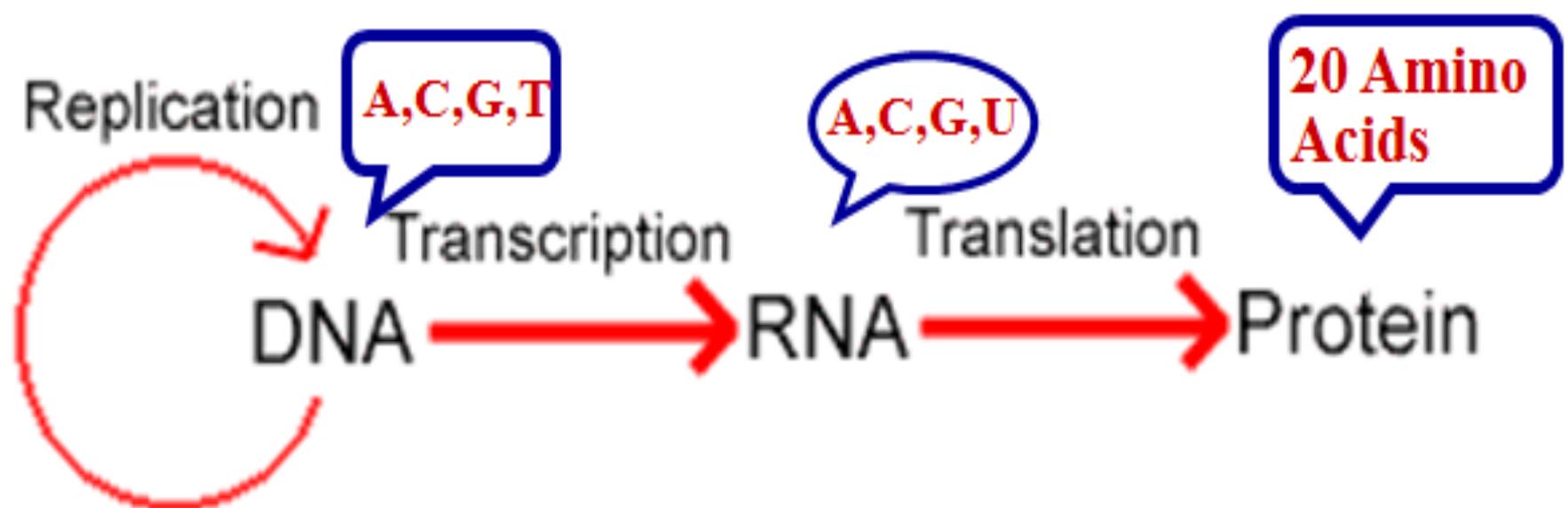
Some Biological Questions (III)

- How do bacteria and viruses cause disease?
 - How can a better understanding of their physiology improve prevention and treatment?
- Can DNA evidence be relied upon in convicting (or exonerating) those accused of crimes?
- What is the risk that a child will inherit a genetic disease or the susceptibility for the disease?

Protein Factory



DNA → RNA → Protein

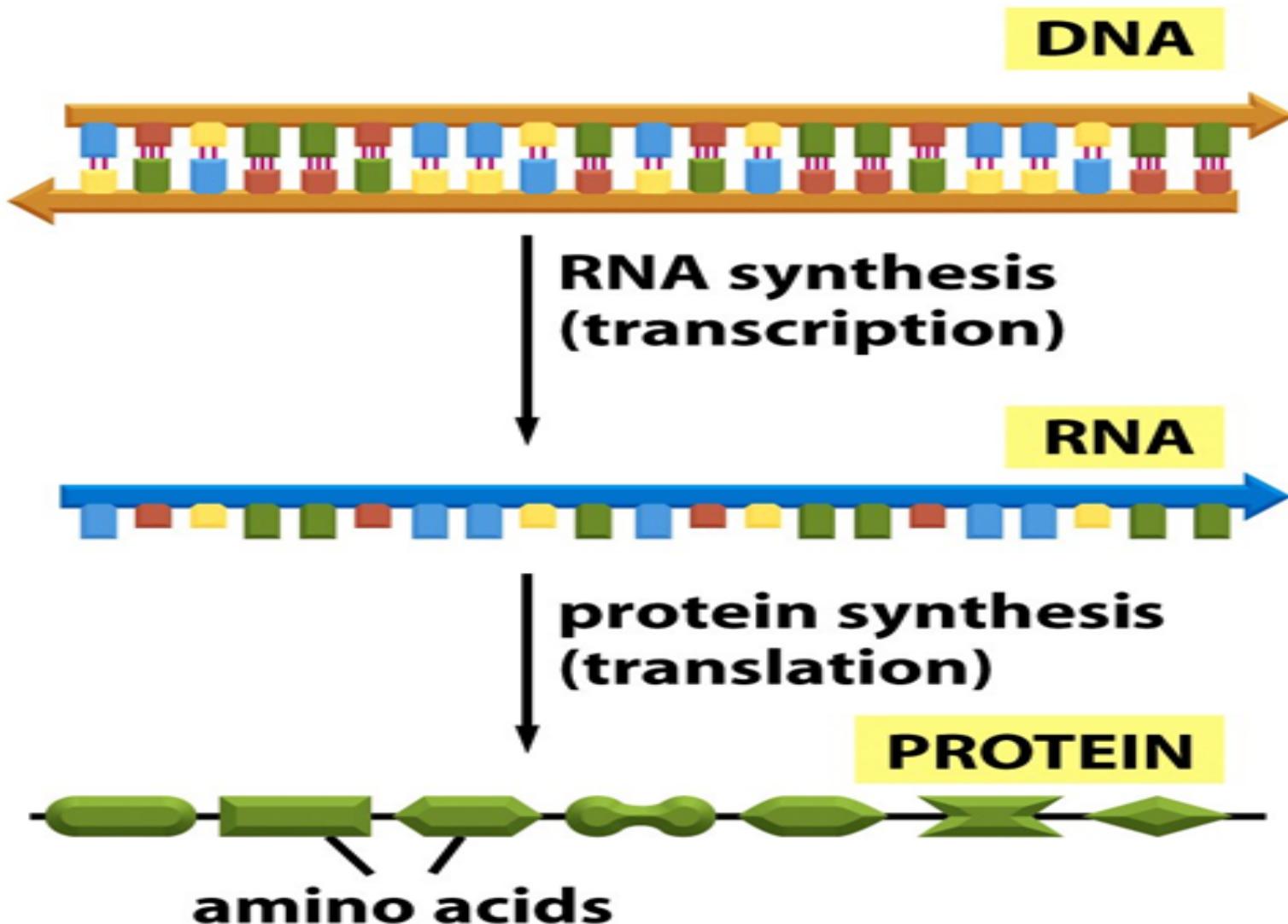


Adenine (A)
Guanine (G)
Cytosine (C)
Thymine (T)

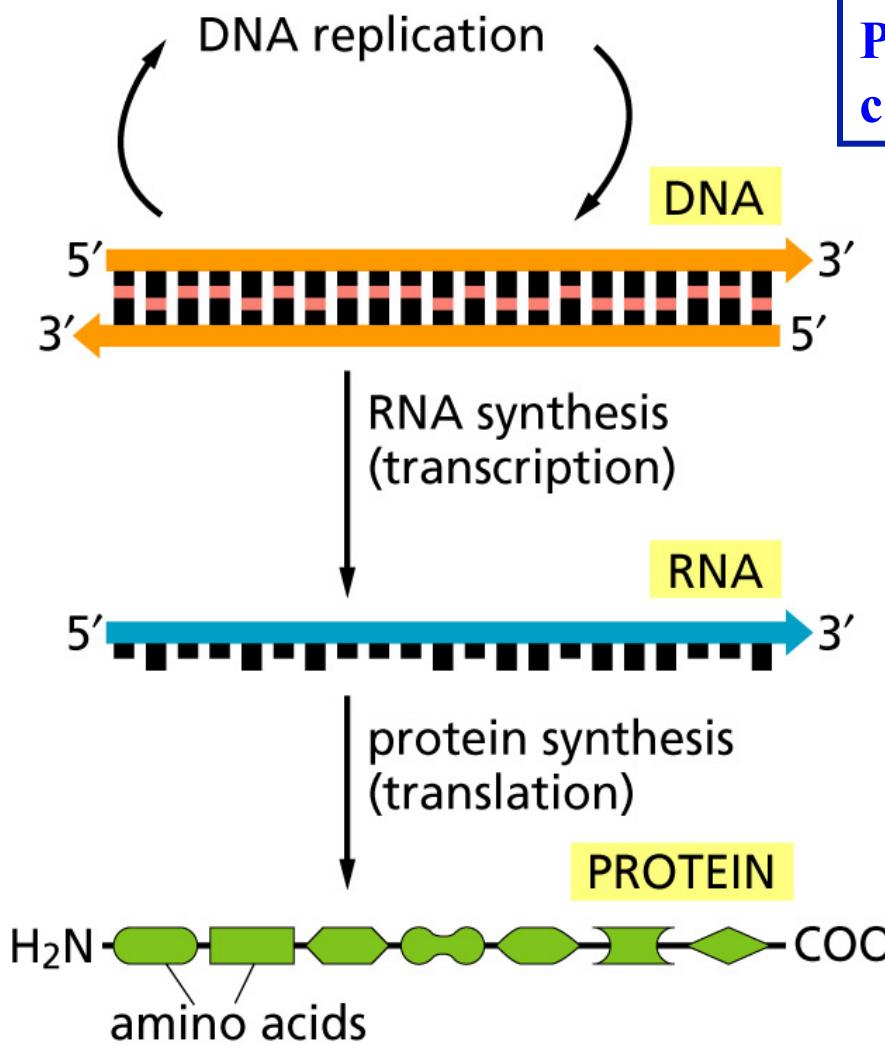
Adenine (A)
Guanine (G)
Cytosine (C)
Uracil (U)

{A-Y} – {BJOUX}

From DNA to Protein



Central Dogma of Molecular Biology



Proteins, are the molecules that put the cell's genetic information into action.

Traits

Diseases

Drug Resistance

Physiology

Metabolism

“We are our Proteins” Doolittle



Source: George Poste

Limitless Diversity From Combinatorial Assemblies of Limited Building Blocks

© 2014 Wendy Lee

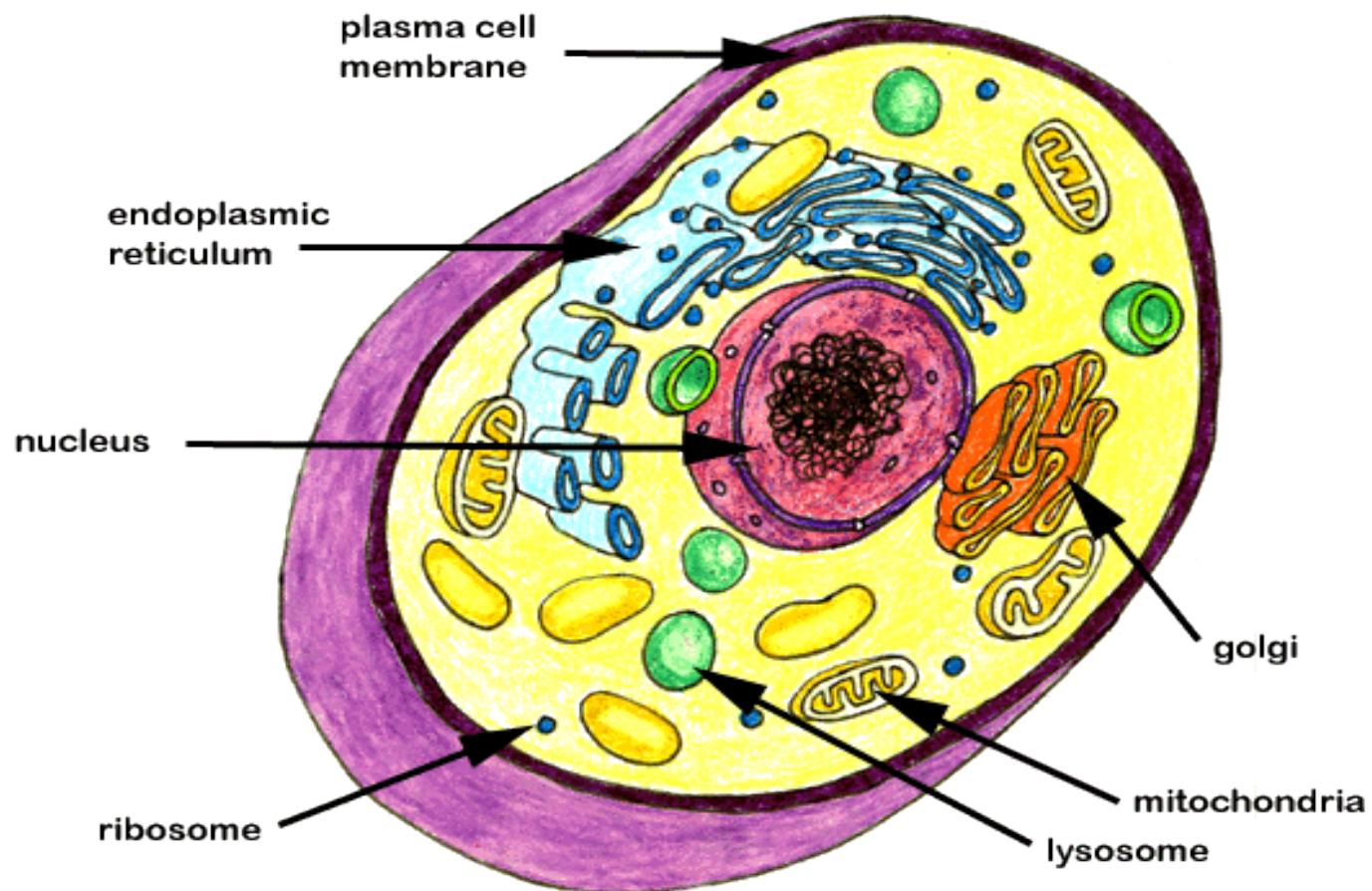
Prokaryotes and Eukaryotes

A **cell** is the fundamental working unit of every living organism.

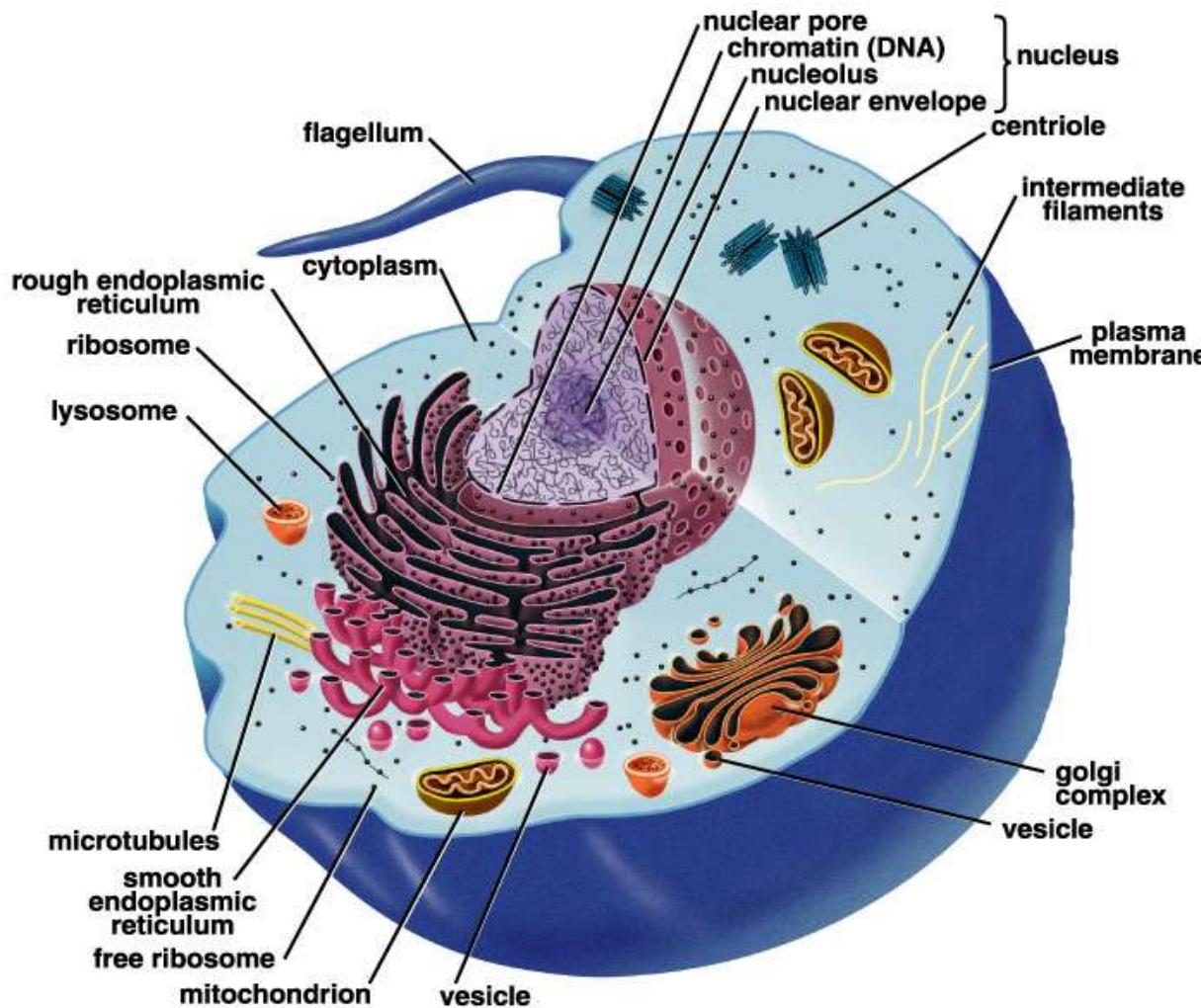
There are two kinds of cells:

- **prokaryotes**, which are single-celled organisms with **no cell nucleus**: archea and bacteria.
- **eukaryotes**, which are higher level organisms, and their cells have **nuclei**: animals and plants.

Eukaryotic Cell (I)

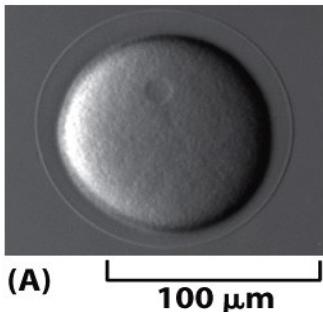


Eukaryotic Cell (II)

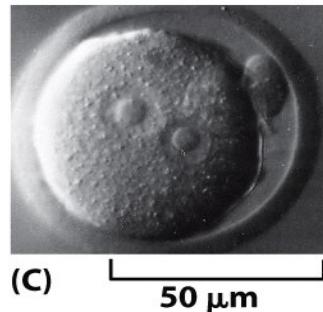


A **cell** carries the entire set of genetic instructions: the genome, that makes an entire organism. The **instructions** are encoded in DNA as genes and packaged as chromosomes in the nucleus.

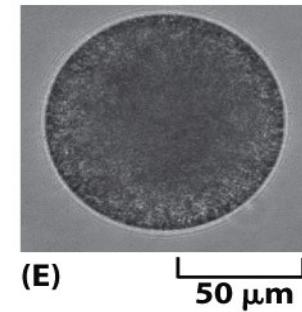
Hereditary Information



Sea Urchin ↓



Mouse ↓



Seaweed ↓



The hereditary information in the fertilized egg cell determines the nature of the whole multicellular organism.

Proteins and Nucleic Acids

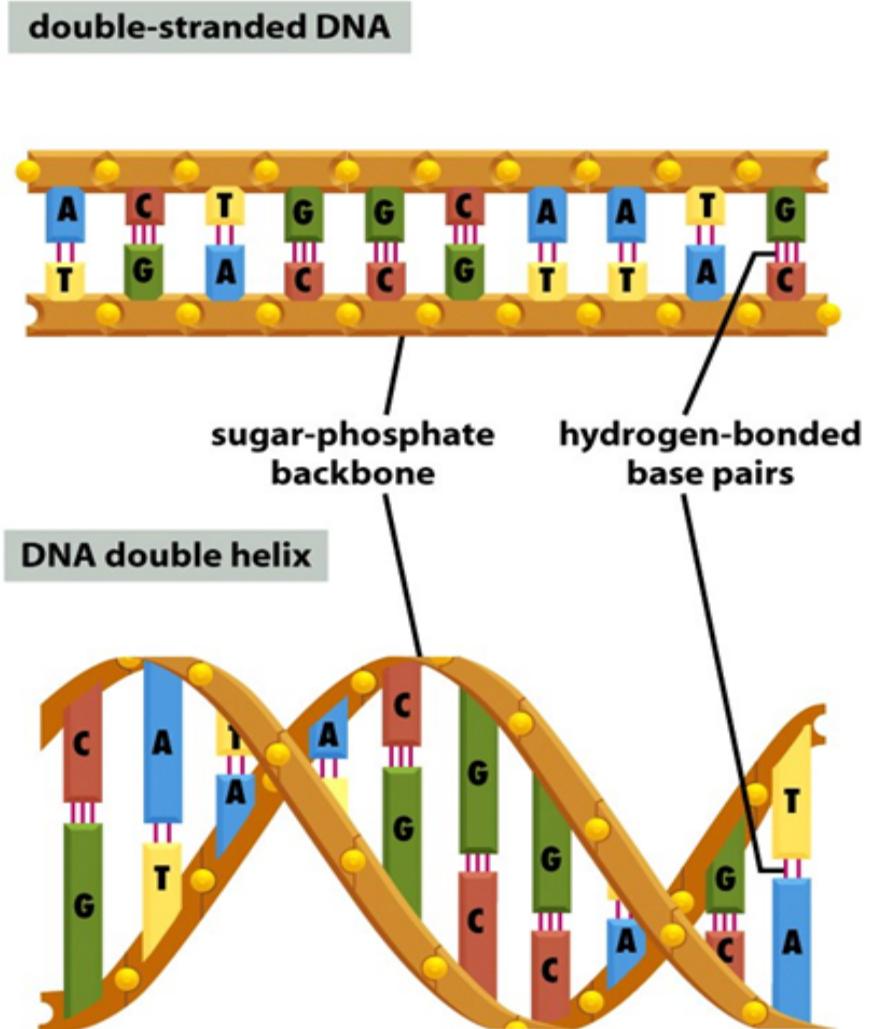
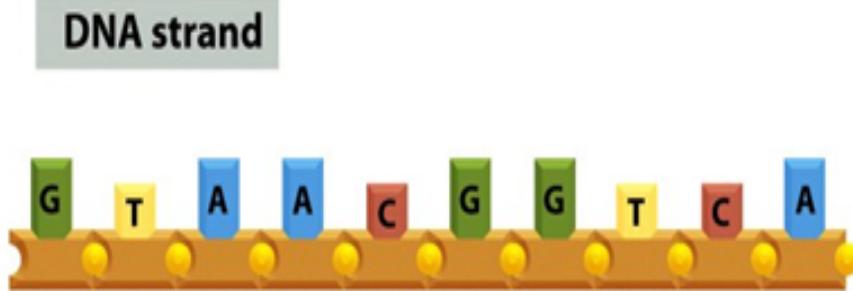
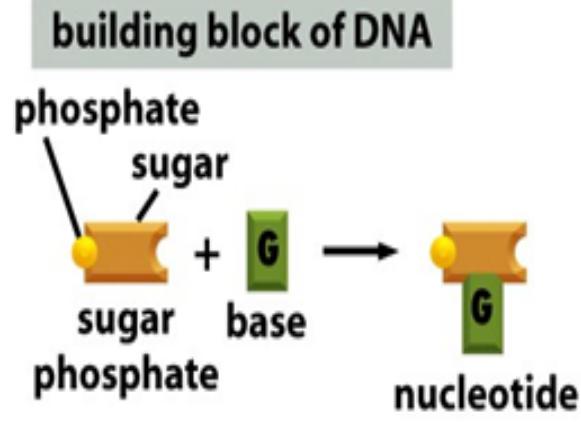
All living organisms have a similar molecular chemistry.
The main actors in the chemistry of life are molecules:

- **proteins**: which are responsible for what a living being is and does in a physical sense.
“We **are** our proteins” R. Doolittle.
- **nucleic acids**: which encode the information necessary to produce proteins and are responsible for passing the “recipe” to subsequent generations.

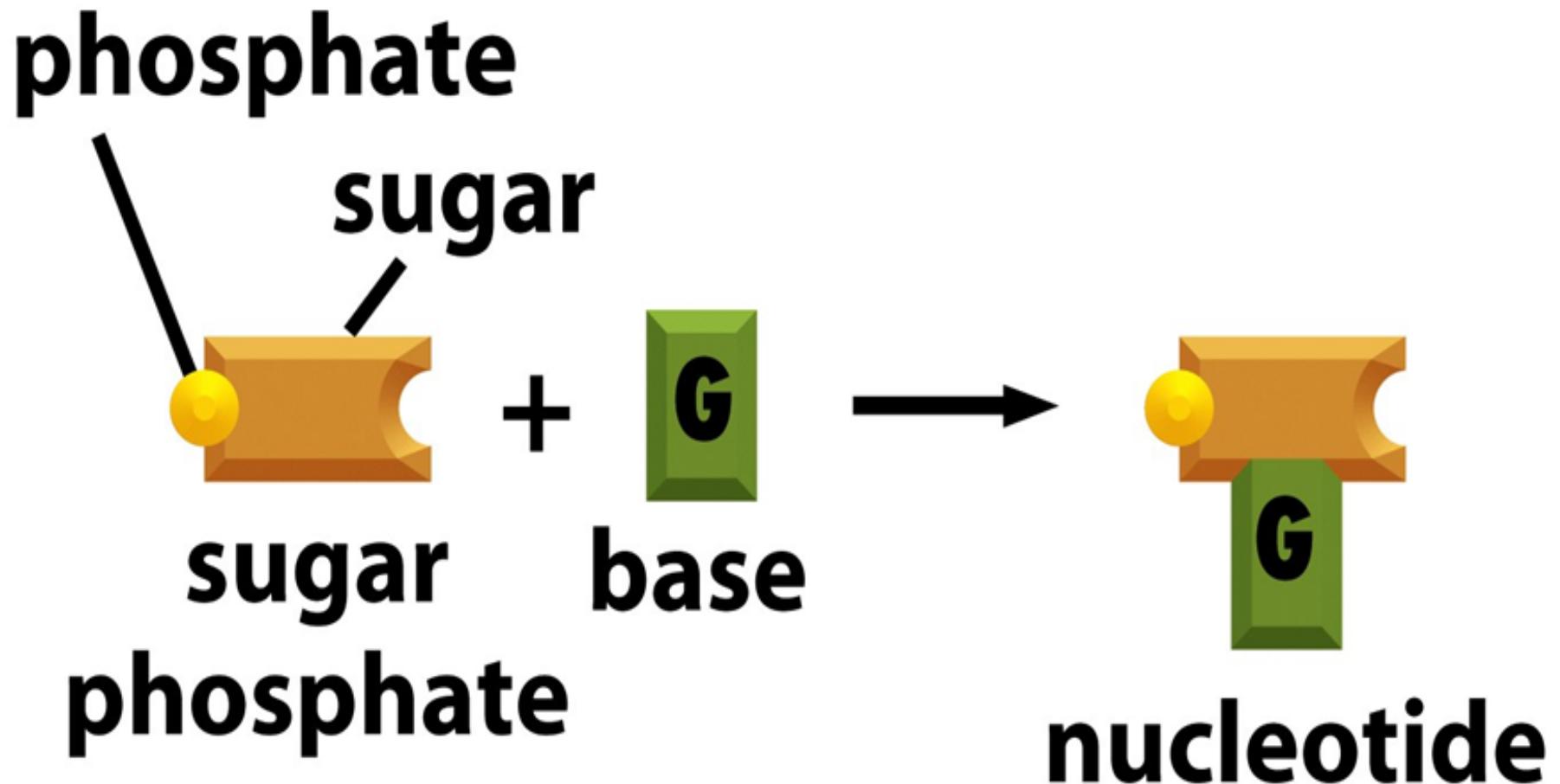
Living organisms contain 2 kinds of nucleic acids:

- **Ribonucleic acid (RNA)**
- **Deoxyribonucleic acid (DNA)**

DNA Double Helix

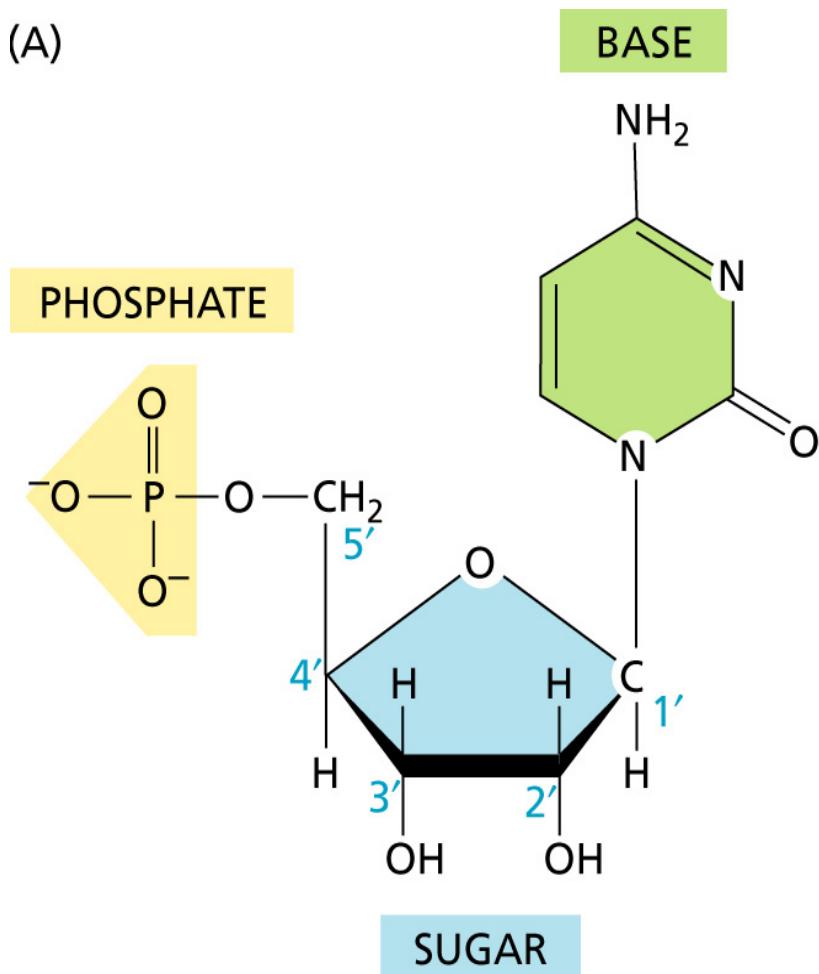


Building Block of DNA

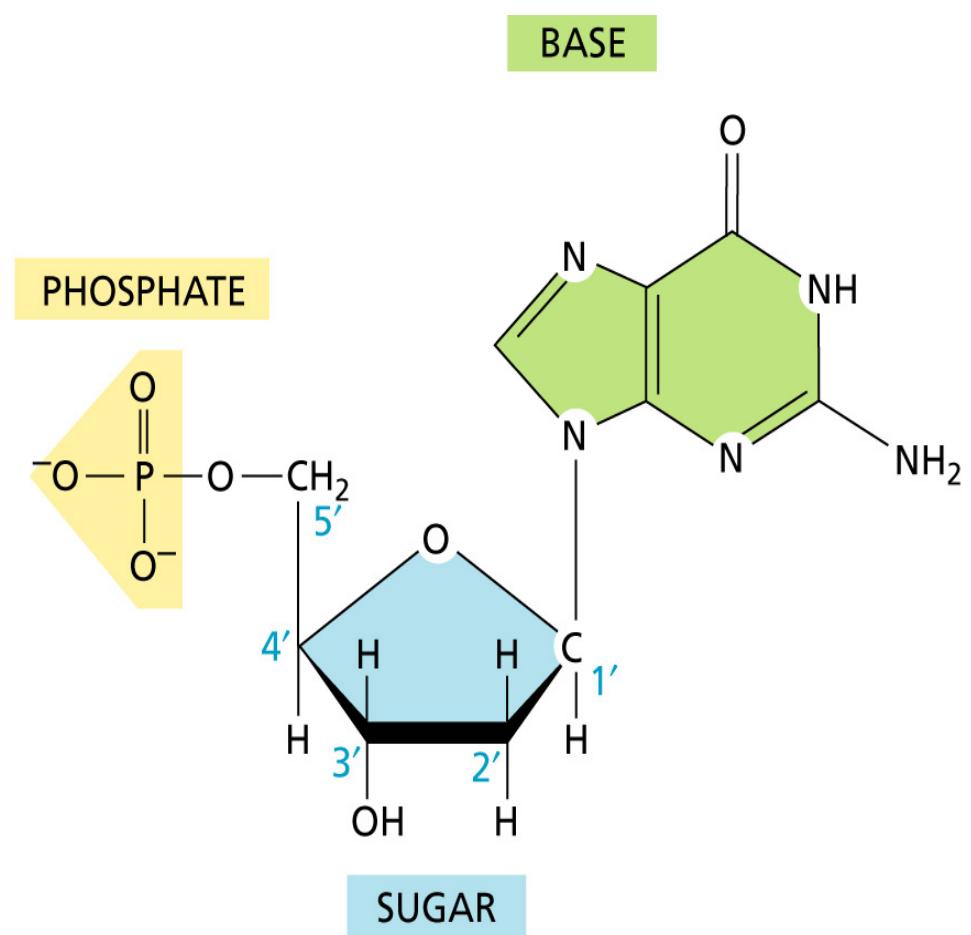


Sugars Found in Nucleic Acids

(A)

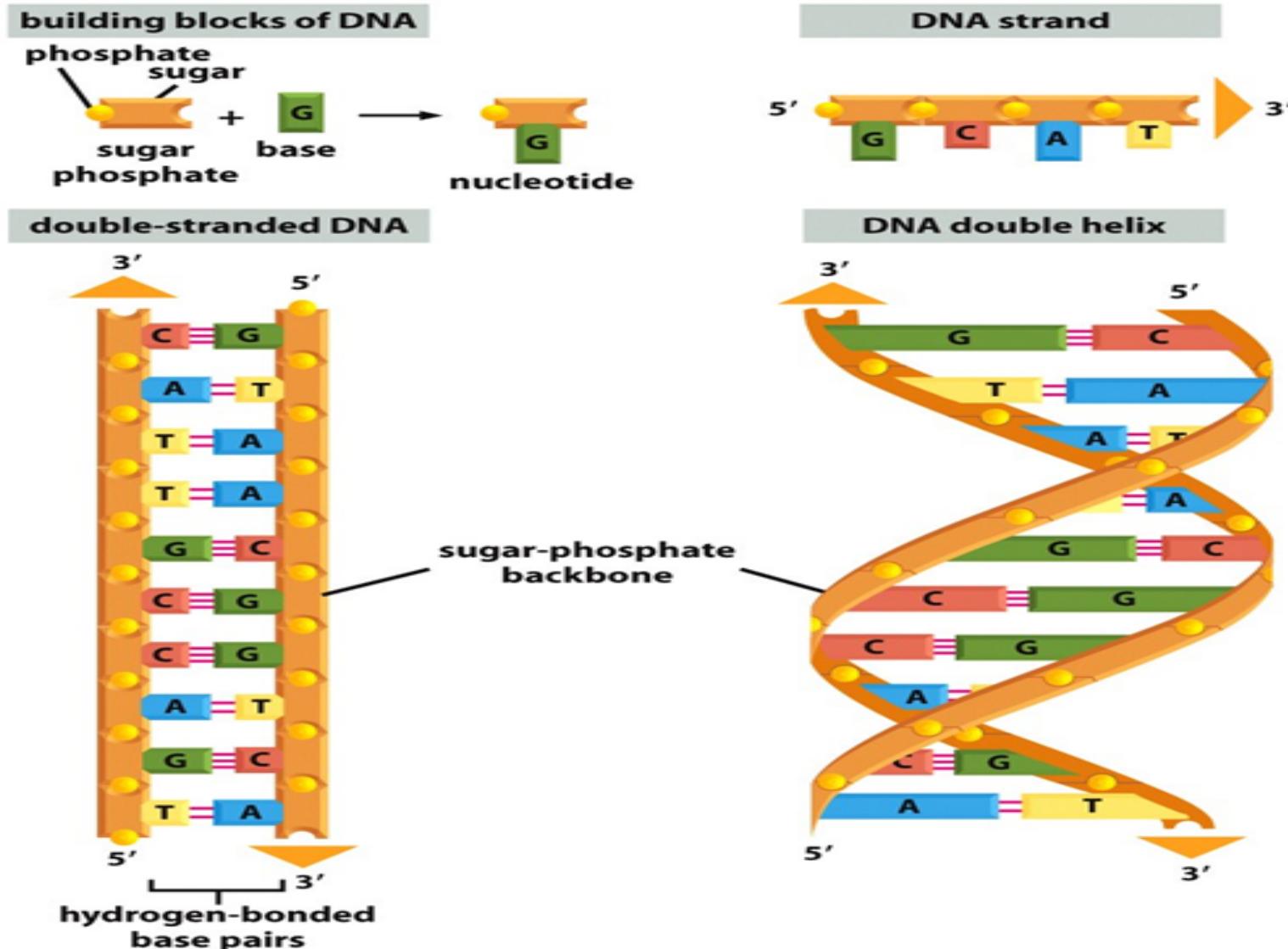


Pentose sugar present in **RNA**



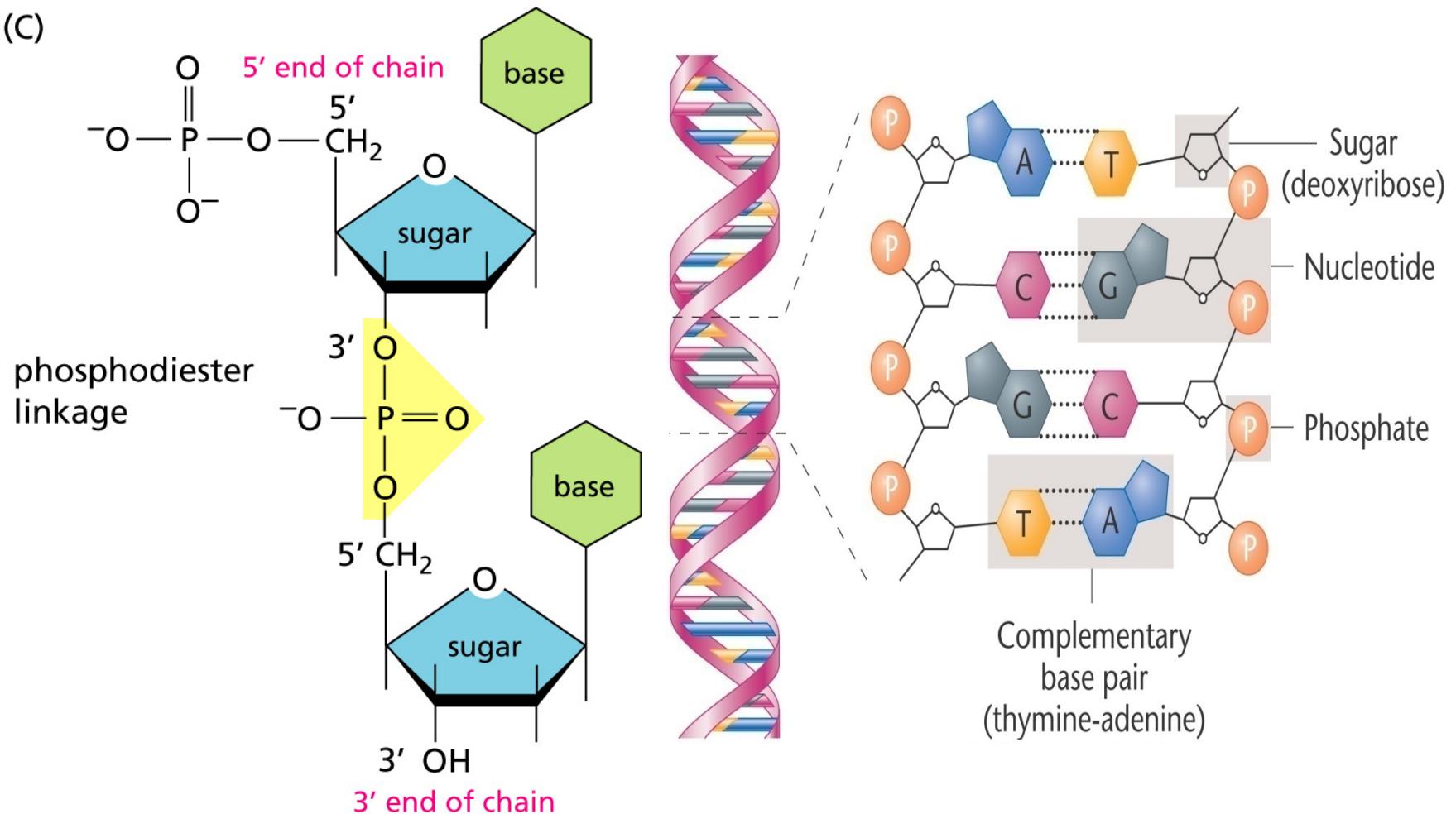
Pentose sugar present in **DNA**

Double-Stranded DNA

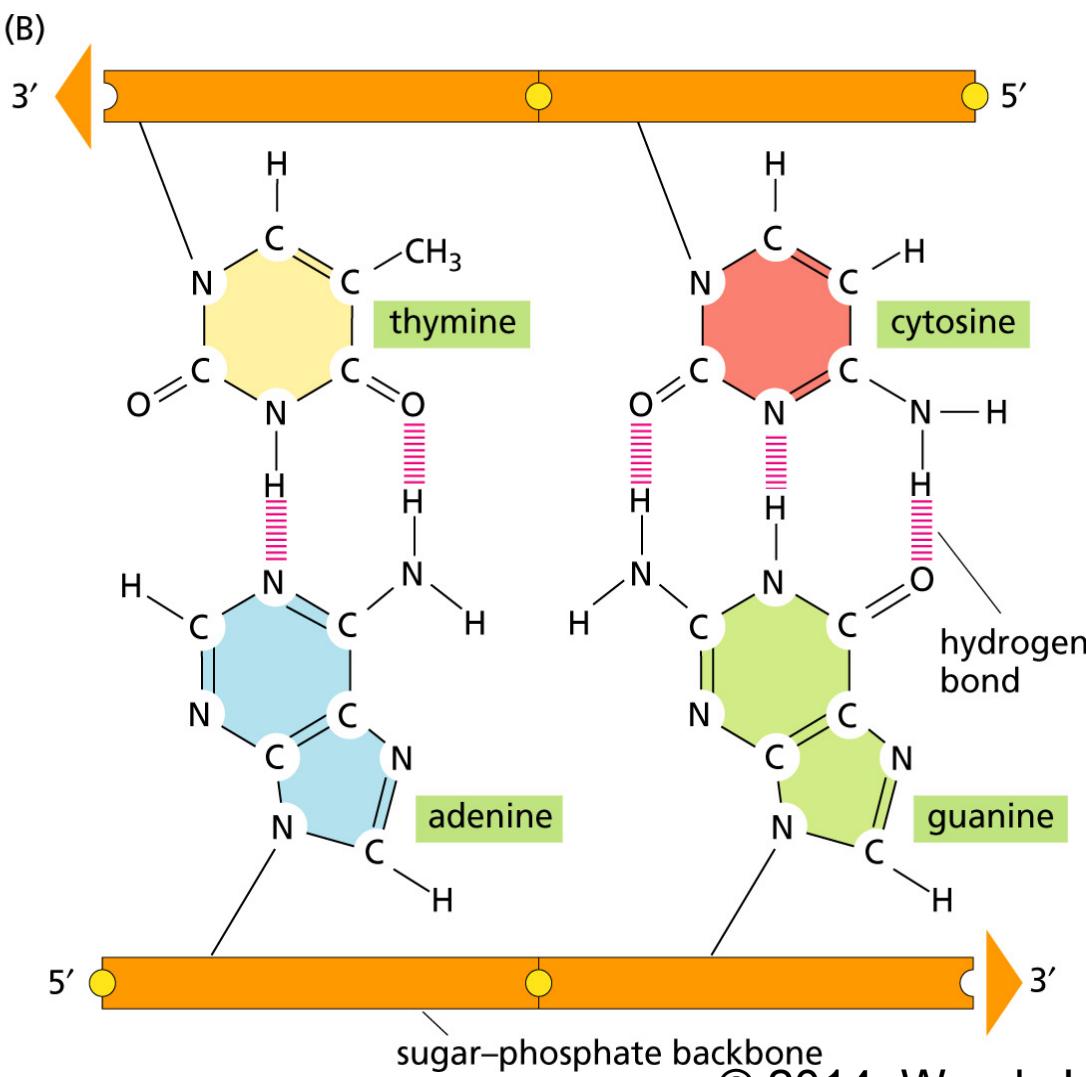
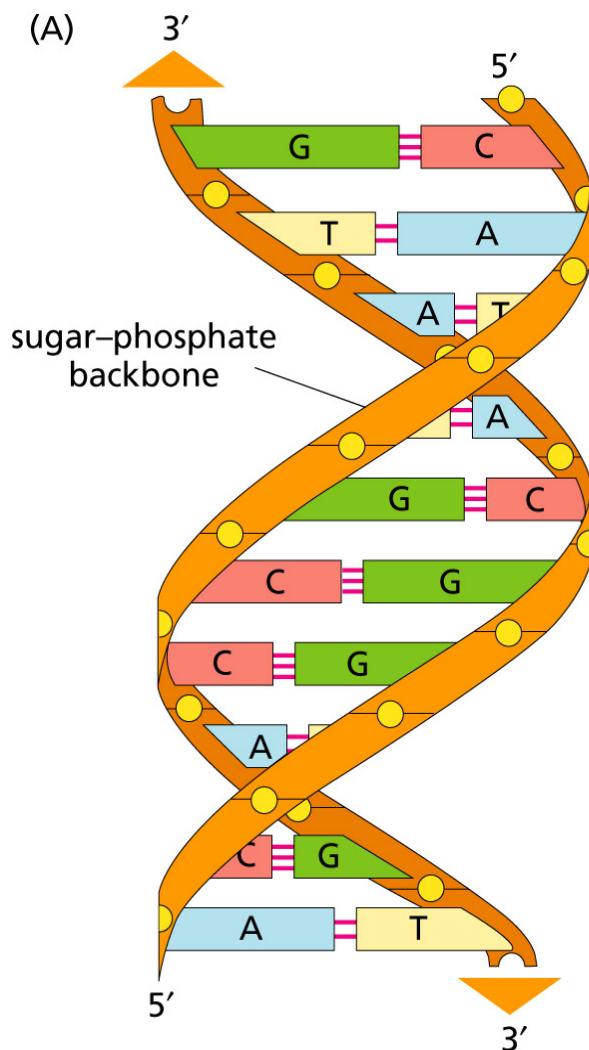


DNA Phosphodiester Backbone

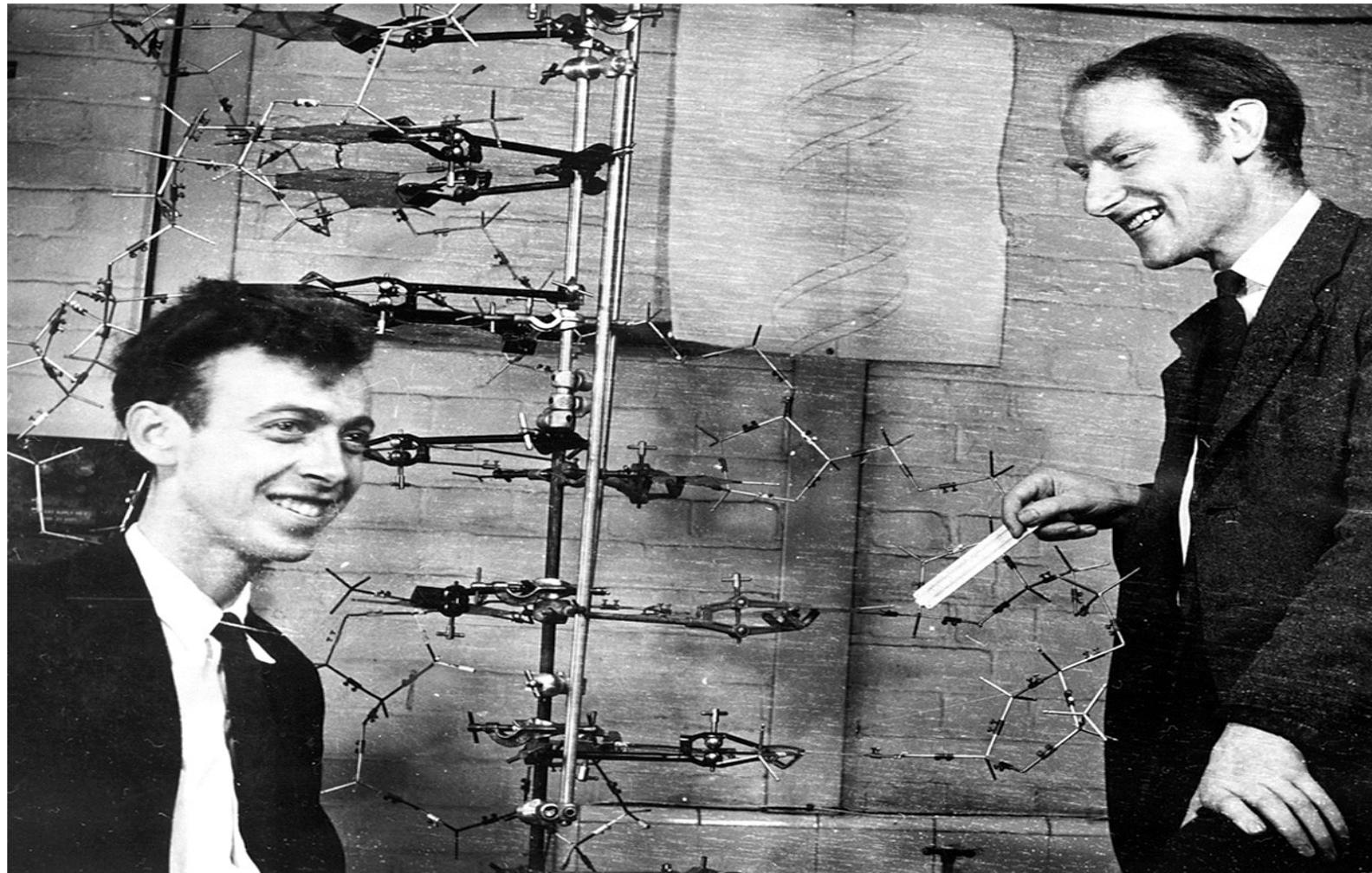
(C)



Double Helical Structure of DNA



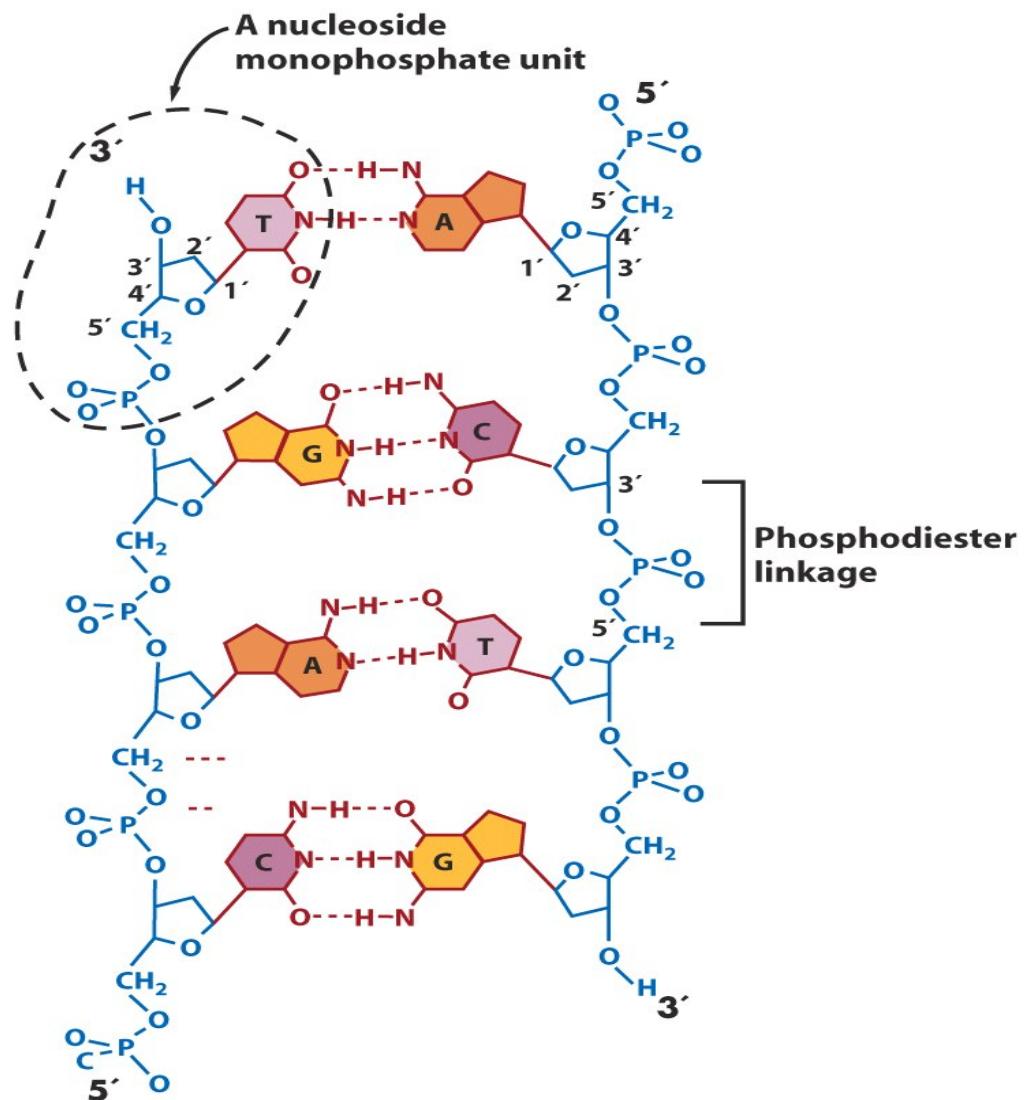
J. Watson and F. Crick



© 2014 Wendy Lee

The Two Backbones of DNA

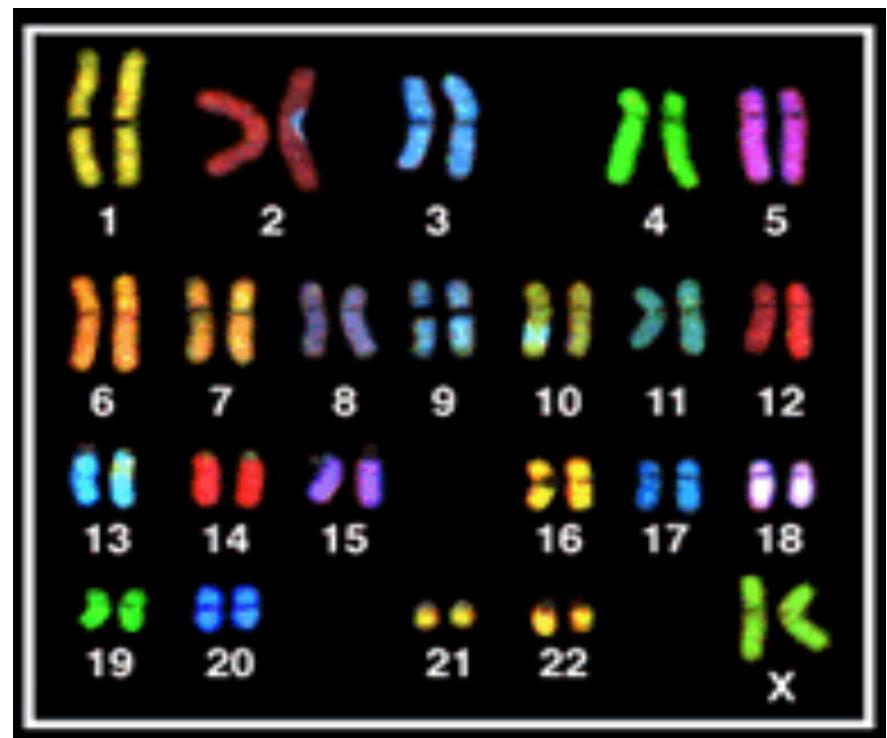
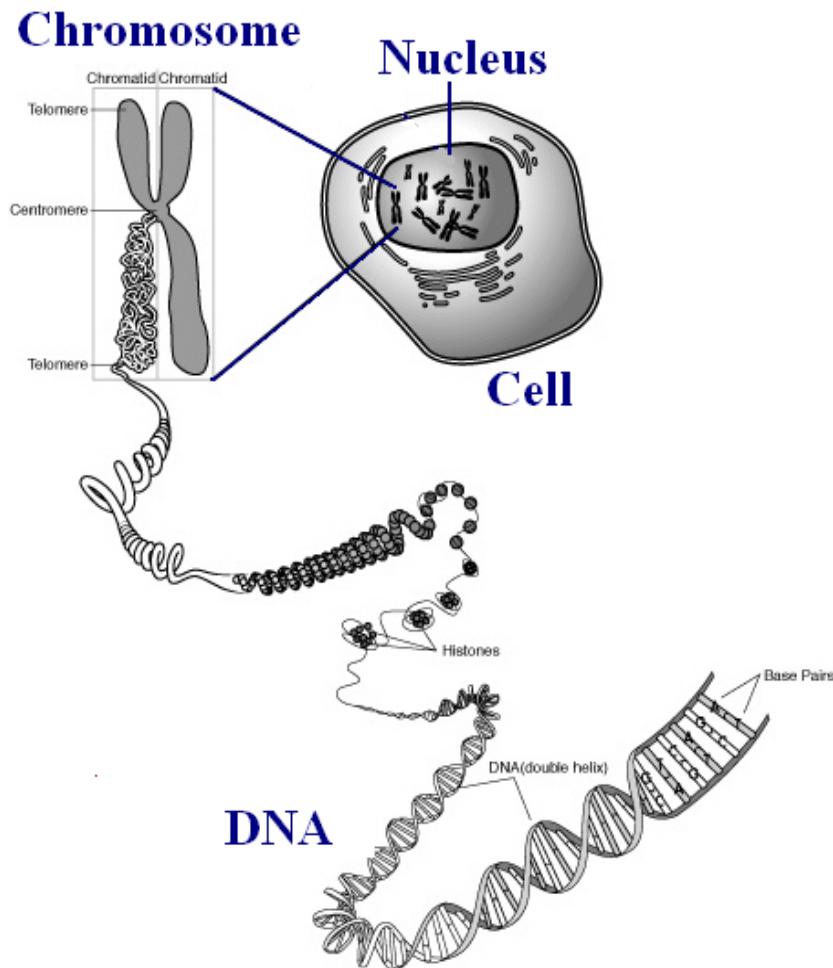
The DNA backbones have alternating sugar-phosphate components. The backbones run in opposite directions.



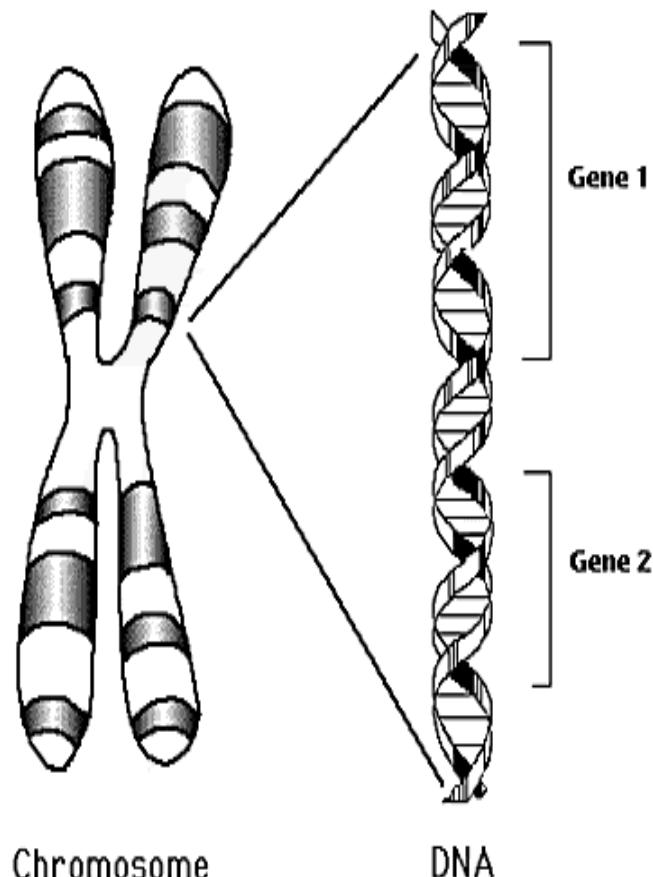
DNA and Chromosomes

- The **genome** is a complete set of instructions for making an organism, consists of tightly coiled threads of **DNA** organized into structures called **chromosomes**.
- Besides the reproductive cell and red blood cell, every single **cell** in the human body contains the **human genome**.

Chromosomes and Genome

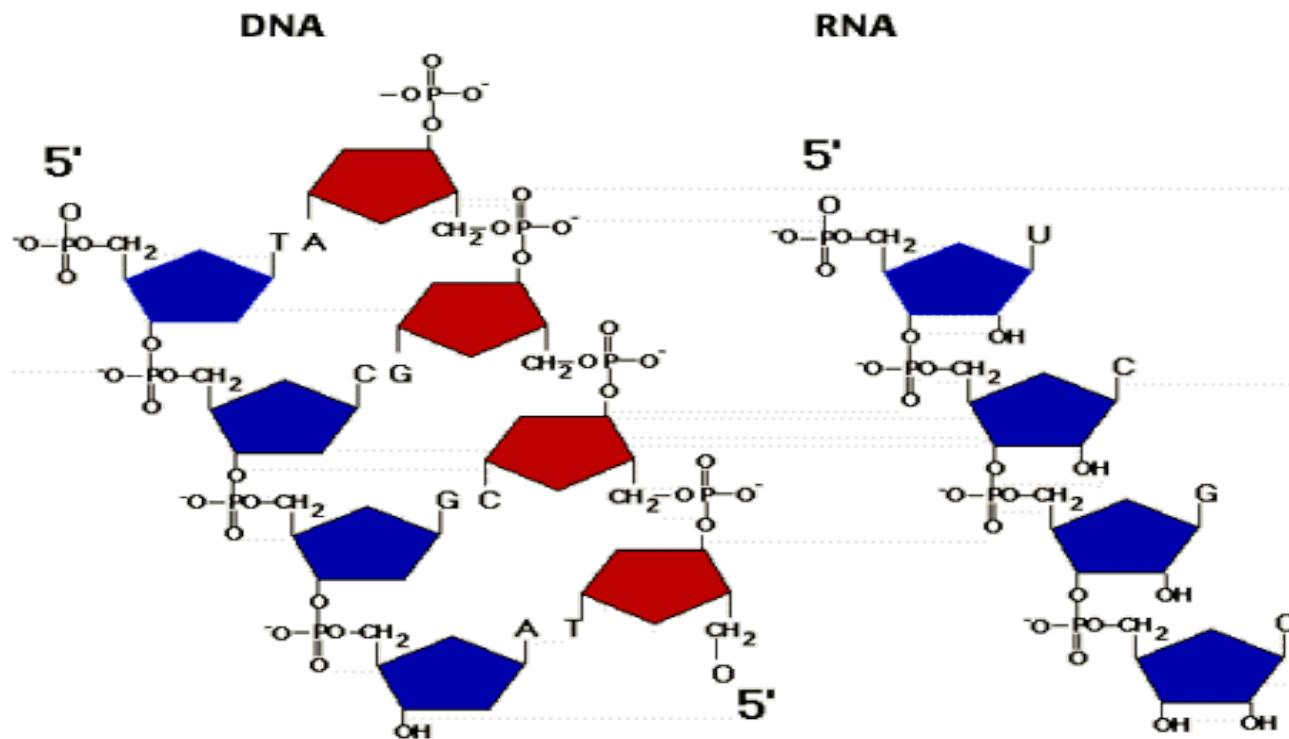


Genes



- A **gene** is a specific sequence of nucleotide bases along a chromosome carrying information for constructing a protein. A gene encodes a protein (or an RNA).
- The distance between **genes** is often much larger than the genes themselves.
- The human genome has around 23,500 genes.

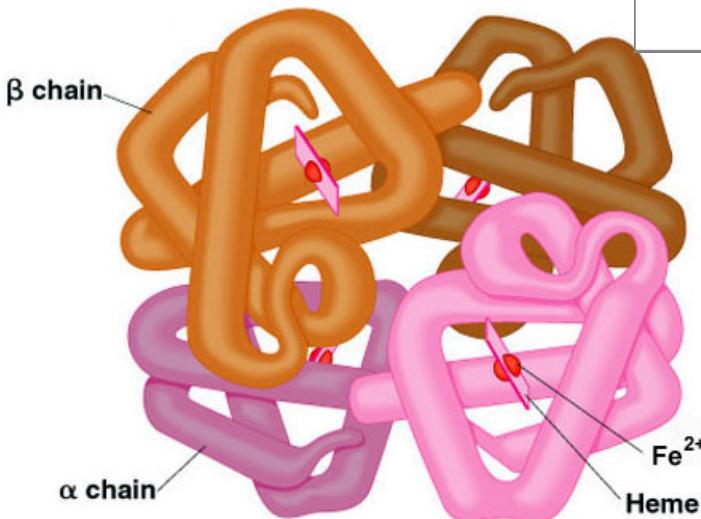
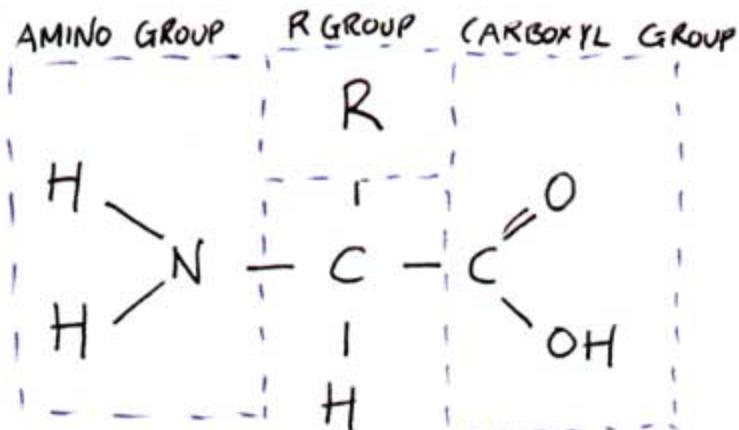
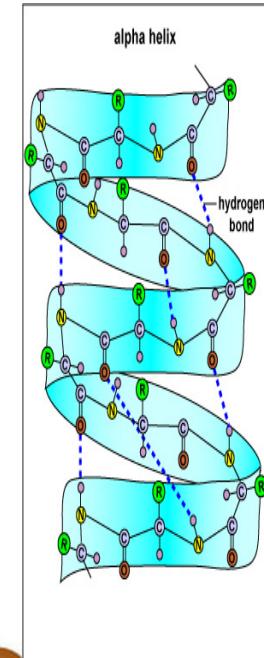
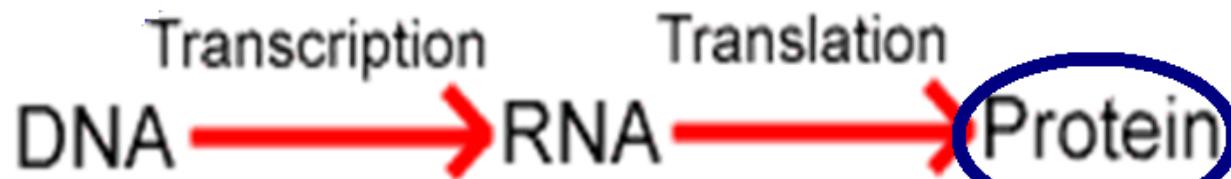
RNA



Ribonucleic Acid - RNA

- **RNA** is found in the cell and can also carry genetic information.
- While DNA is located primarily in the nucleus, **RNA** can also be found in the **cytoplasm**.
- **RNA** is built from the nucleotides **cytosine**, **guanine**, **adenine** and **uracil (U)** (instead of thymine).
- **RNA** has its sugar phosphate backbone containing **ribose**.
- **RNA** forms a **single strand**.
- **RNA** molecules tend to have a less-regular three-dimensional structure than DNA.

Proteins

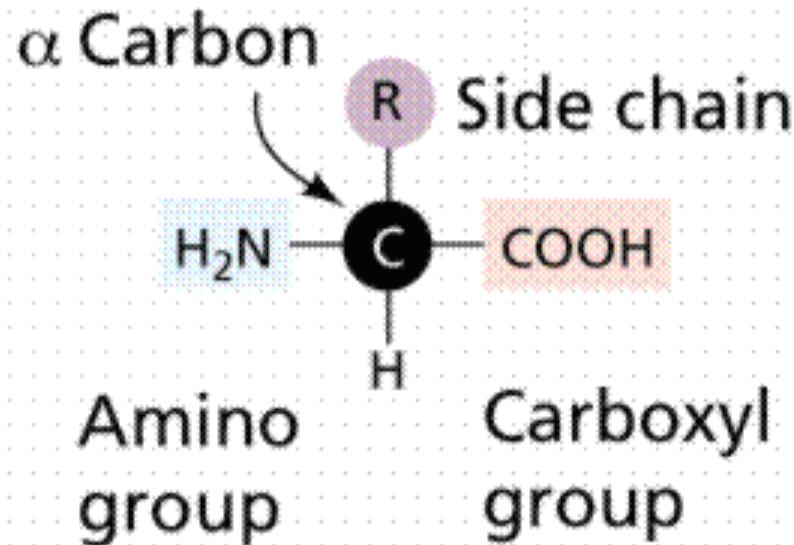


Proteins

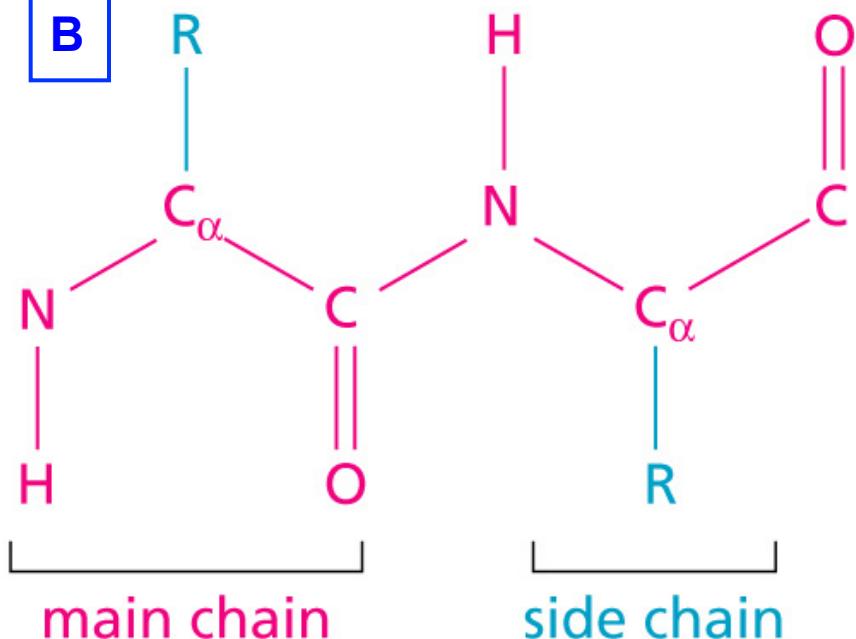
- 20 different **amino acids** are used to synthesize **proteins**.
- The shape and other properties of each **protein** is dictated by the precise sequence of **amino acids** in it.
- The function of a **protein** is determined by its unique three-dimensional structure.

Structure of the Amino Acid

A



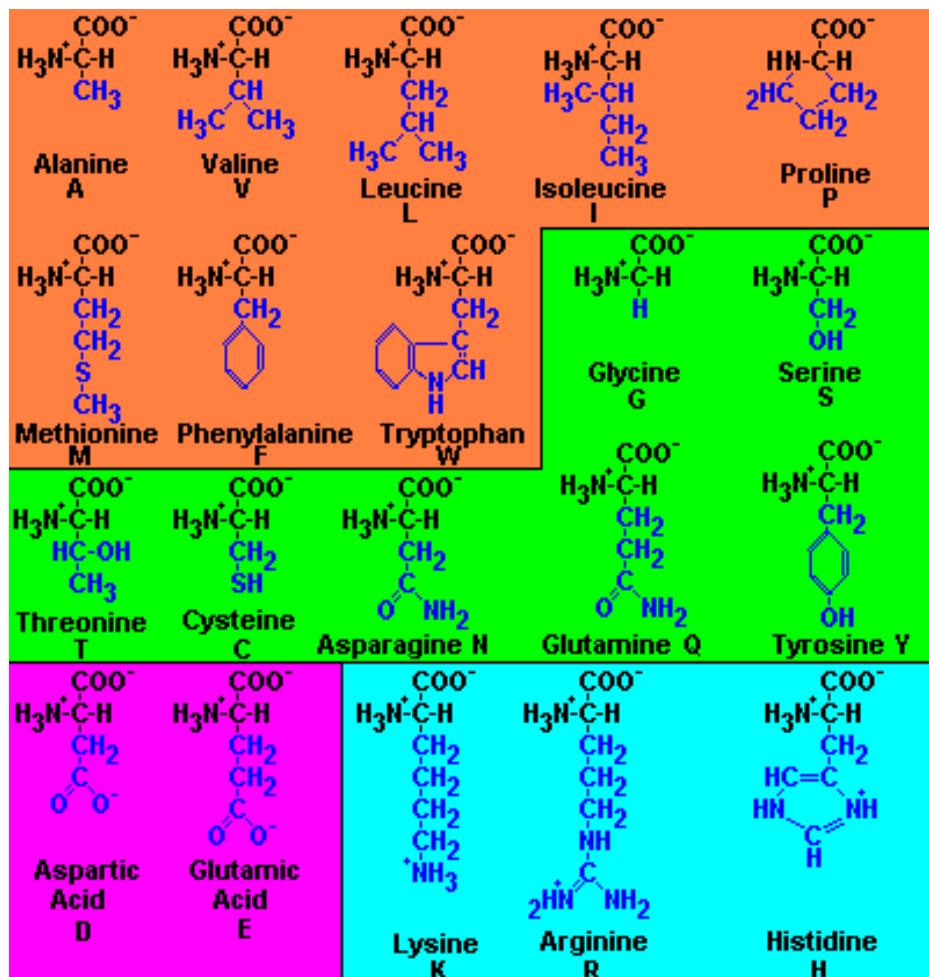
B



A) The functional properties of proteins are almost entirely due to the side chains in the amino acids.

B) Two amino acids:
the main chain is in red and
the side chain in blue.

The Twenty Amino Acids



Orange:
nonpolar and hydrophobic.

The other amino acids are:
polar and hydrophilic - "water loving".

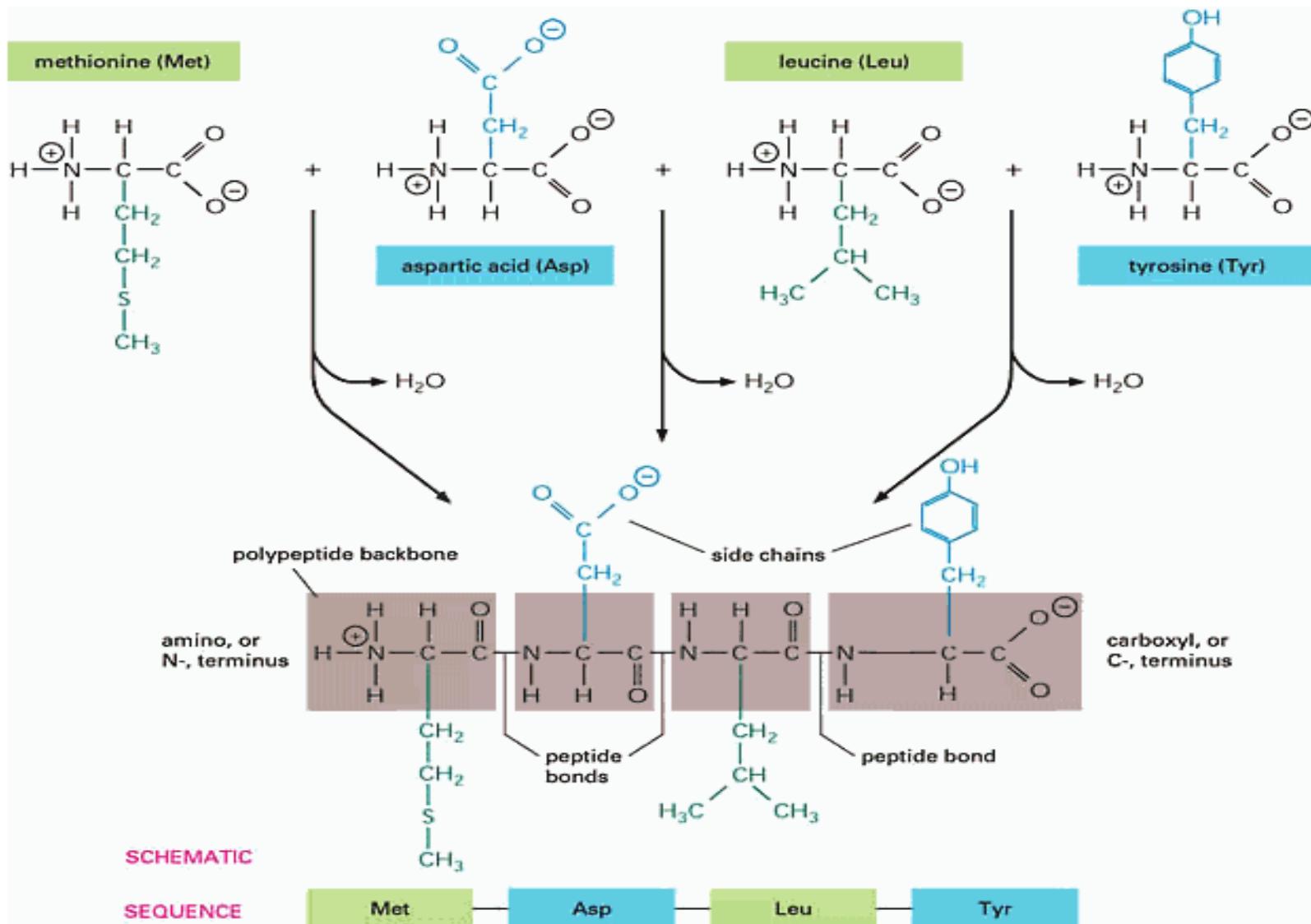
Magenta:
acidic - "carboxy" group in
the side chain.

Light blue:
basic - "amine" group in the
side chain.

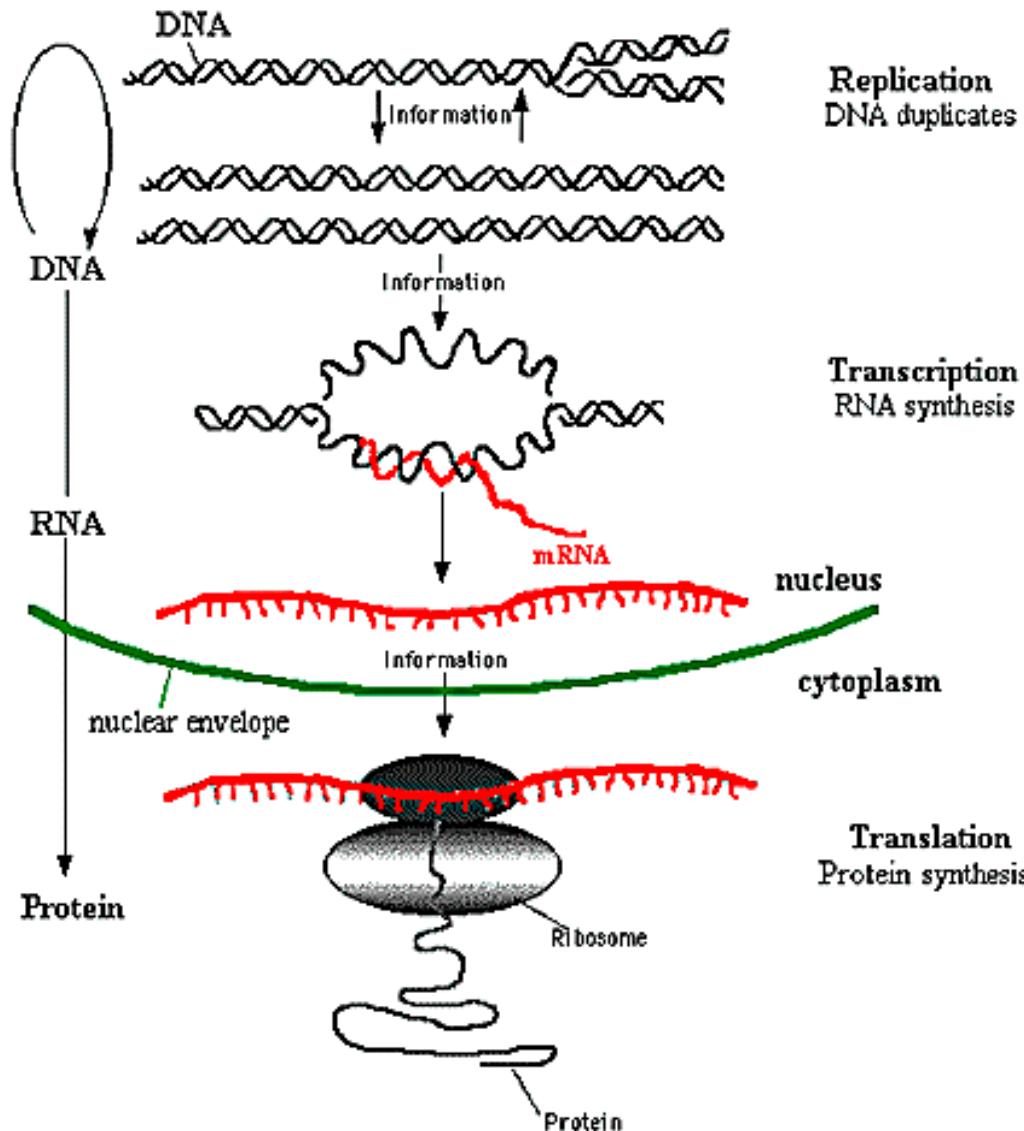
The 20 Amino Acids

1-letter	3-letter	Amino acid	1-letter	3-letter	Amino Acid
A	Ala	Alanine	M	Met	Methionine
C	Cys	Cysteine	N	Asn	Asparagine
D	Asp	Aspartic Acid	P	Pro	Proline
E	Glu	Glutamic Acid	Q	Gln	Glutamine
F	Phe	Phenylalanine	R	Arg	Arginine
G	Gly	Glycine	S	Ser	Serine
H	His	Histidine	T	Thr	Threonine
I	Ile	Isoleucine	V	Val	Valine
K	Lys	Lysine	W	Trp	Tryptophan
L	Leu	Leucine	Y	Tyr	Tyrosine

Protein Structure



Steps of the Central Dogma



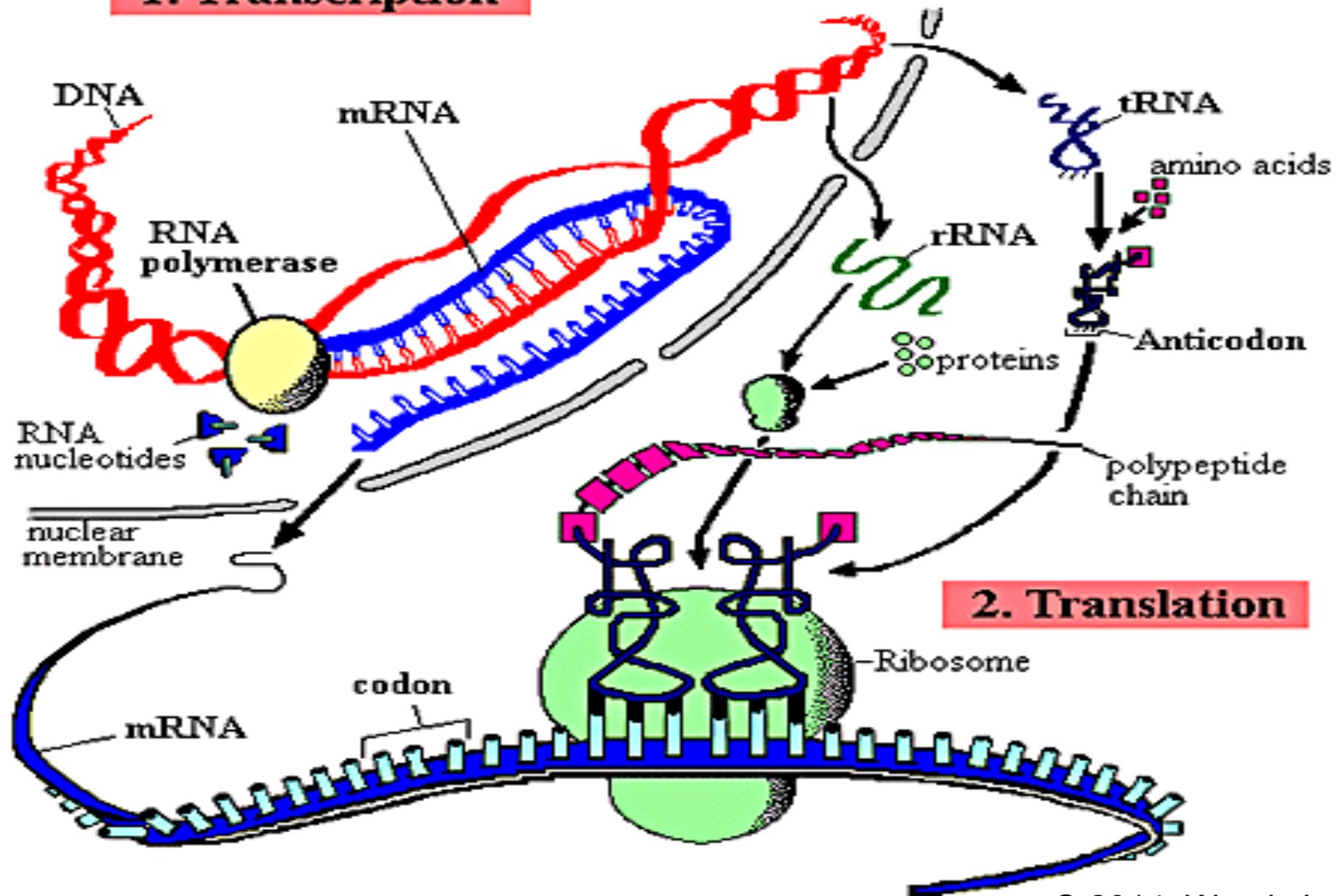
Genetic information embodied in DNA is **replicated** into more DNA

The synthesis of an RNA from a sequence of DNA.
The resulting RNA is **mRNA**.

In eukaryotic cells, the mRNA is **spliced** and it migrates from the nucleus to the cytoplasm.

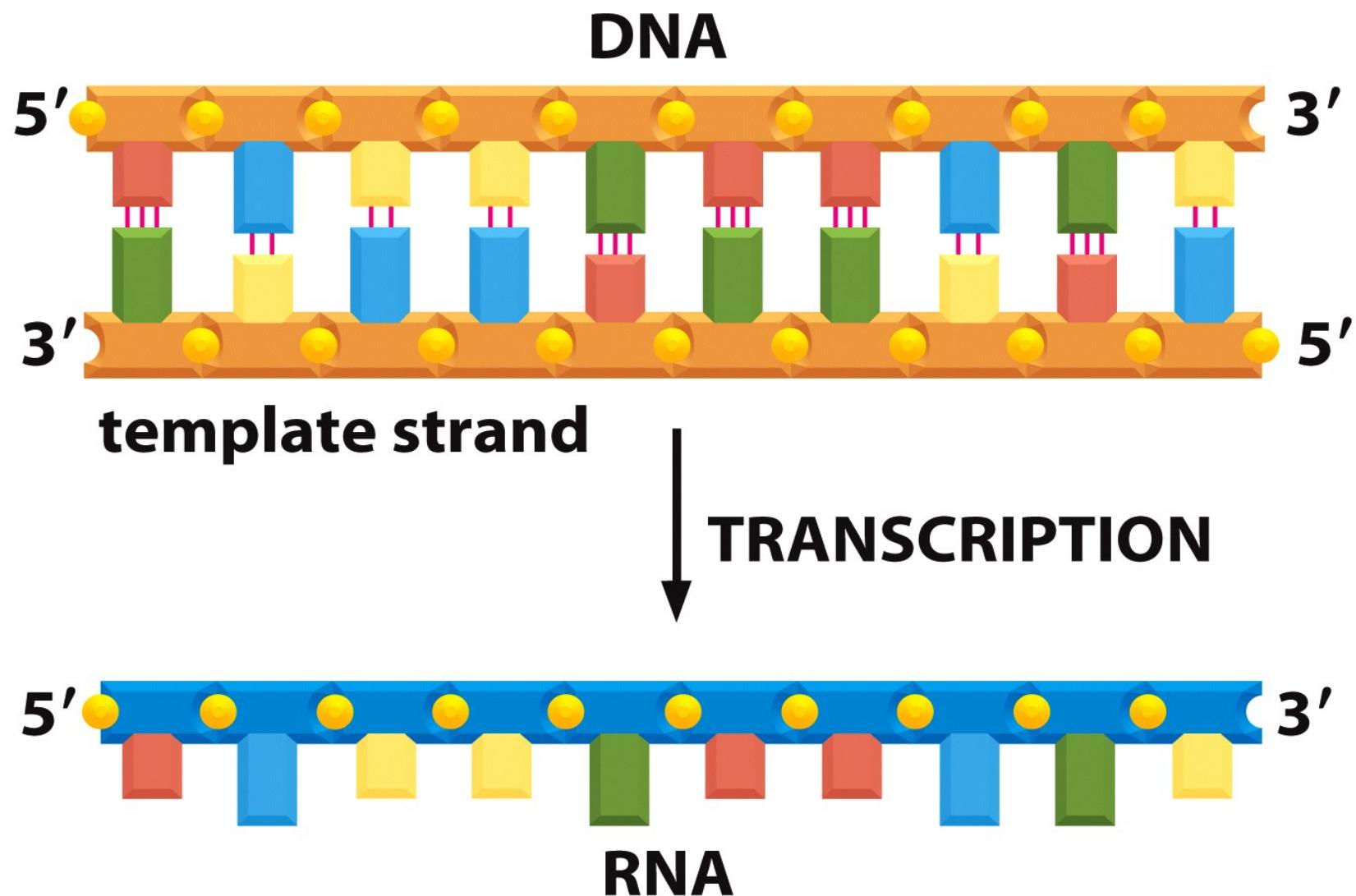
Messenger RNA carries coded information to ribosomes that "read" and use it for **protein synthesis**.

1. Transcription

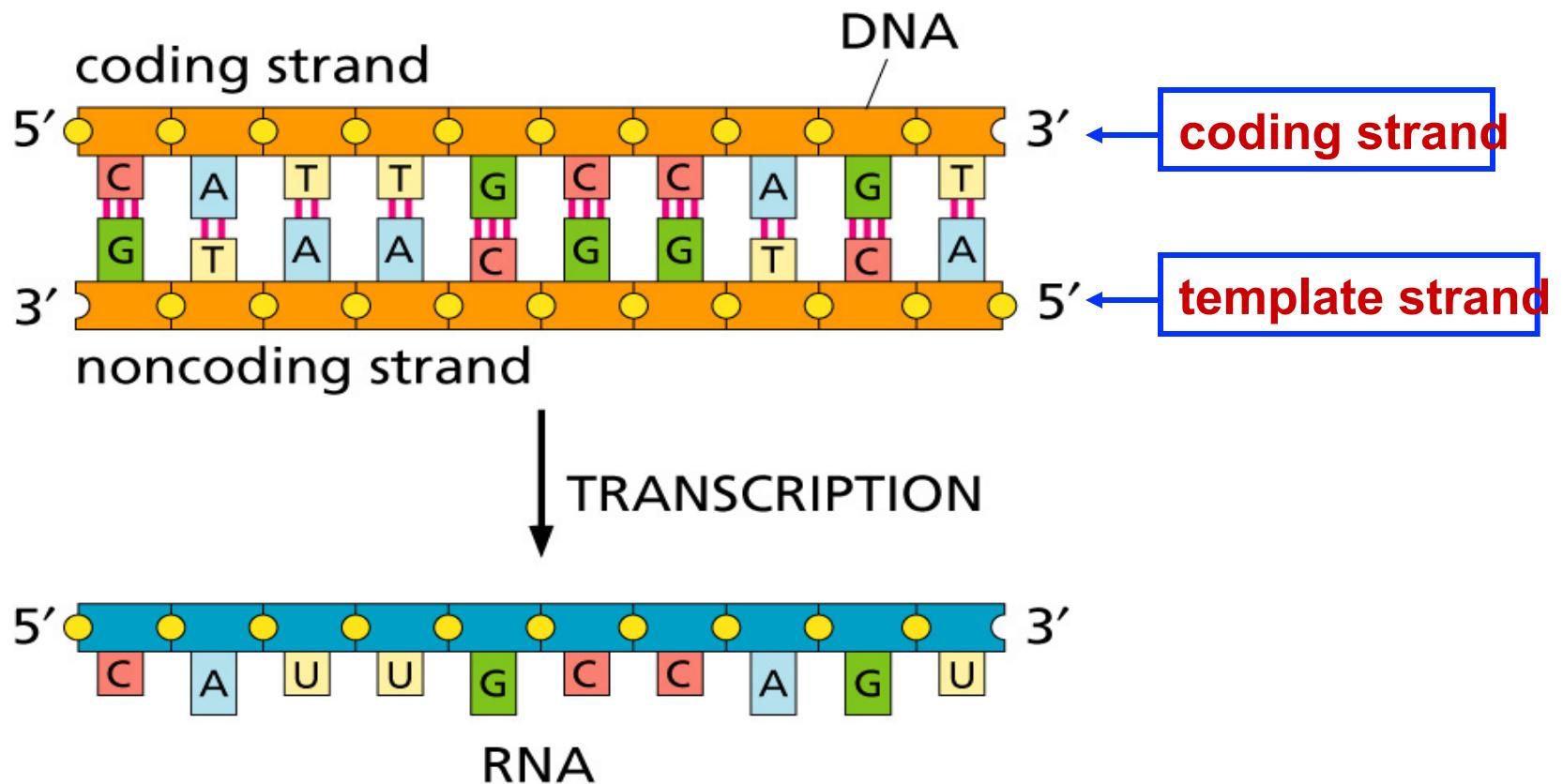


2. Translation

Transcription

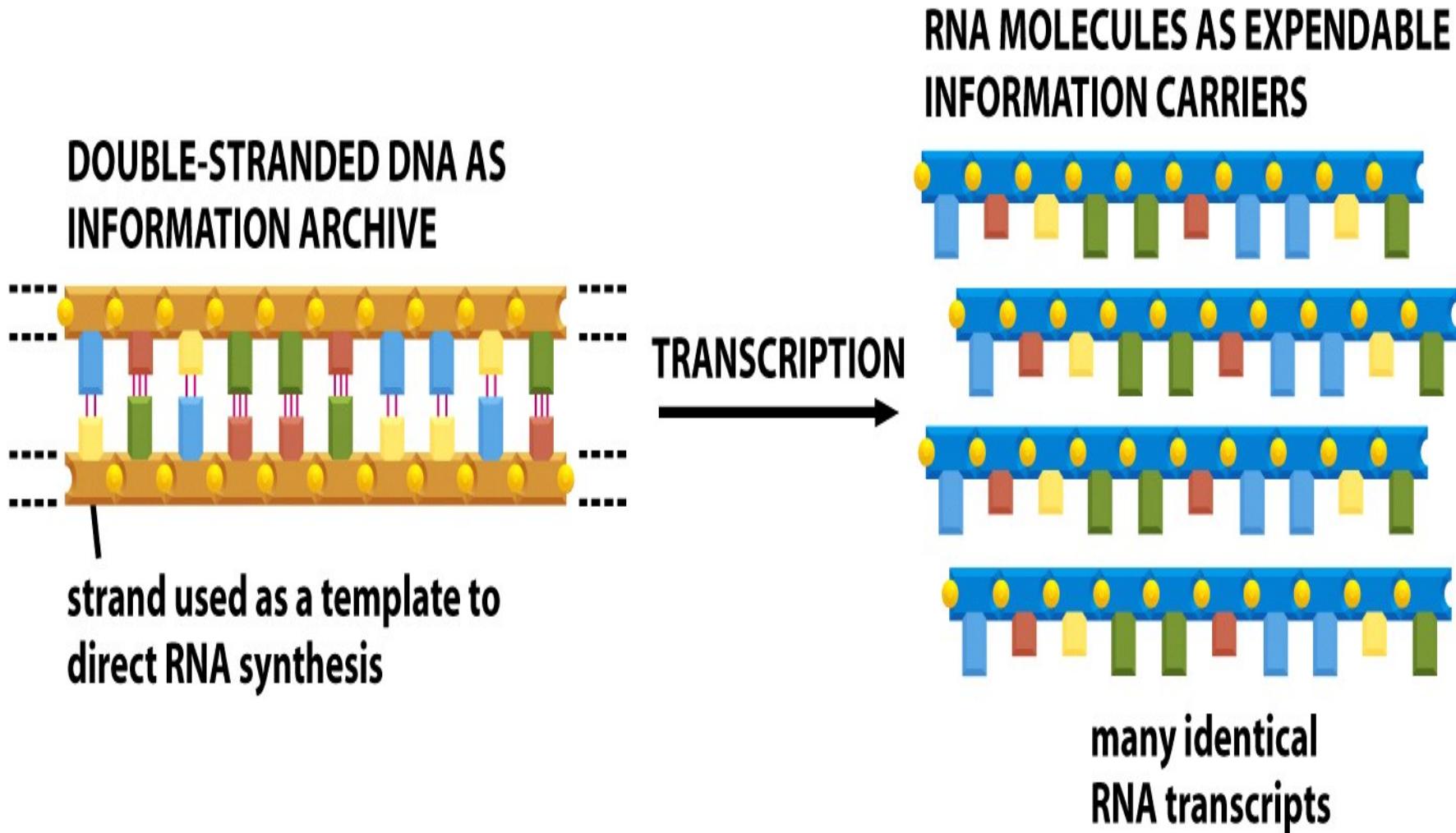


Transcription

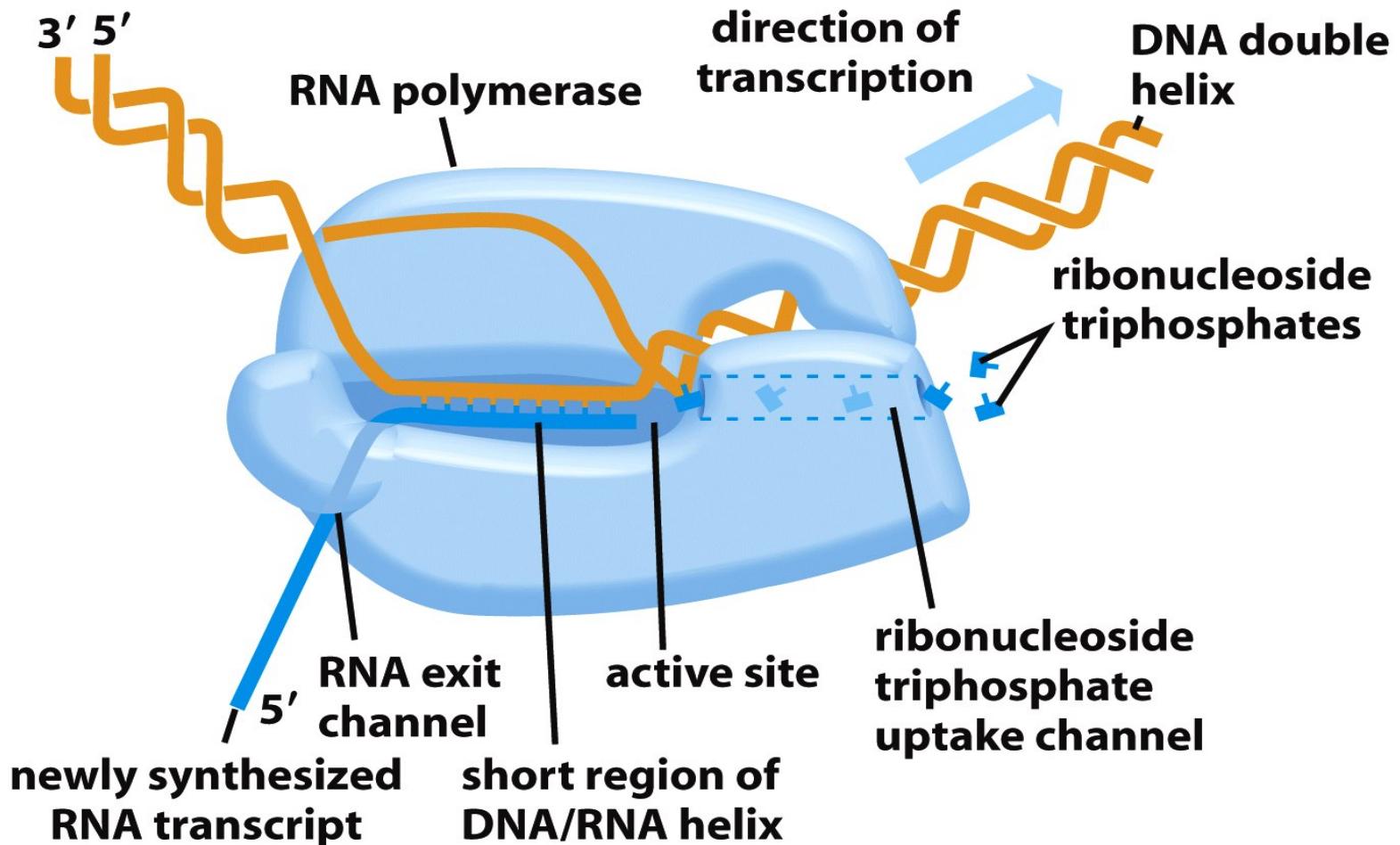


Transcription is the process in which one DNA strand: the **template strand**, is used to synthesize a complementary RNA.

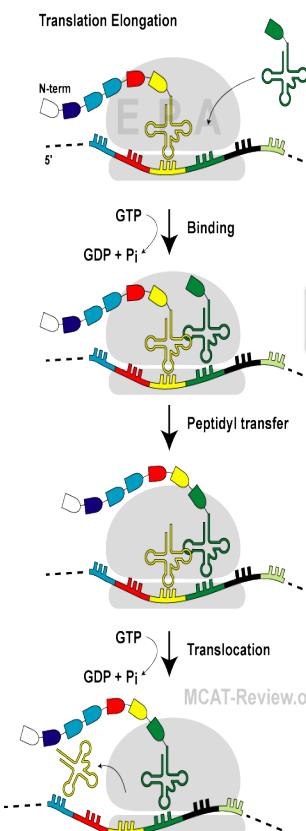
RNA: Carrier of Information



Synthesizing RNA from 5' to 3'

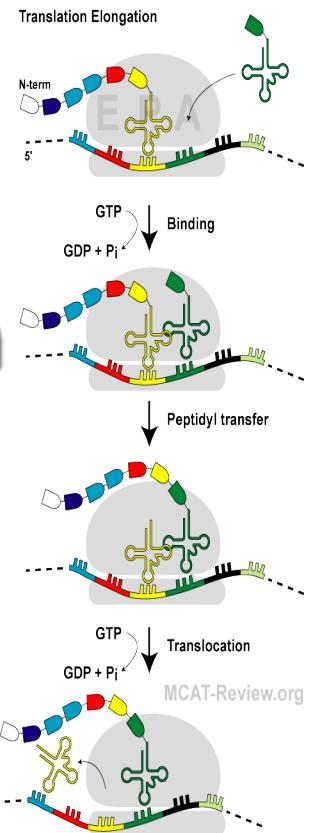
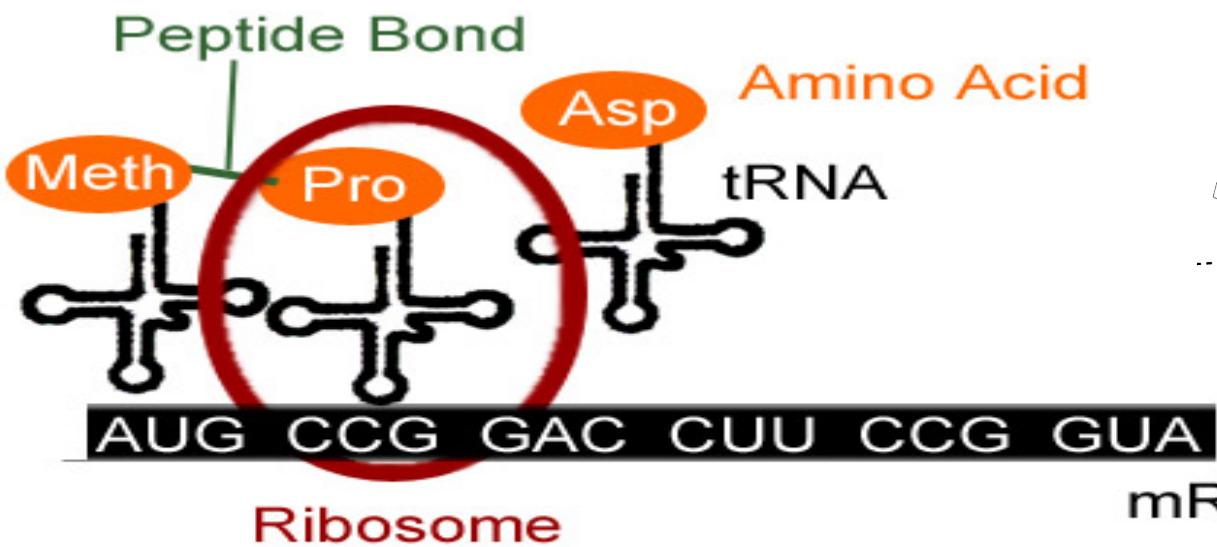


Translation



Transcription
DNA → RNA → Protein

Translation



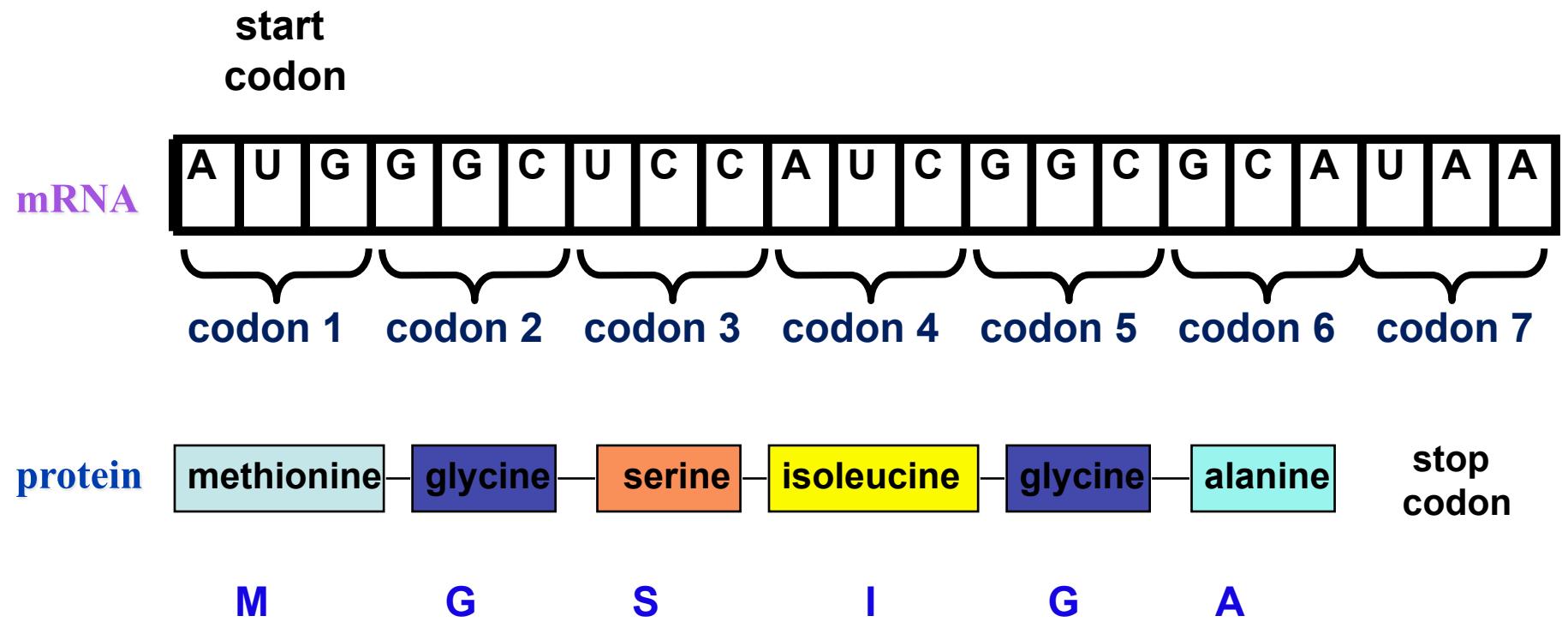
The Genetic Code

		SECOND BASE					
		U	C	A	G		
FIRST BASE	U	UUU } Phe UUC } UUA } Leu UUG }	UCU } Ser UCC } UCA } Ser UCG }	UAU } Tyr UAC } UAA } Stop UAG }	UGU } Cys UGC } UGA Stop UGG Trp	U	C
	C	CUU } Leu CUC } CUA } Leu CUG }	CCU } Pro CCC } CCA } Pro CCG }	CAU } His CAC } CAA } Gln CAG }	CGU } Arg CGC } CGA } Arg CGG }	U	C
	A	AUU } AUC } Ile AUA } AUG Met	ACU } Thr ACC } ACA } Thr ACG }	AAU } Asn AAC } AAA } Lys AAG }	AGU } Ser AGC } AGA } Arg AGG }	U	C
	G	GUU } Val GUC } GUA } Val GUG }	GCU } Ala GCC } GCA } Ala GCG }	GAU } Asp GAC } GAA } Glu GAG }	GGU } Gly GGC } GGA } Gly GGG }	U	C
		THIRD BASE					

The Genetic Code

		The Genetic Code					
		Second Codon Position					
First Codon Position (5' End)	U	C	A	G			Third Codon Position (3' End)
	U	UUU Phe (F) UUC Phe (F) UUA Leu (L) UUG Leu (L)	UCU Ser (S) UCC Ser (S) UCA Ser (S) UCG Ser (S)	UAU Tyr (Y) UAC Tyr (Y) UAA Stop UAG Stop	UGU Cys (C) UGC Cys (C) UGA Stop UGG Trp (W)	U C A G	
	C	CUU Leu (L) CUC Leu (L) CUA Leu (L) CUG Leu (L)	CCU Pro (P) CCC Pro (P) CCA Pro (P) CCG Pro (P)	CAU His (H) CAC His (H) CAA Gln (Q) CAG Gln (Q)	CGU Arg (R) CGC Arg (R) CGA Arg (R) CGG Arg (R)	U C A G	
	A	AUU Ile (I) AUC Ile (I) AUA Ile (I) AUG Met (M)	ACU Thr (T) ACC Thr (T) ACA Thr (T) ACG Thr (T)	AAU Asn (N) AAC Asn (N) AAA Lys (K) AAG Lys (K)	AGU Ser (S) AGC Ser (S) AGA Arg (R) AGG Arg (R)	U C A G	
	G	GUU Val (V) GUC Val (V) GUA Val (V) GUG Val (V)	GCU Ala (A) GCC Ala (A) GCA Ala (A) GCG Ala (A)	GAU Asp (D) GAC Asp (D) GAA Glu (E) GAG Glu (E)	GGU Gly (G) GGC Gly (G) GGA Gly (G) GGG Gly (G)	U C A G	

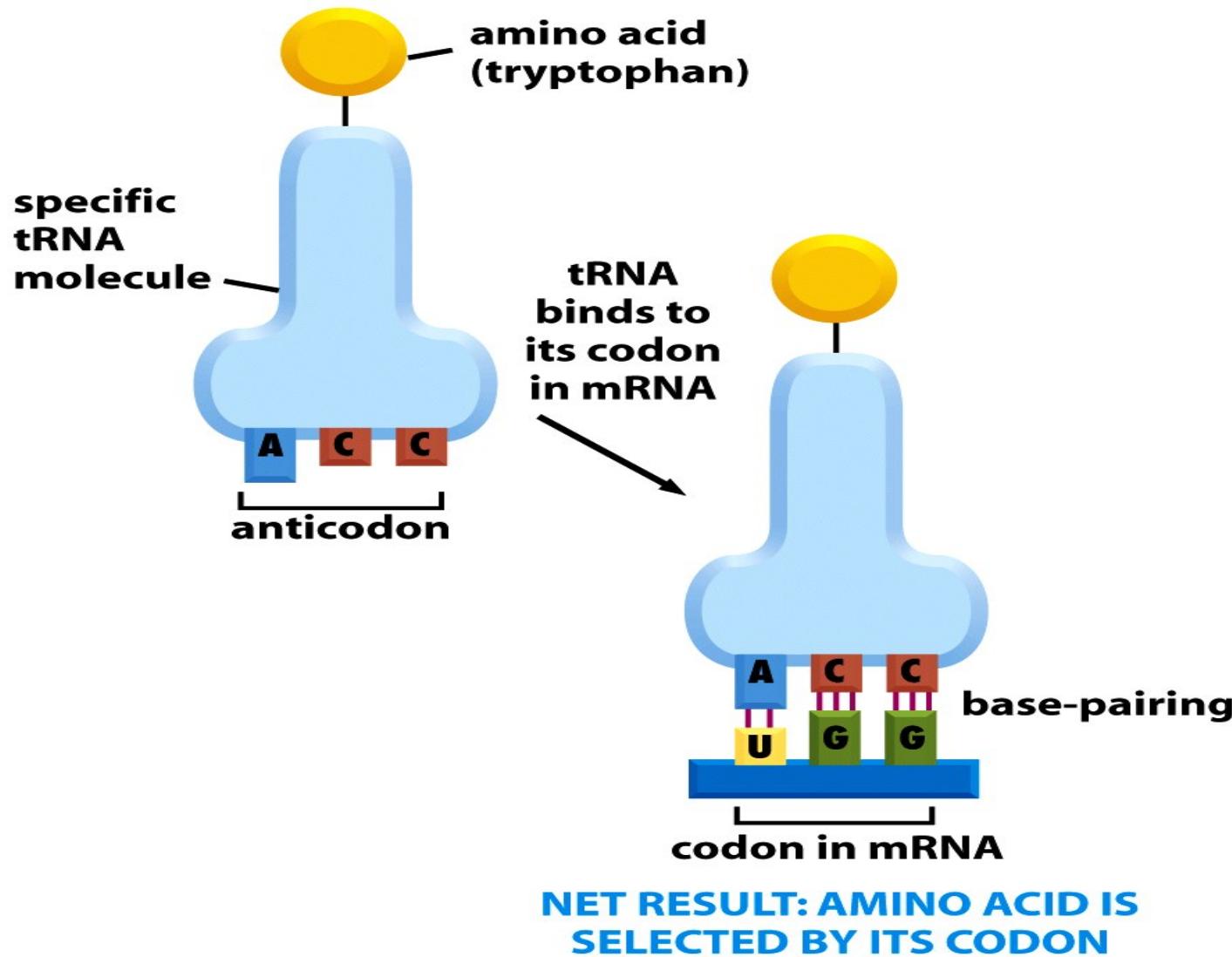
Translation: An Example



Transfer RNA and Translation

- The translation from nucleotides to amino acid is done by means of **transfer RNA (tRNA)** molecules, each specific for one amino acid and for a particular **triplet** of nucleotides in mRNA called a **codon**.
- The family of tRNA molecules enables the codons in a mRNA molecule to be **translated** into the sequence of amino acids in the protein.

tRNA, Anticodon and Codon



Codons and Anticodons

At least one kind of tRNA is present for each of the 20 amino acids used in protein synthesis.

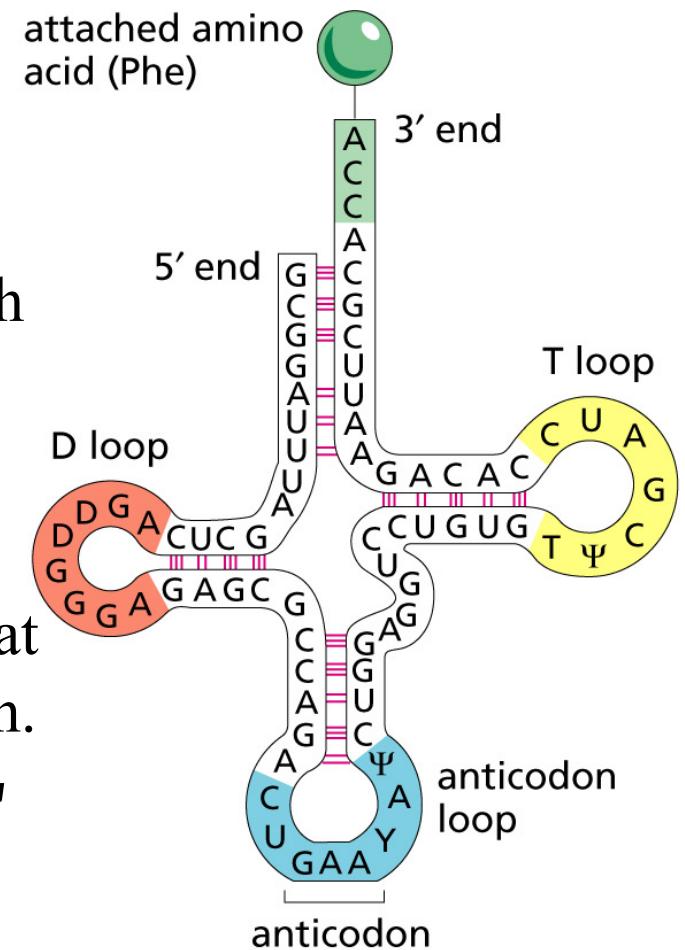
Each kind of tRNA has a sequence of 3 unpaired nucleotides - the **anticodon** - which can bind to the complementary triplet of nucleotides - the **codon** - in an mRNA molecule.

The reading of codons in mRNA requires that the anticodons bind in the opposite direction.

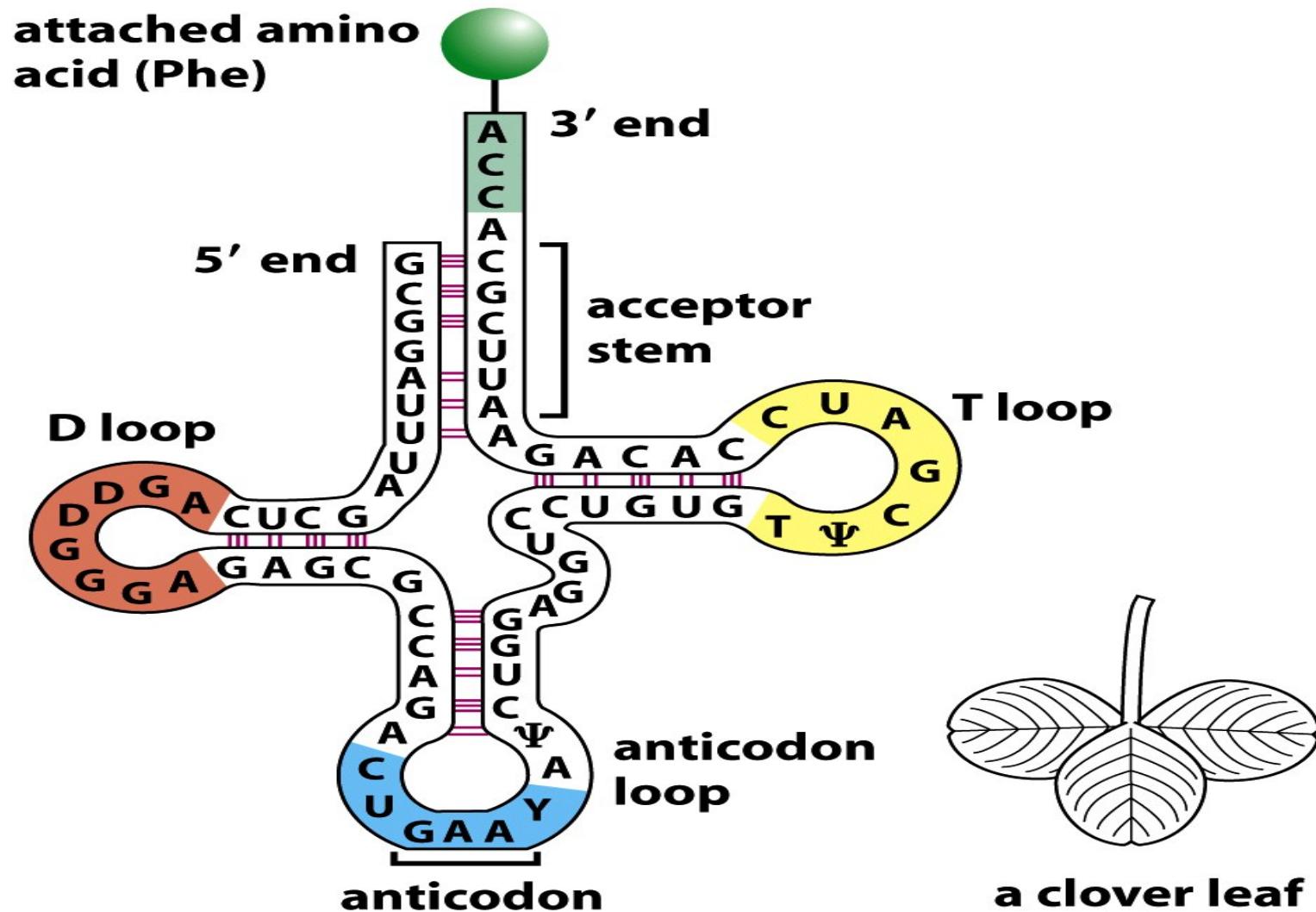
Anticodon : 3' AAG 5'

Codon : 5'

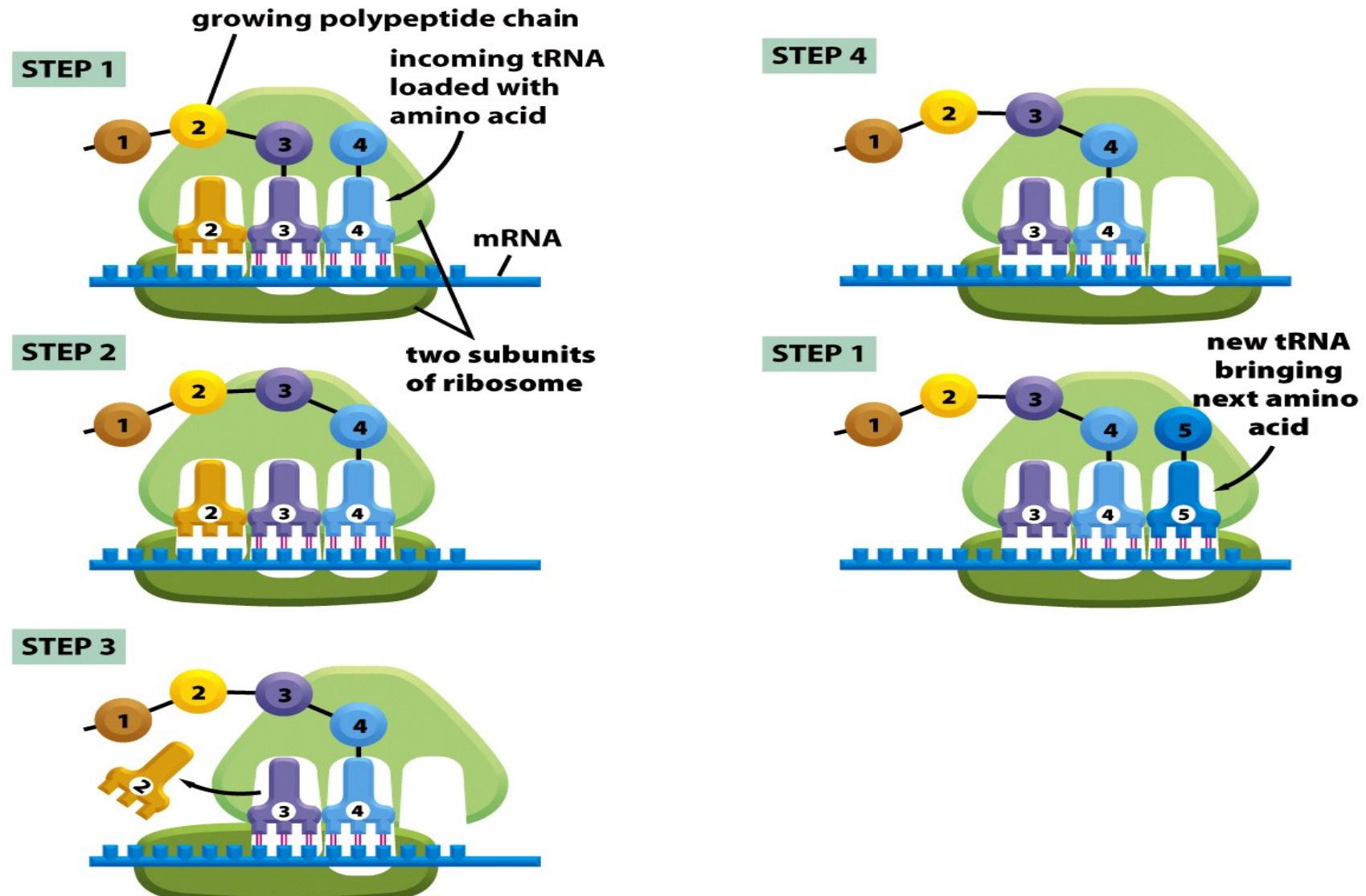
UUC 3'



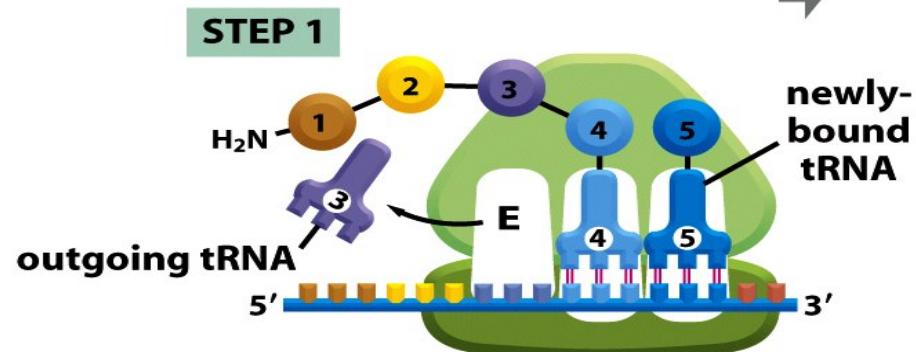
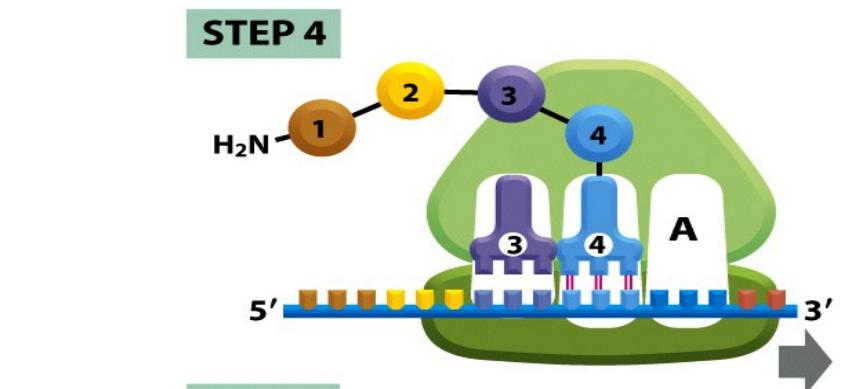
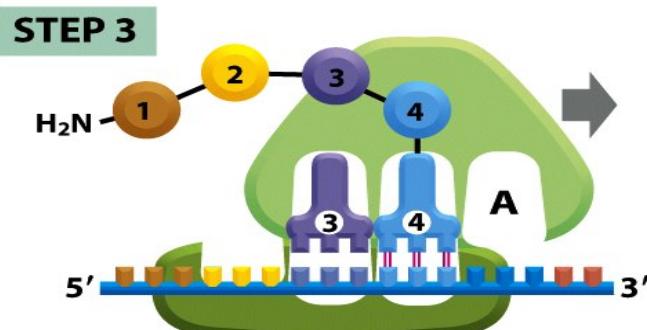
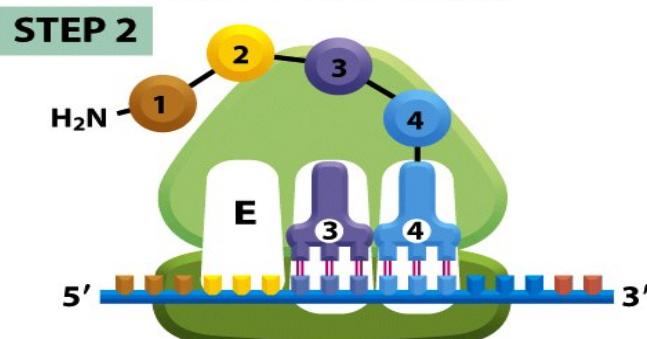
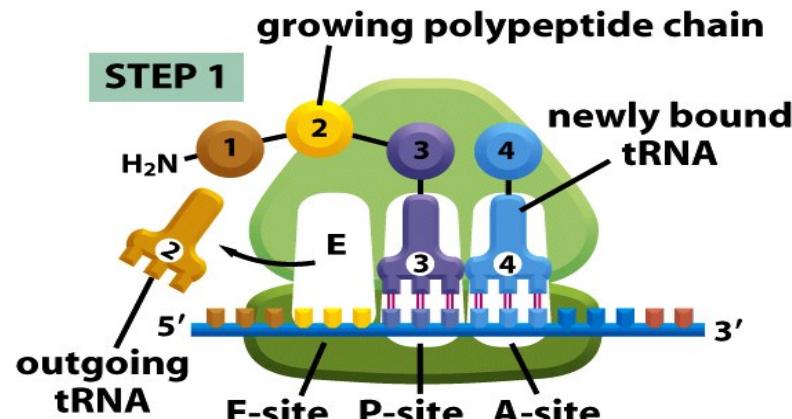
tRNA and Anticodon



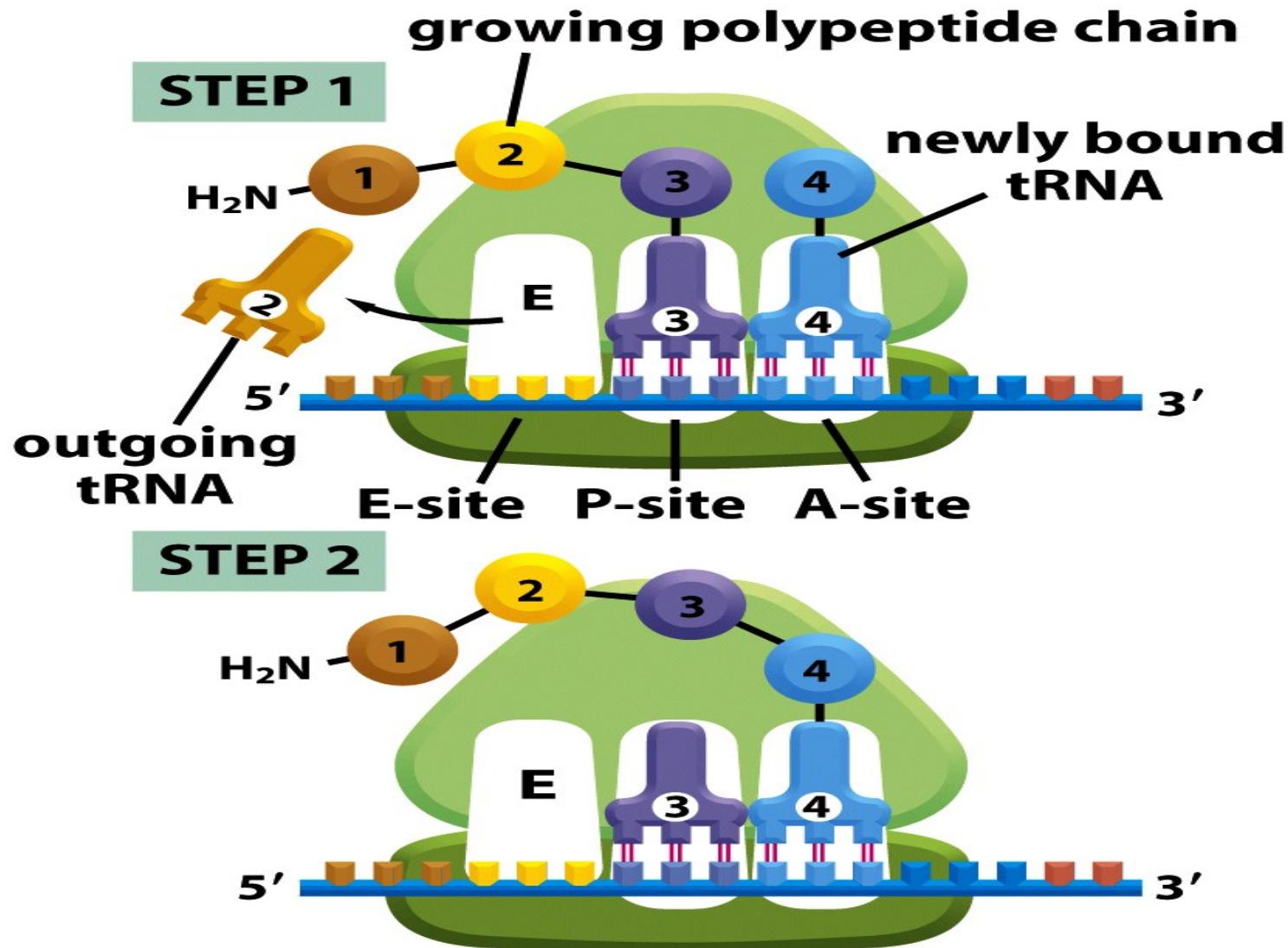
Steps of Translation



Translation Showing Ends

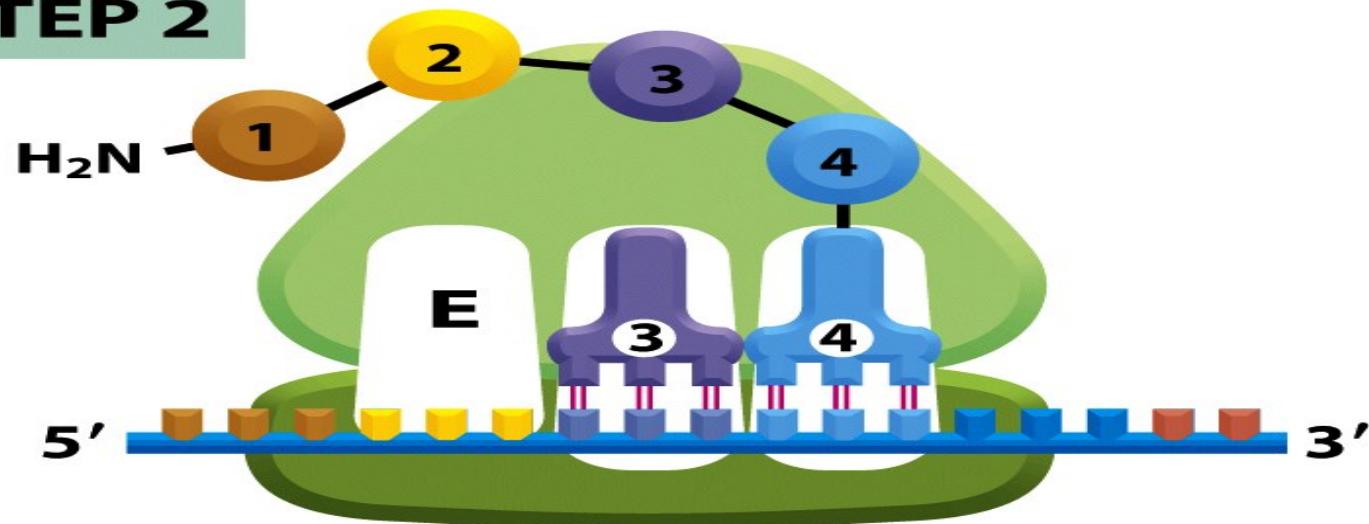


Translation: Steps 1 and 2

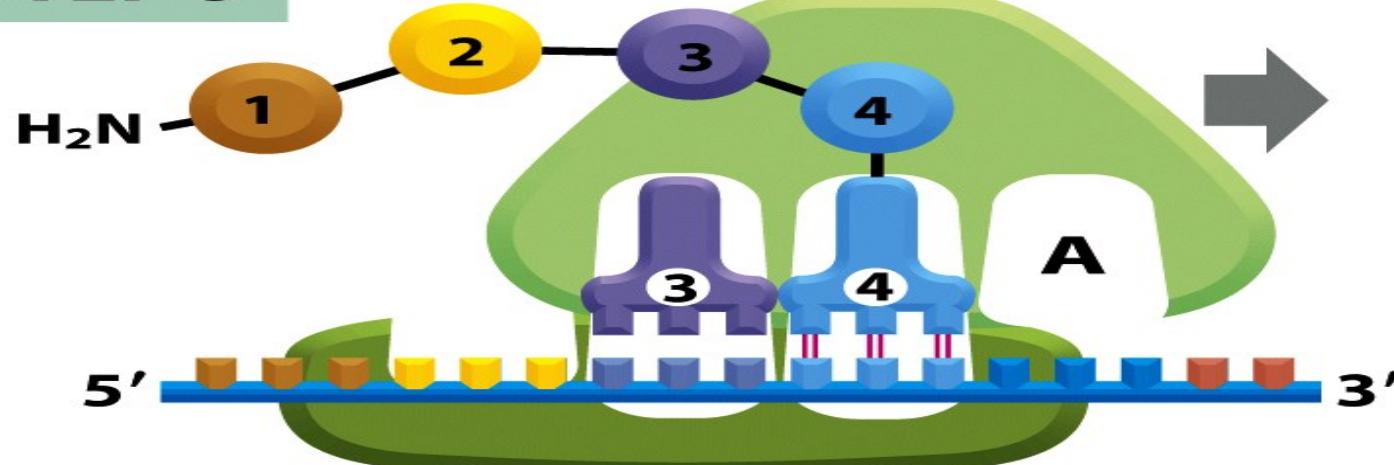


Translation: Steps 2 and 3

STEP 2

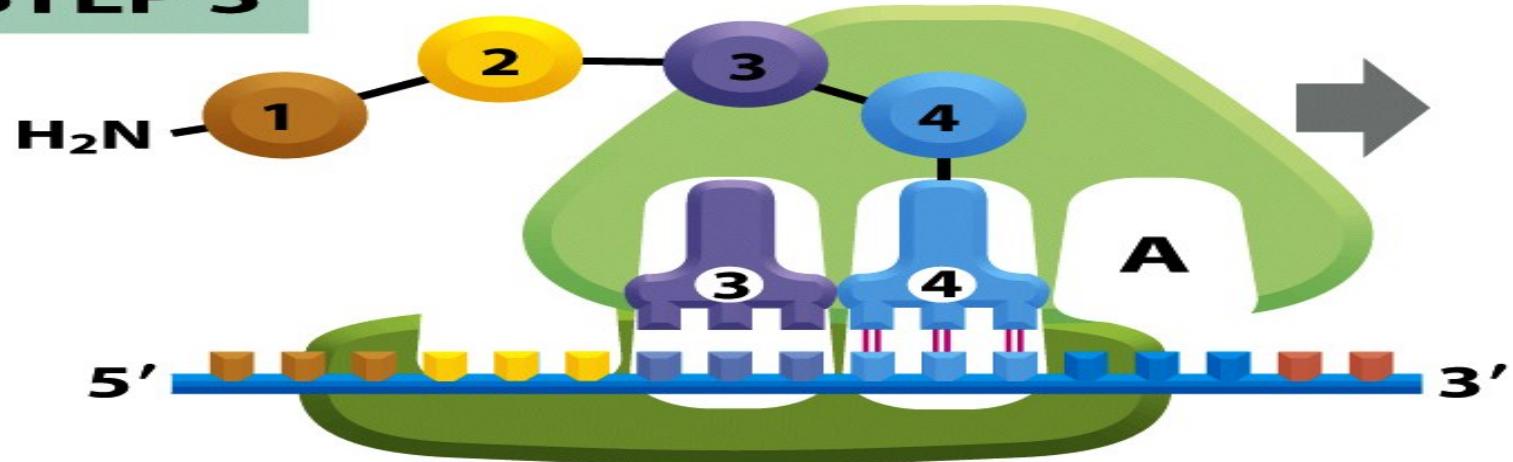


STEP 3

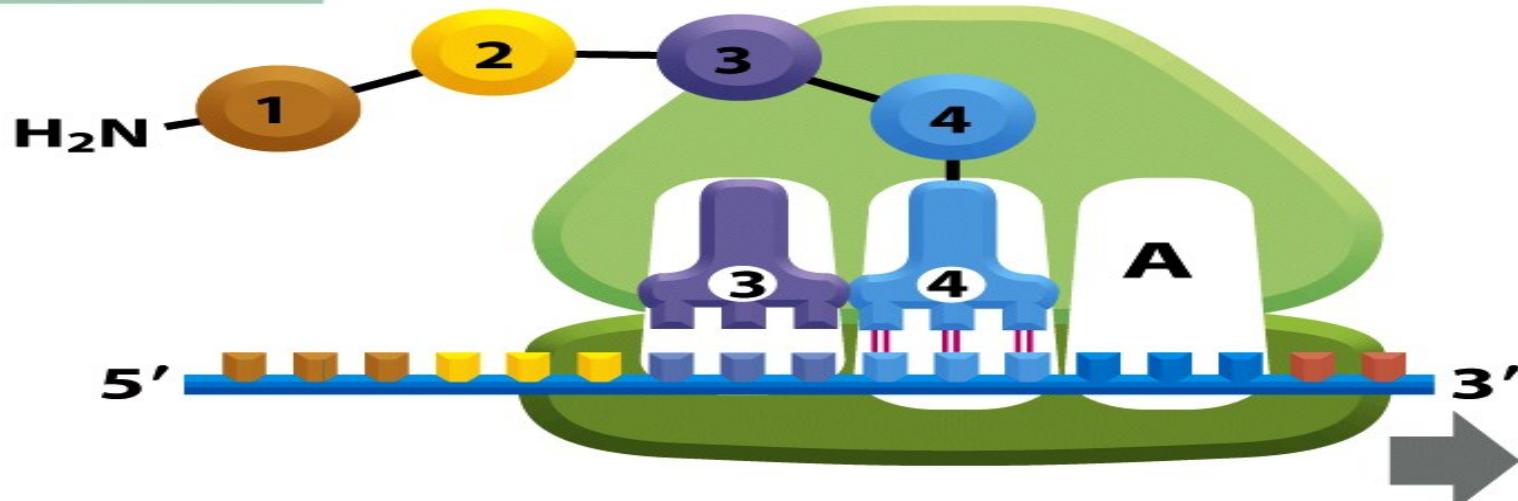


Translation: Steps 3 and 4

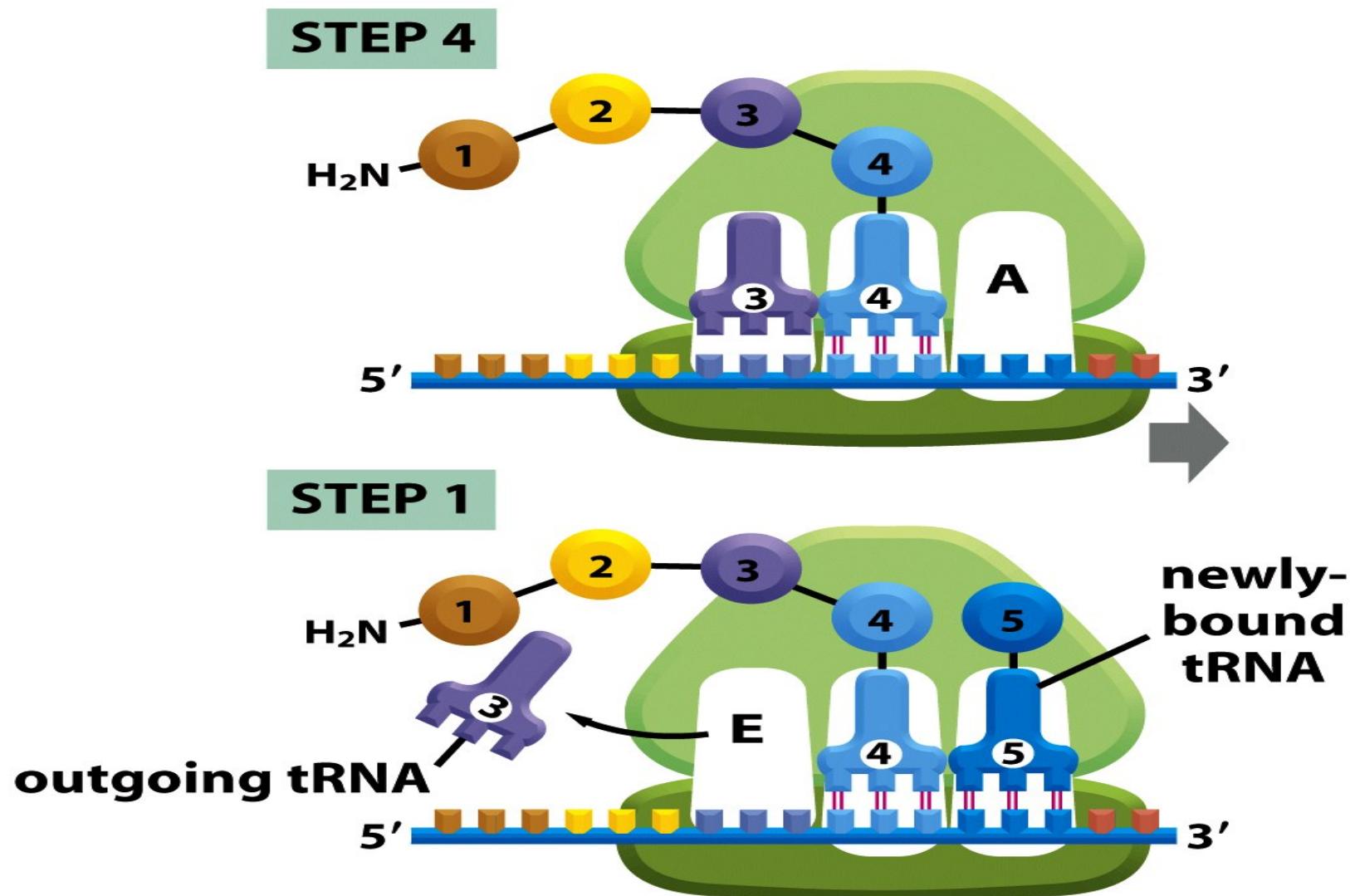
STEP 3



STEP 4



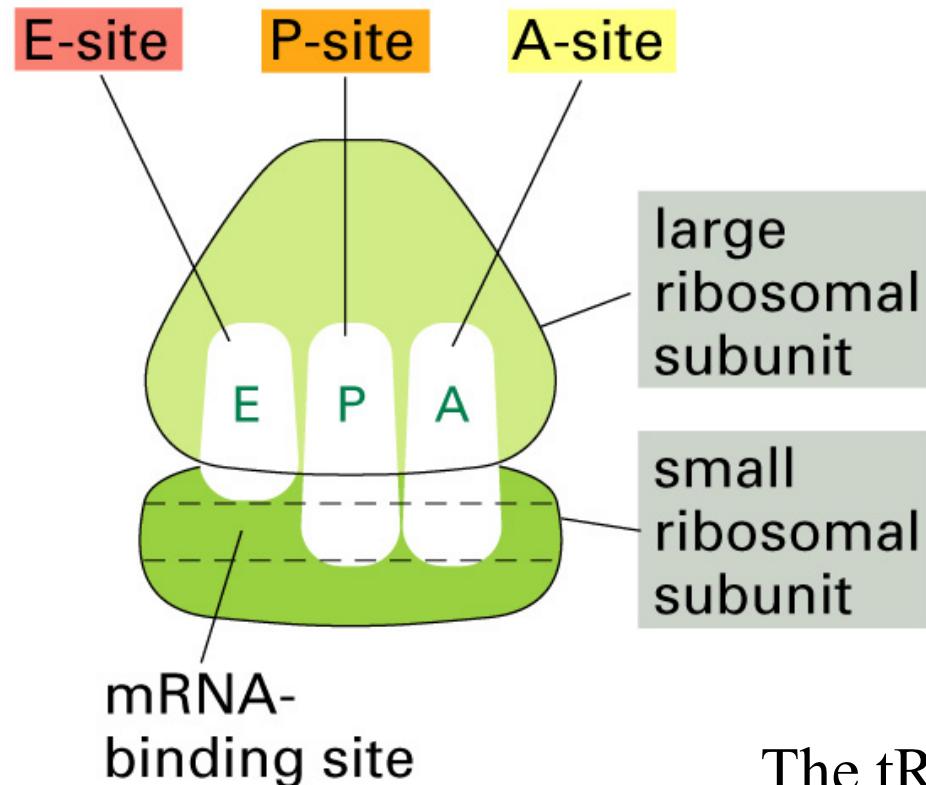
Translation: Steps 4 and 1



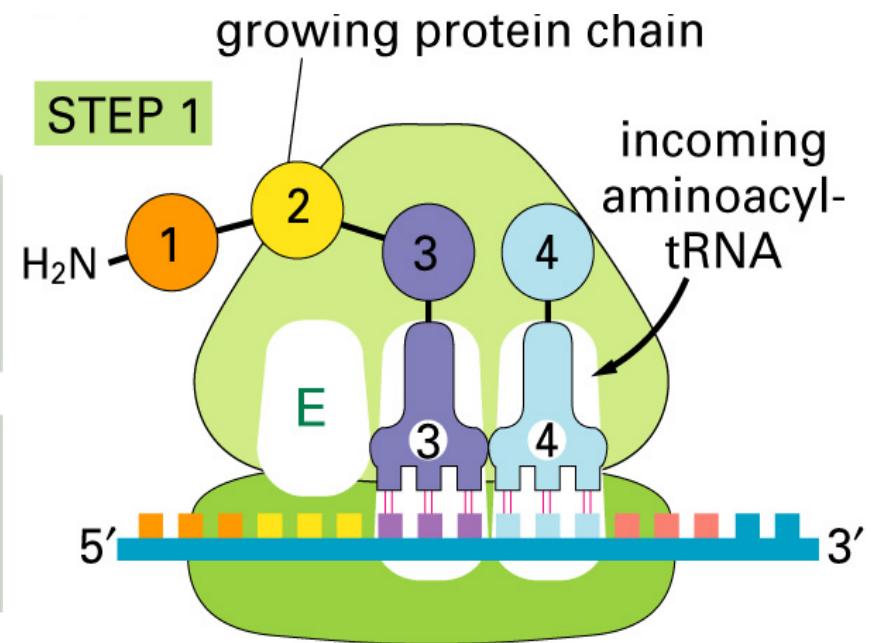
Start and Stop Codons

- The codon **AUG** serves two related functions
 - It begins most messages; that is, it signals the start of translation placing the amino acid methionine at the amino terminal of the polypeptide to be synthesized.
 - When it occurs within the message, it guides the incorporation of methionine.
- Three **codons**, UAA, UAG, and UGA, act as signals to terminate translation. They are called **STOP codons**.

Translation with Details



Binding site of ribosome
for the mRNA and the three
tRNA binding sites.



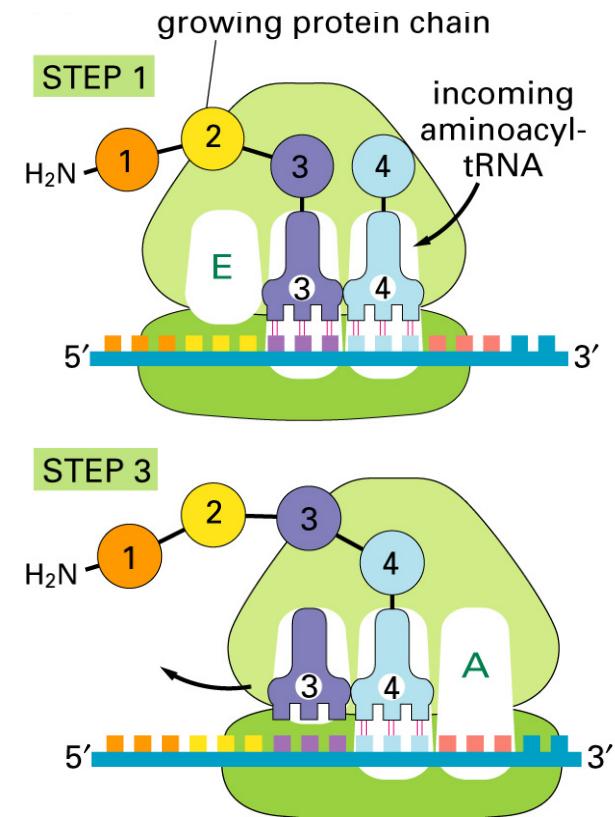
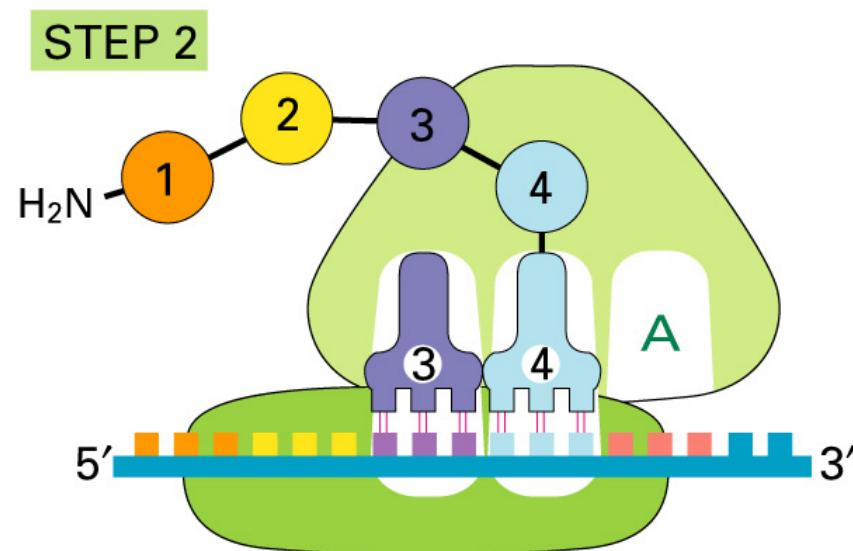
The tRNA molecules bind to the
ribosome and are the physical link
between the mRNA and the growing
protein chain.

Steps of Translation: Initiation

- The small subunit of the ribosome binds to a site “upstream” of the start of the message.
- It proceeds downstream until it encounters the **start codon AUG**.
- It is then joined by the large subunit and a special **initiator tRNA**. The initiator tRNA binds to the **P site** on the ribosome.
- In eukaryotes, **initiator tRNA** generally carries methionine (Met).

Steps of Translation: Elongation

An aminoacyl-tRNA able to base pair with the next codon on the mRNA arrives at the A site.

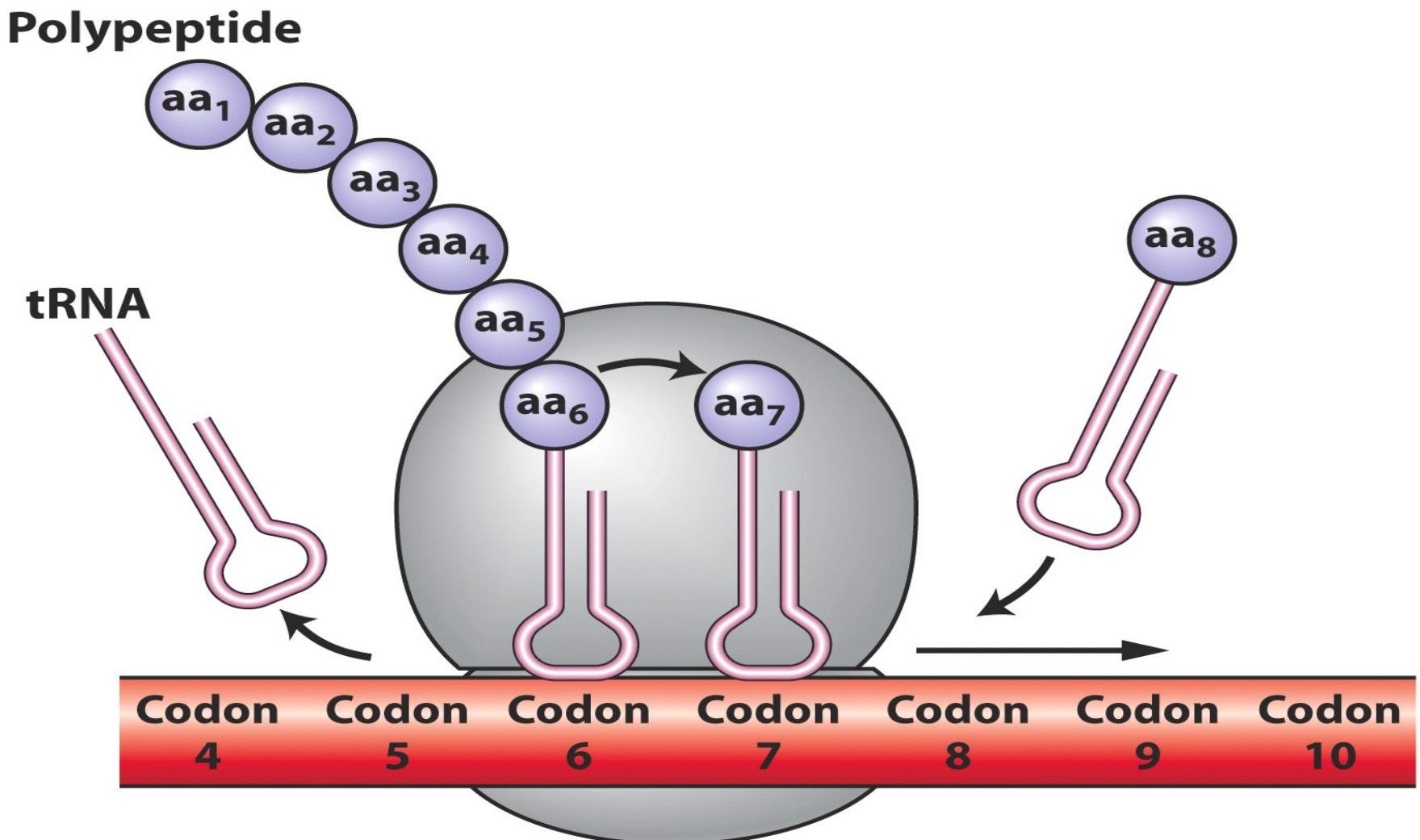


The preceding amino acid is linked to the incoming amino acid with a peptide bond.

Steps of Translation: Termination

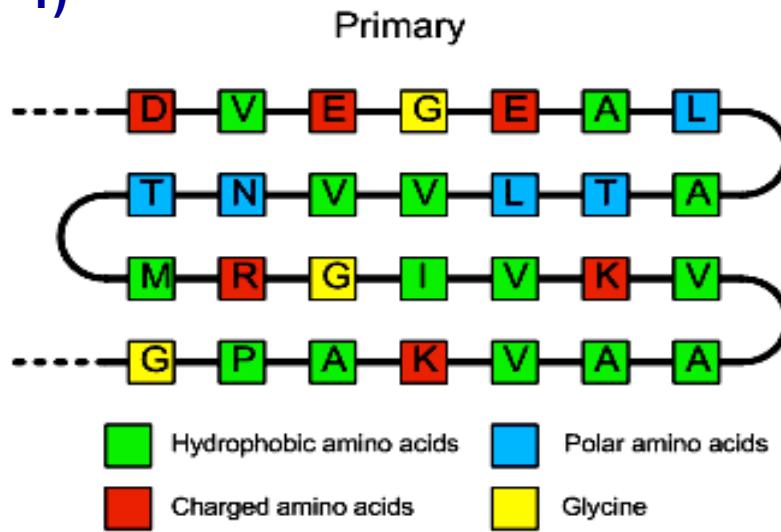
- The end of the message is marked by a **STOP codon**: **UAA, UAG, UGA**.
- No **tRNA** molecules have anticodons for **STOP codons**. Instead, protein release factor recognizes these codons when they arrive at the **A site**.
- Binding of this protein releases the **polypeptide** from the ribosome.
- The **ribosome** splits into its subunits, which can later be reassembled for another round of **protein synthesis**.

Chain of Amino Acids

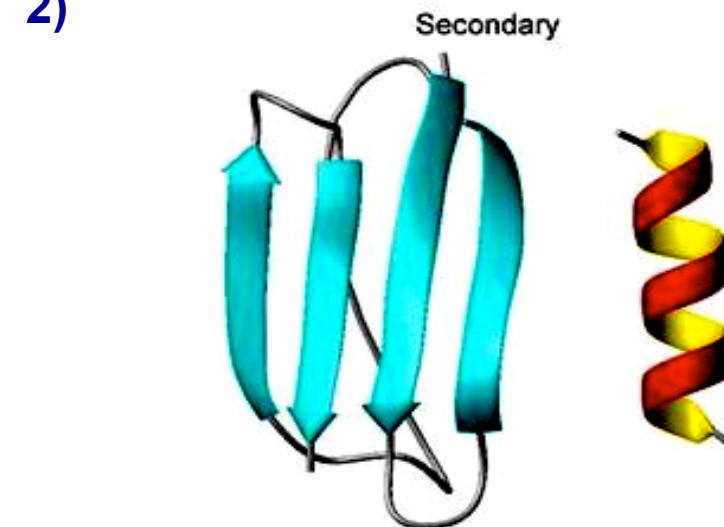


Four Structures of Proteins

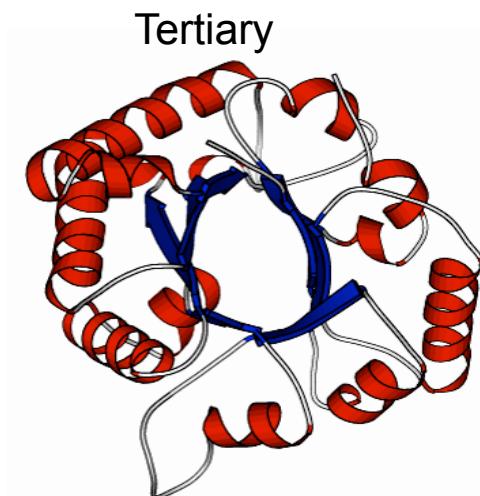
1)



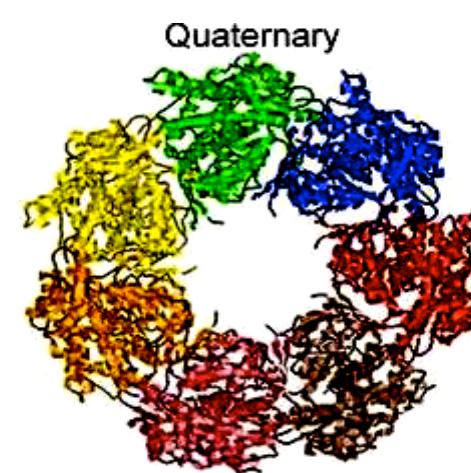
2)



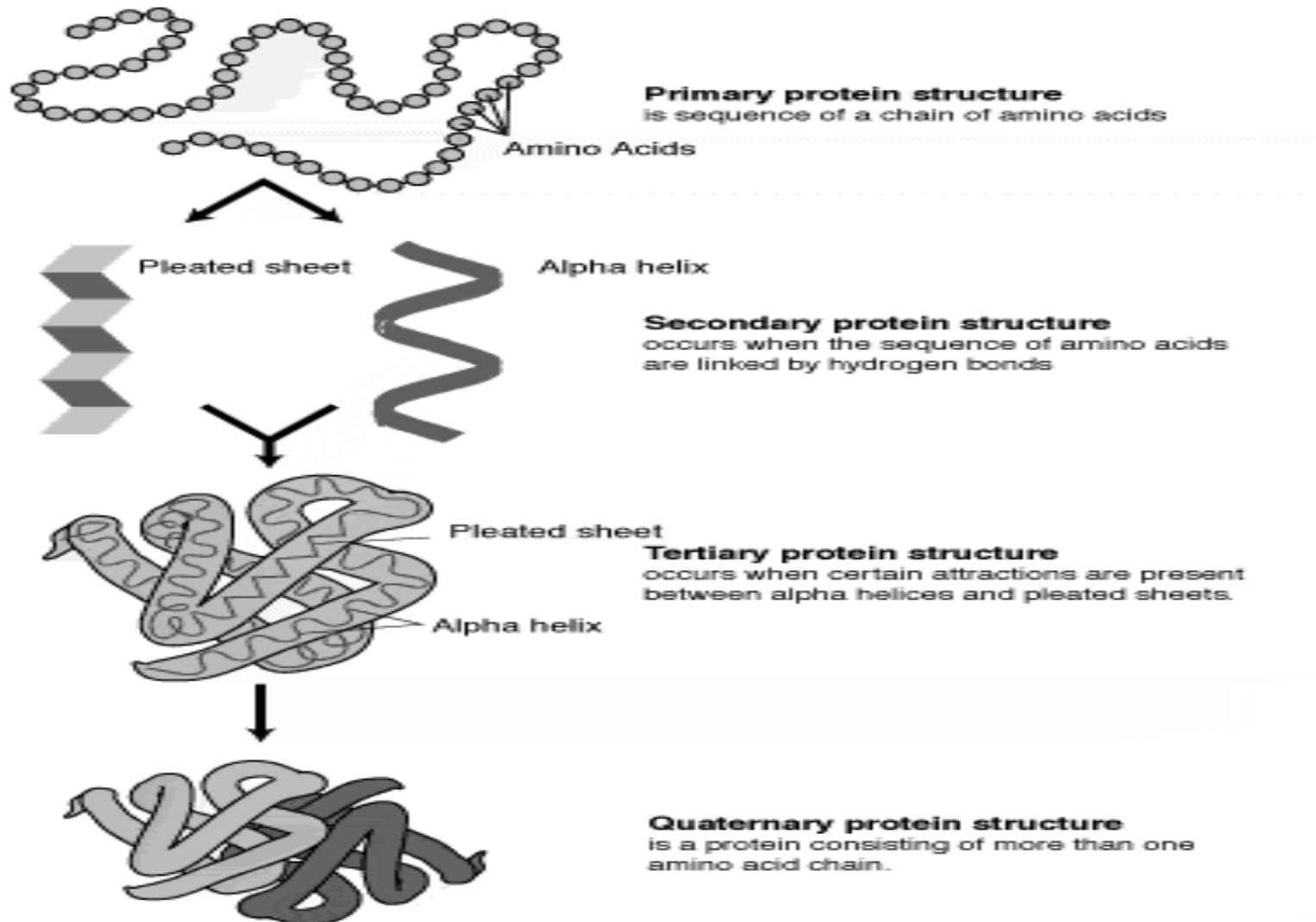
3)



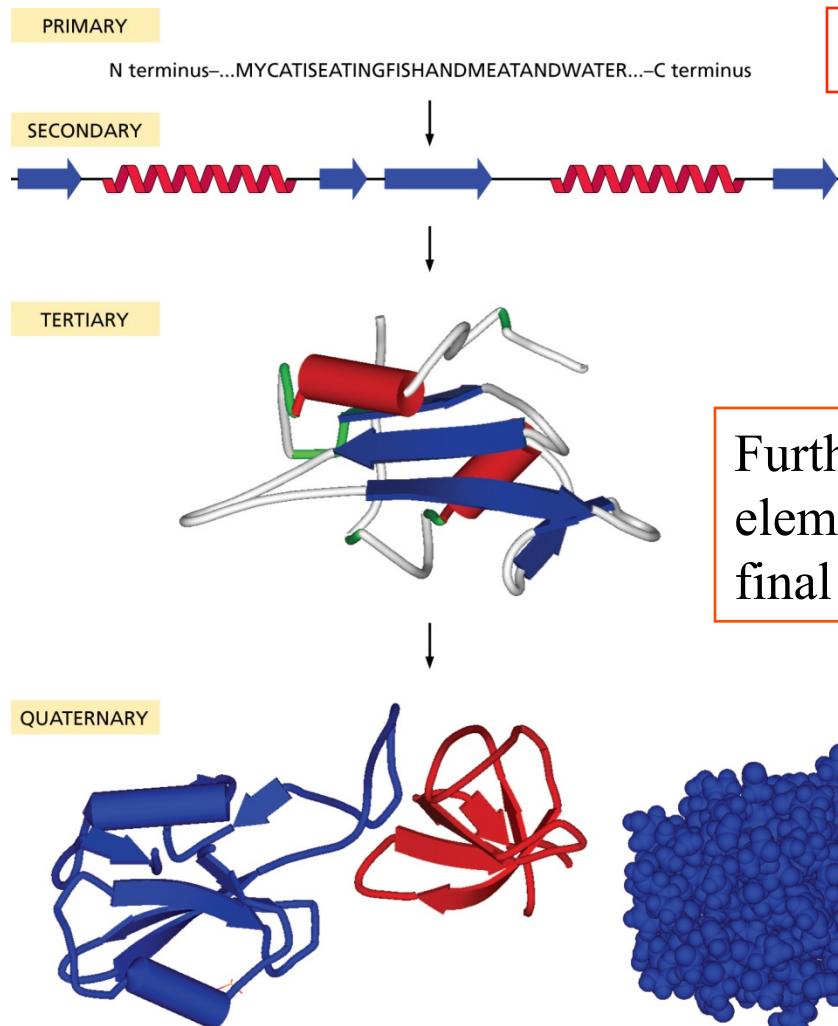
4)



Four Protein Structures



Four Levels of Protein Structure



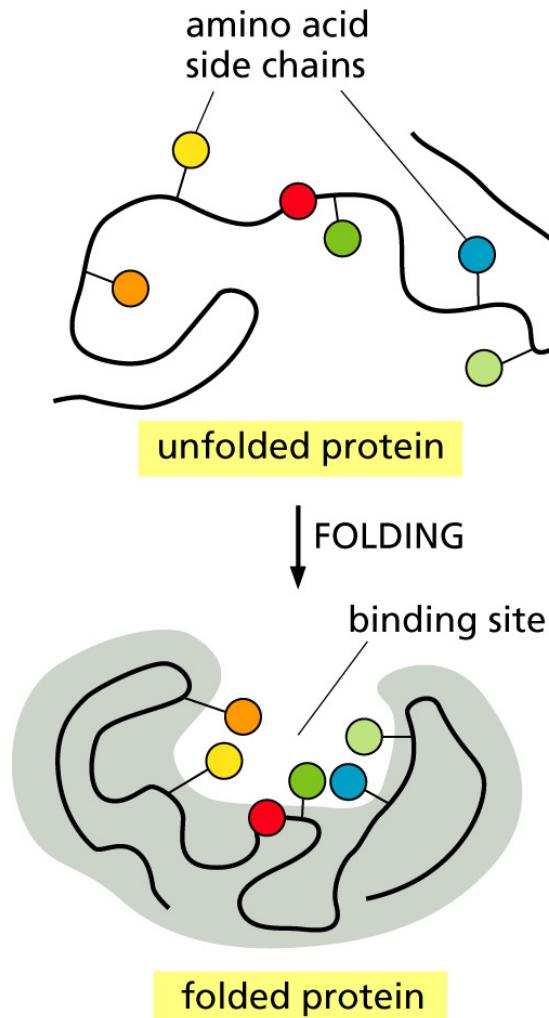
Chain of amino acids

First level of protein folding
Parts of the chain fold to form genetic structures found in all proteins

Further folding and packing together of the elements of secondary structure to produce the final 3-D conformation unique to the protein

Multisubunit protein formed of more than one protein chain.
Each protein chain is called a subunit

Folded Protein



Distant residues (in the primary structure of a protein) can come close in the folded structure.

When a polypeptide chain (primary structure) folds into a tertiary structure, residues that are far apart from each other in the sequence can come close together to form a binding site.

Protein Function

- The amino acid sequence of the protein is critical to its function in 2 ways:
 - The protein can only **interact** with other molecules if it has the correct amino acids with the correct side chains to **bind** those molecules.
 - The sequence of amino acids also determines how the protein can **fold**.

The ability of the protein to **interact** with specific molecules is dependent on the **folding** of its amino acid chain into a very specific three-dimensional shape.