AmirKabir University of Technology

Computational Biology Research Center

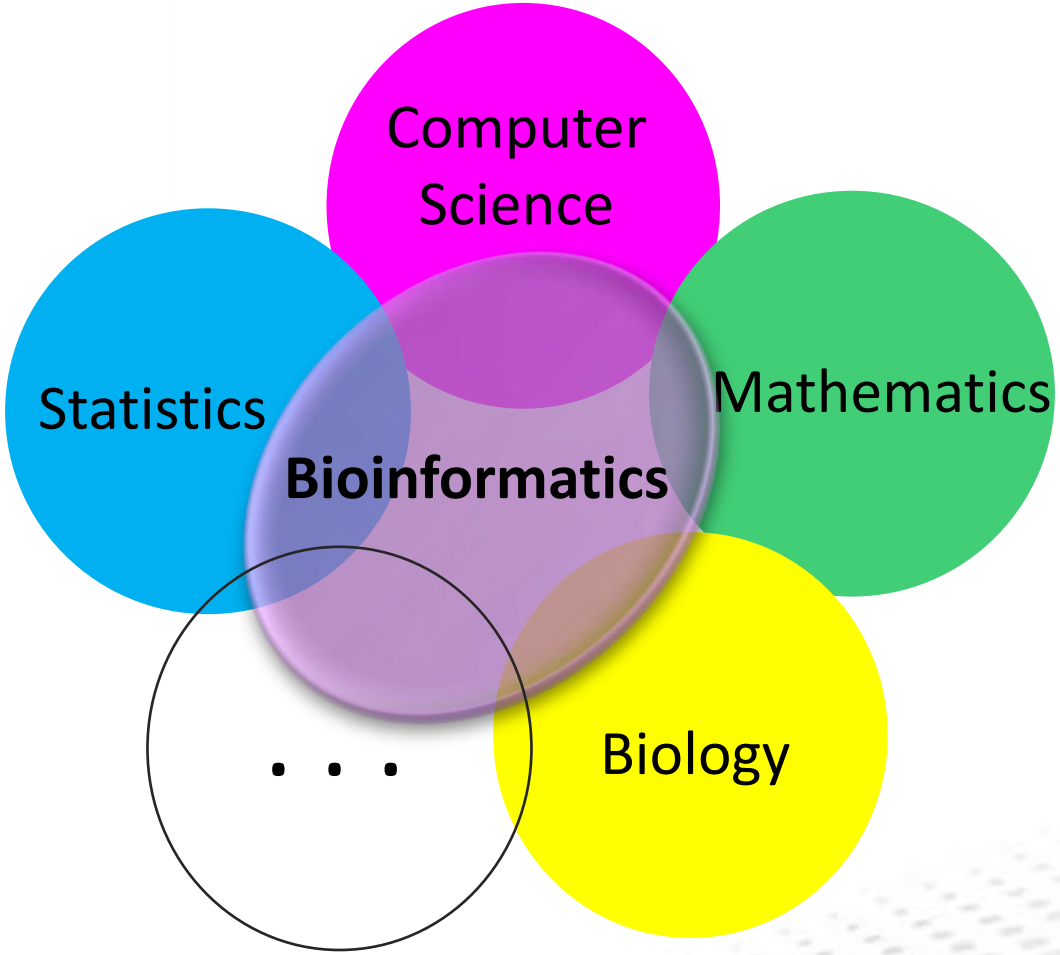
# Introduction to Cellular Biology and Bioinformatics

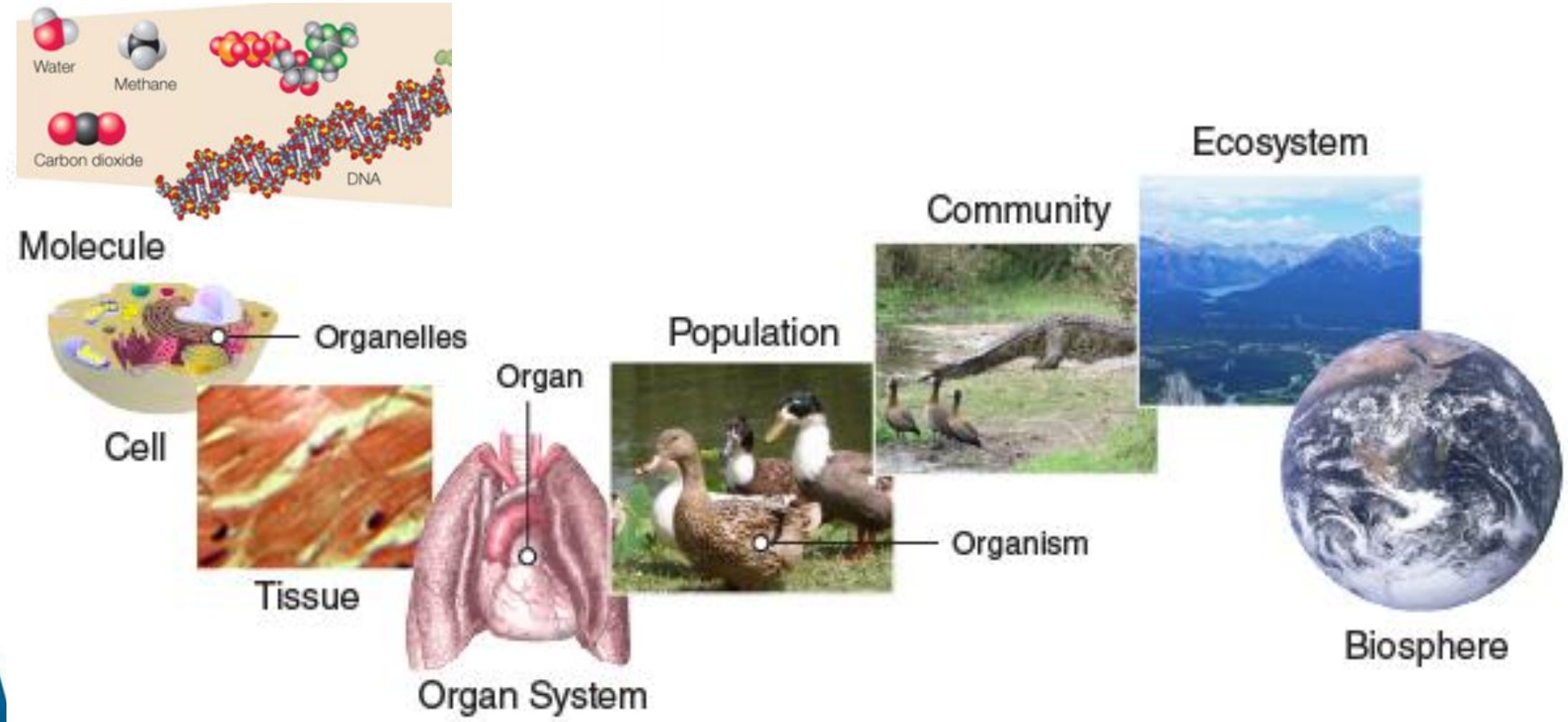
Farzaneh Salari

# Outline

- Bioinformatics
- Cellular Biology
- A Bioinformatics Problem

# What is bioinformatics?



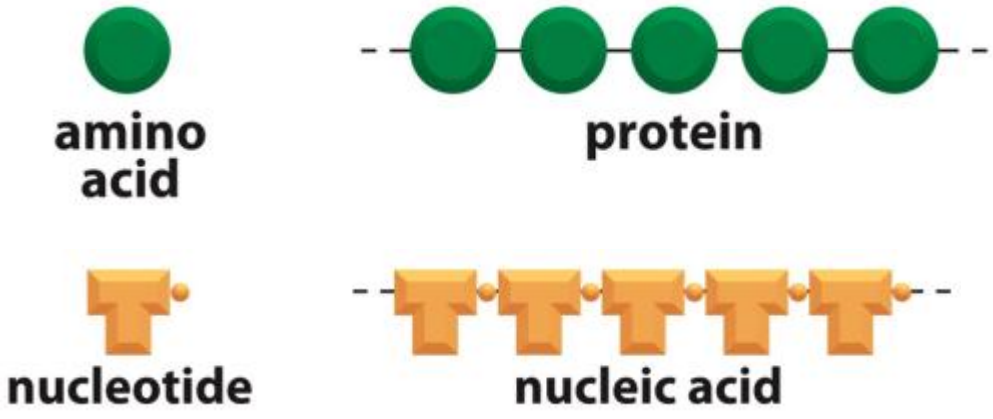

Water
Methane
Carbon dioxide
DNA
Molecule
Organelles
Cell
Tissue
Organ
Organ System
Population
Organism
Community
Ecosystem
Biosphere

# Macromolecules

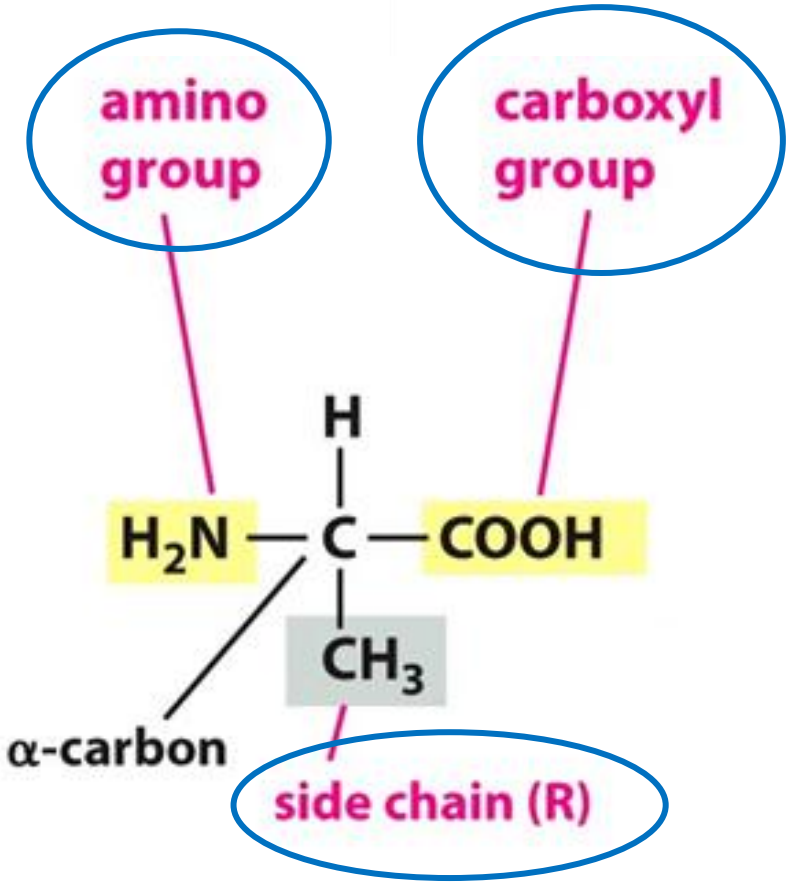
- Proteins
- DNA
- RNA

- bonds
  - Strong bond: Covalent bond
  - Weak bond: Hydrogen bond

- Structures
  - Primary structure: sequence
  - Secondary structure
  - Tertiary structure : function
  - Quaternary structure:  interaction

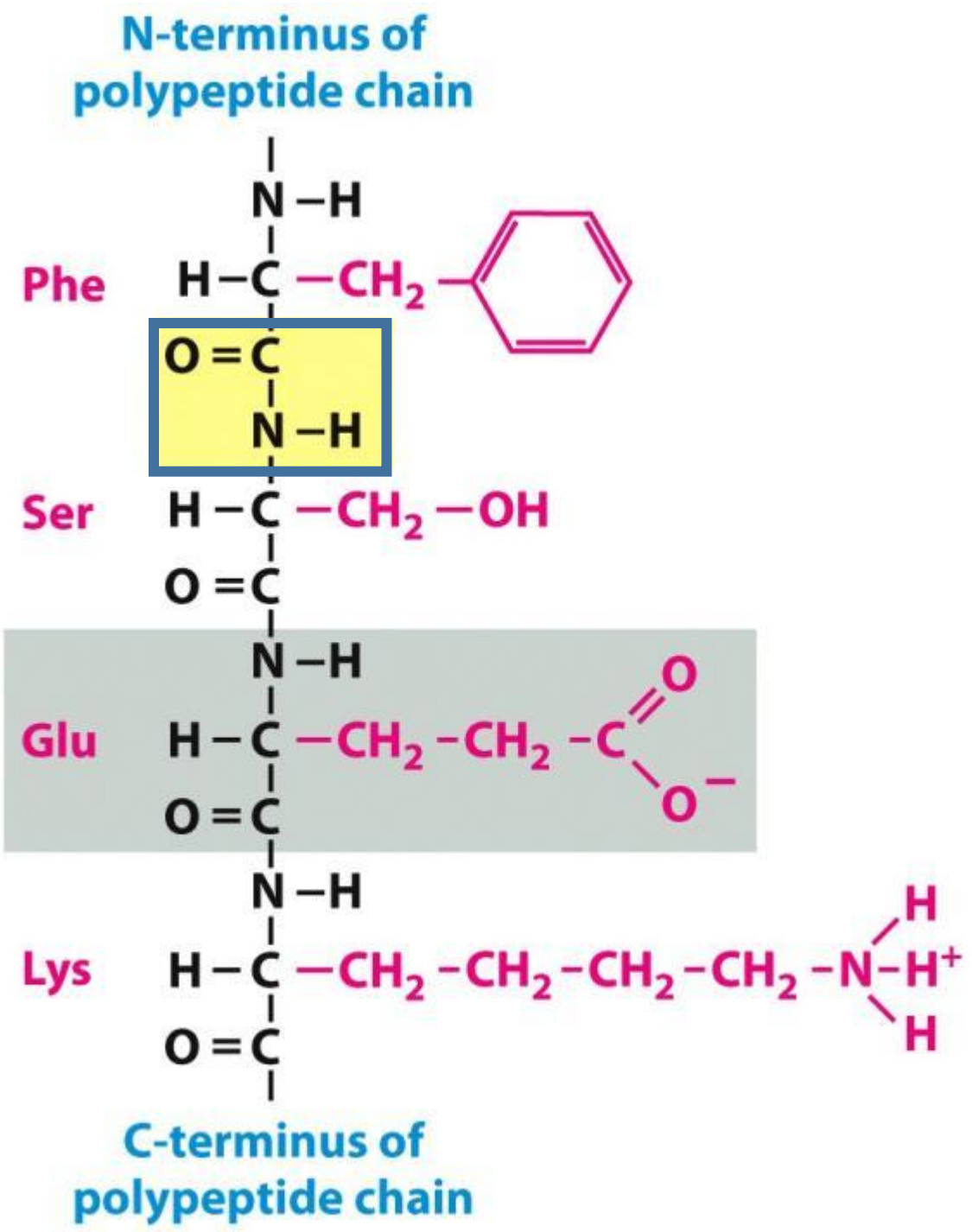


# Protein subunit (Amino acid)



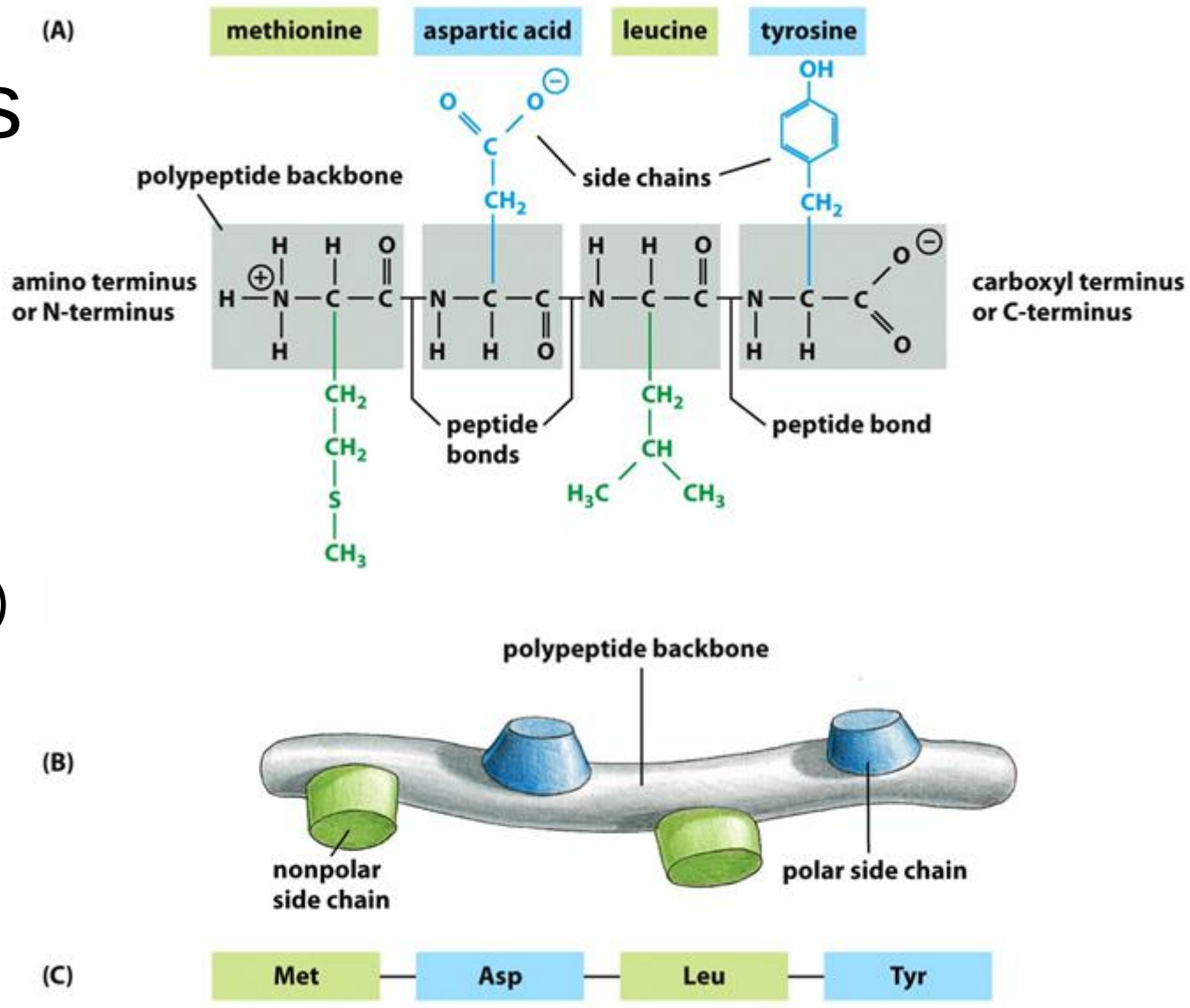
# Polypeptide chain

Peptide bond



# Protein Structures

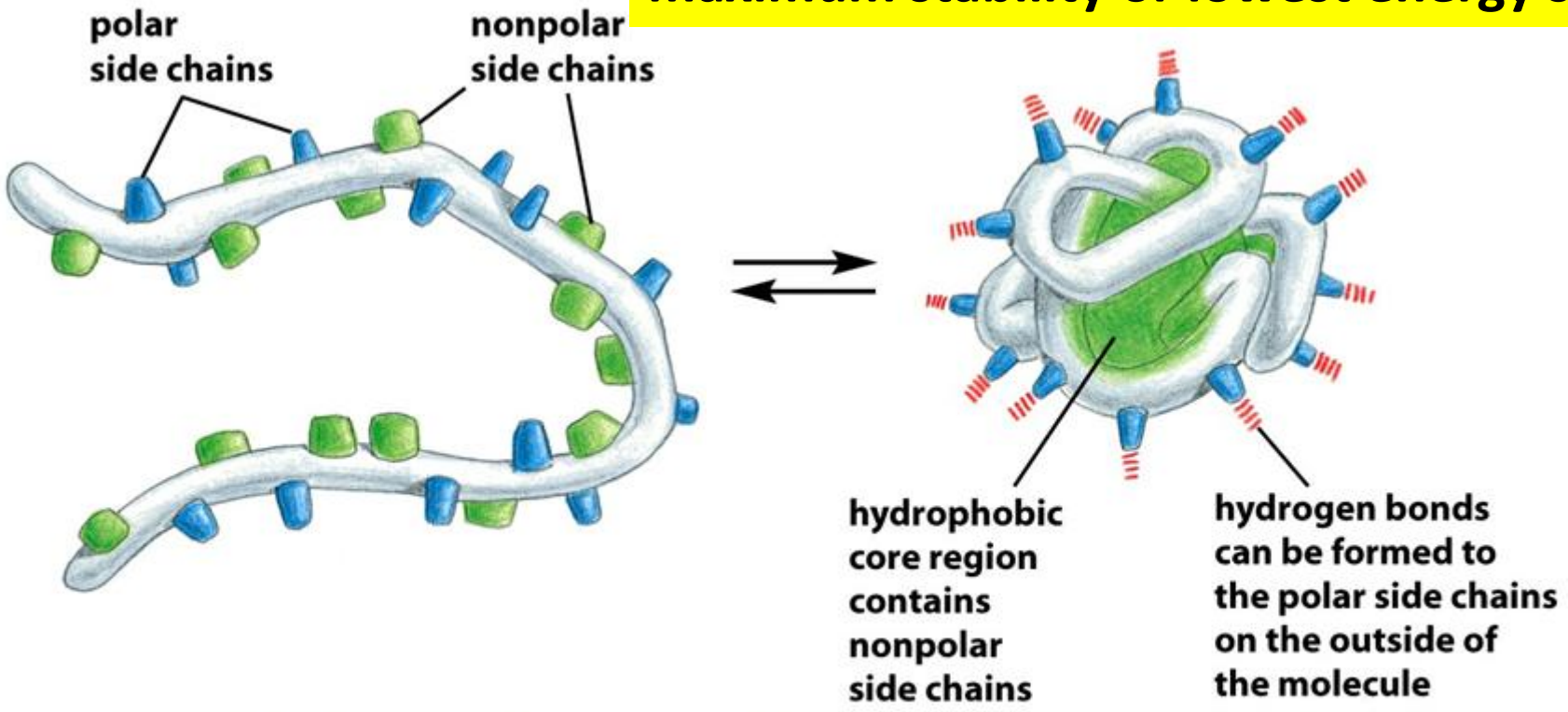
- 20 different  Amino acids
  - Hydrophilic (polar)
  - Hydrophobic (nonpolar)




Sequence

# Protein Structures

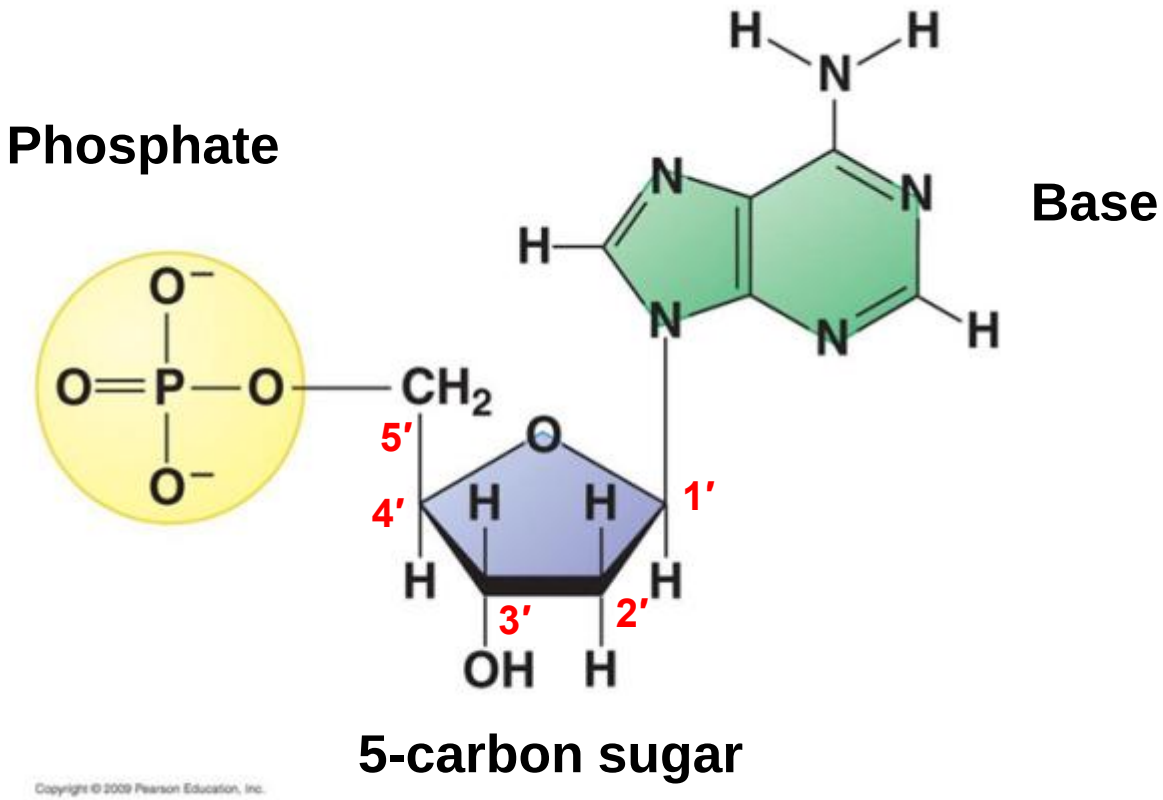
**maximum stability or lowest energy state**




**unfolded polypeptide**

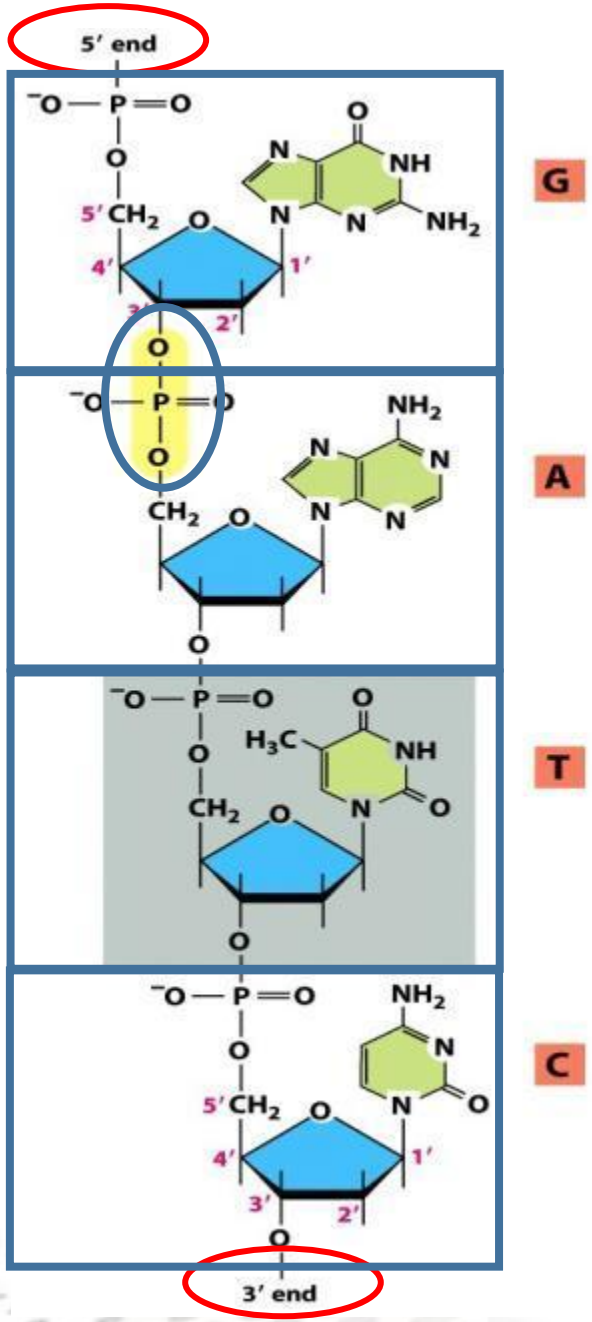
**folded conformation in aqueous environment**

# DNA subunit (Nucleotide)

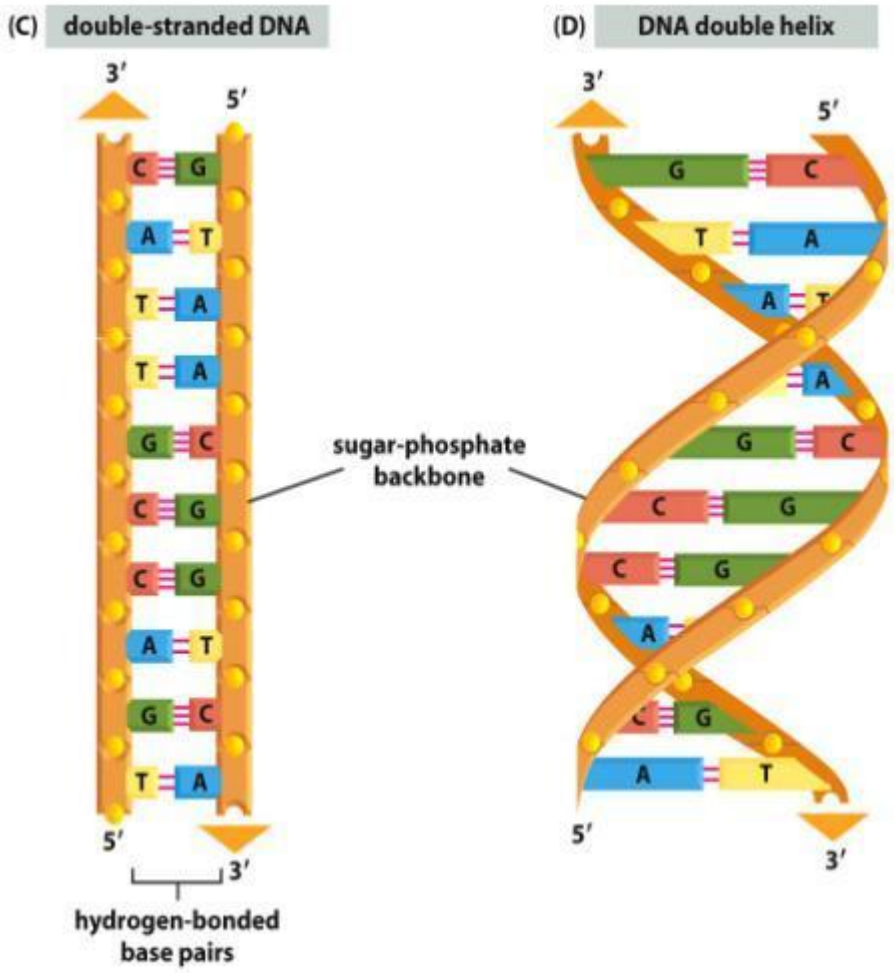


# Deoxyribonucleic acid (DNA)

- 4 different bases
  - Guanine (G)
  - Adenine (A)
  - Thymine (T)
  - Cytosine (C)



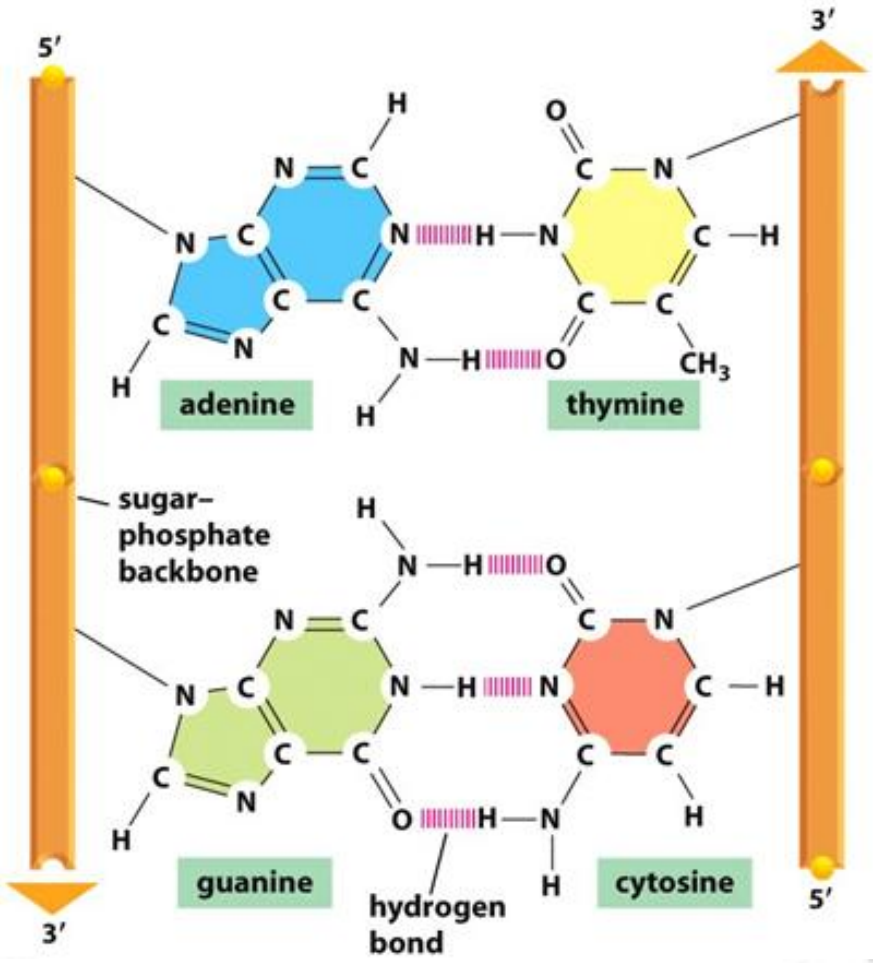
# DNA as a double helix



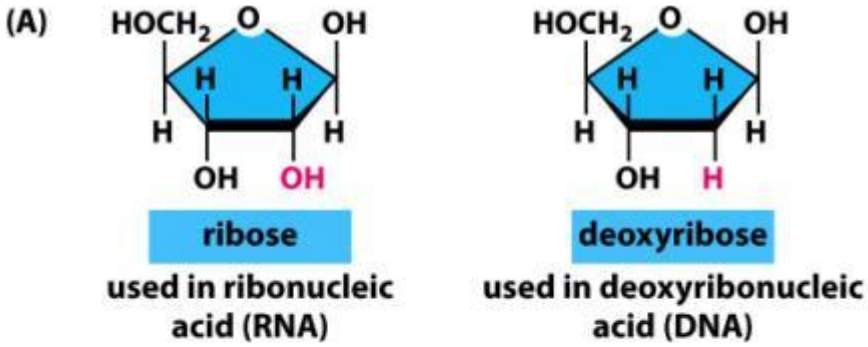
# Complementary base-pairing

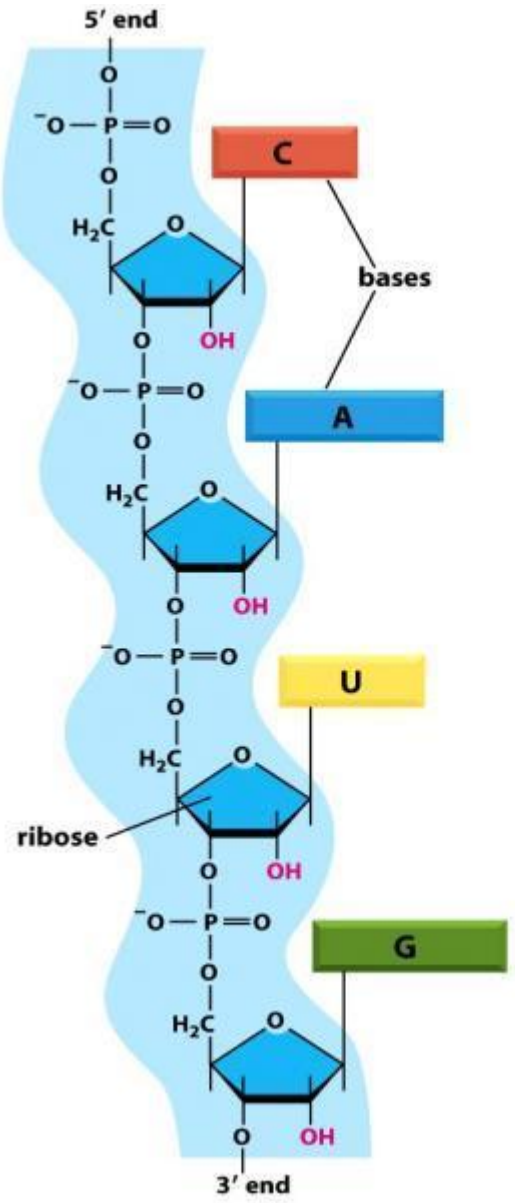
A - T

C - G

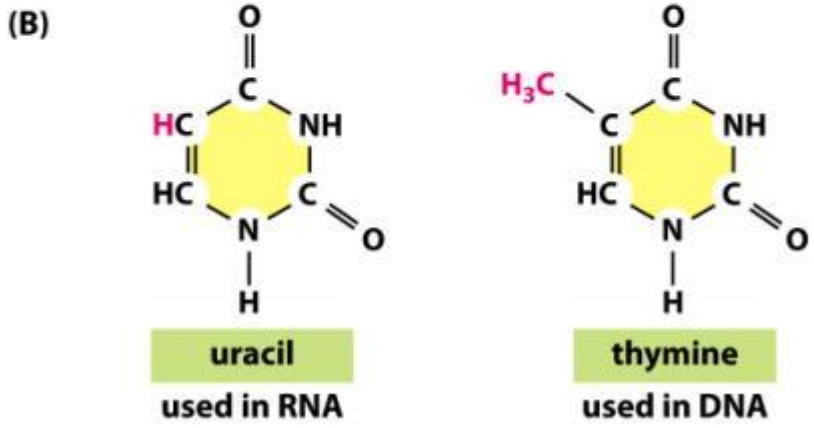


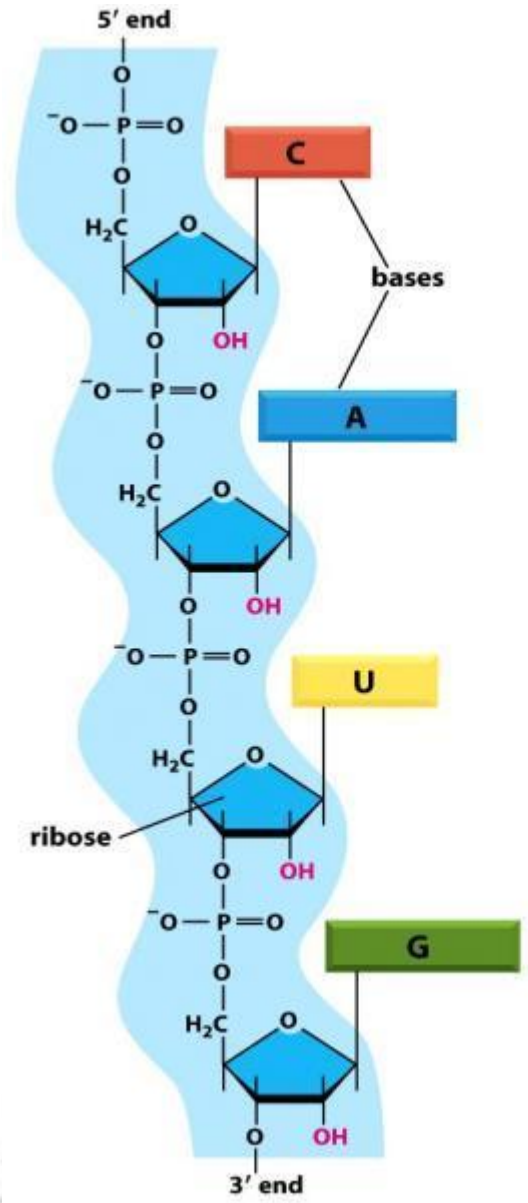
# Ribonucleic acid (RNA)



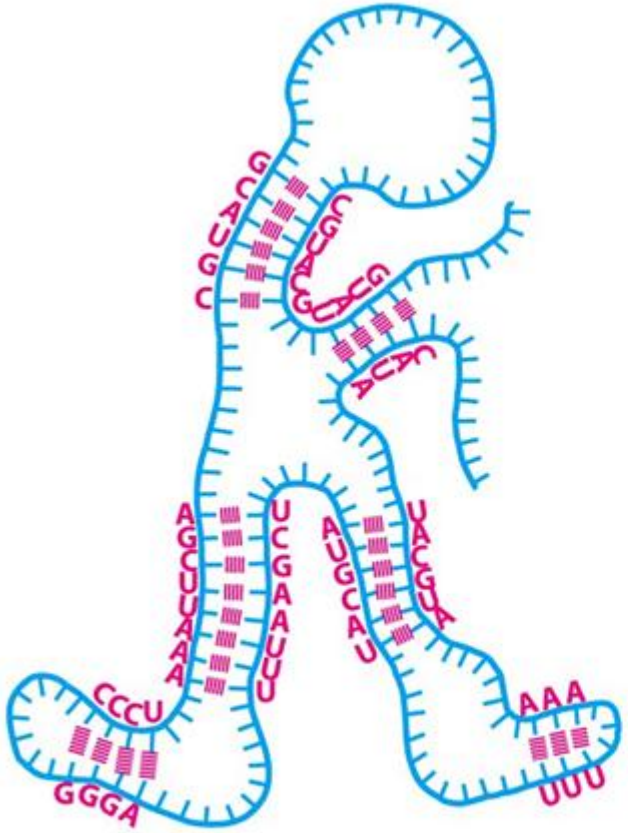



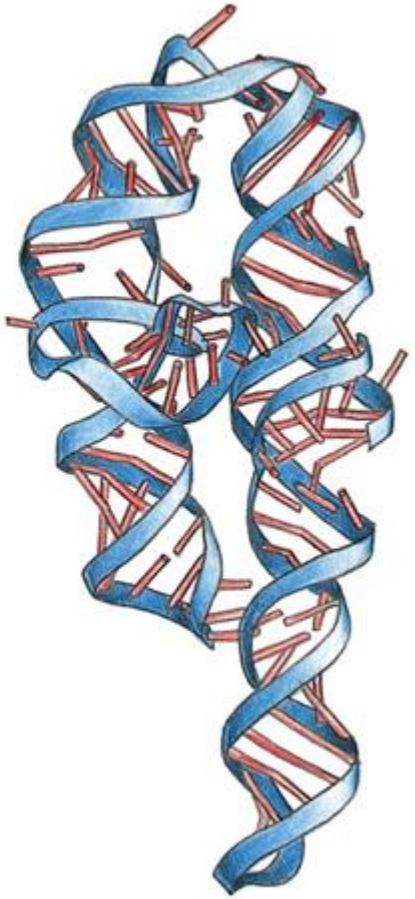
# Ribonucleic acid (RNA)

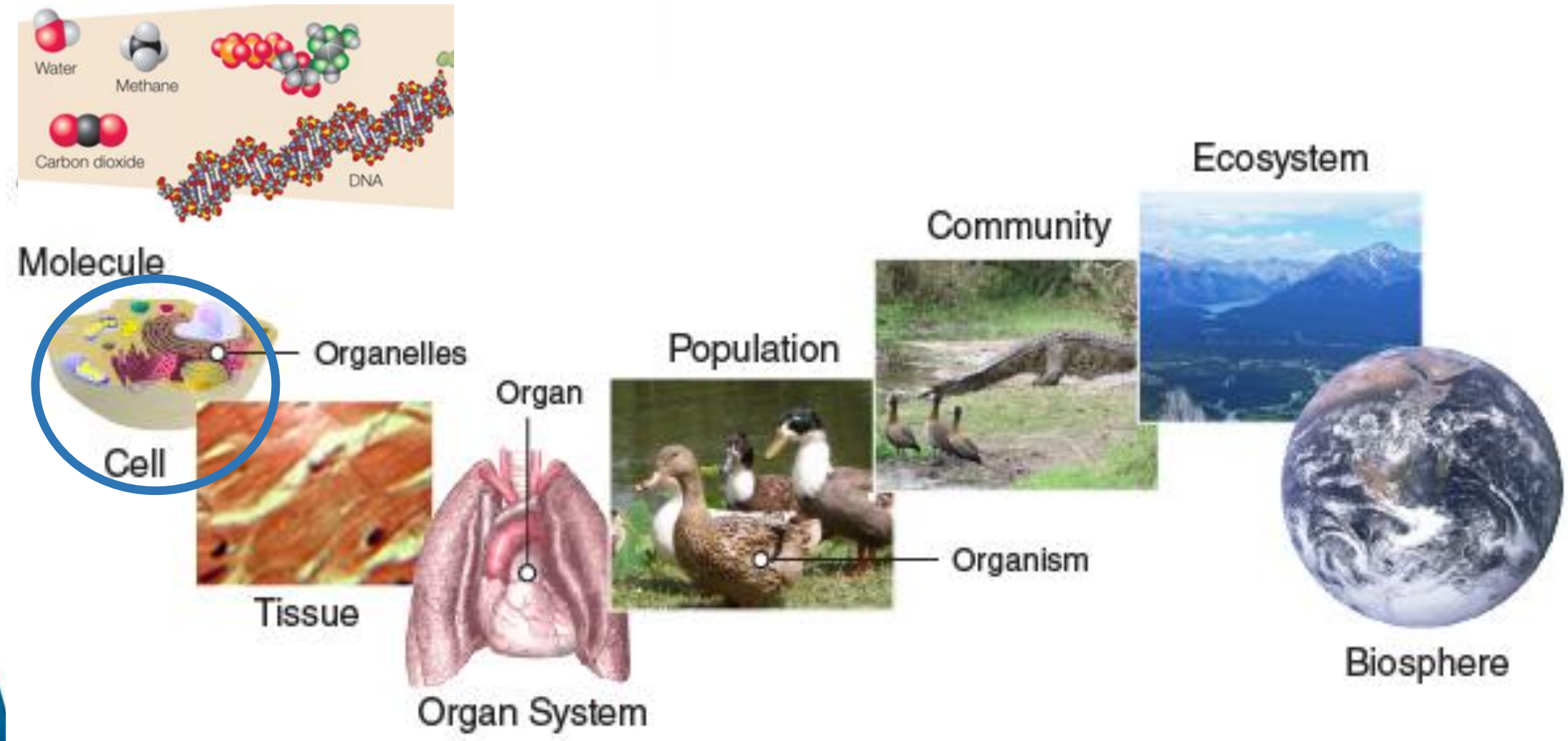


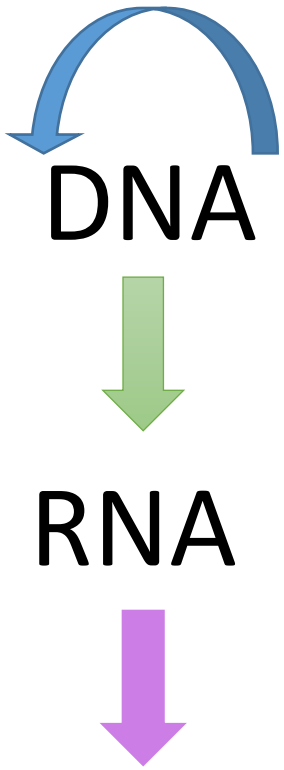
# RNA Structures

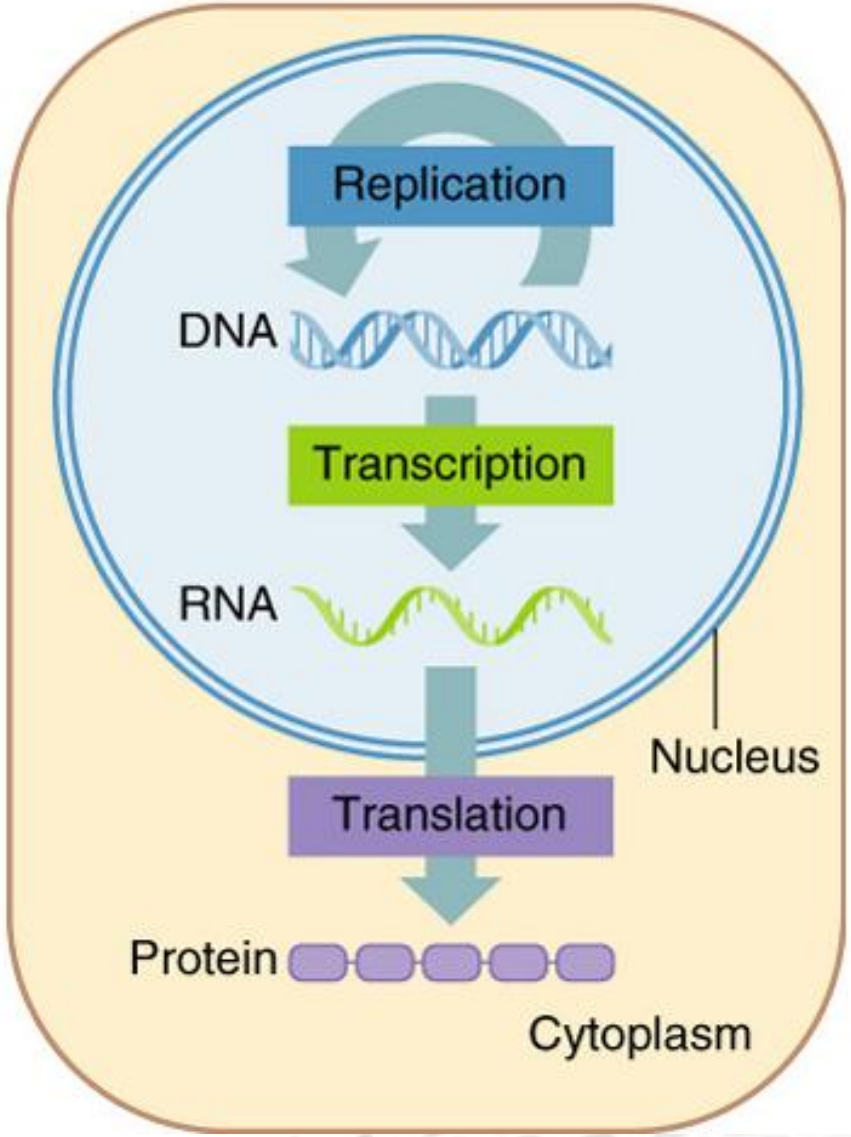



Water
Methane
Carbon dioxide
DNA
Molecule
Organelles
Cell
Tissue
Organ
Organ System
Population
Organism
Community
Ecosystem
Biosphere

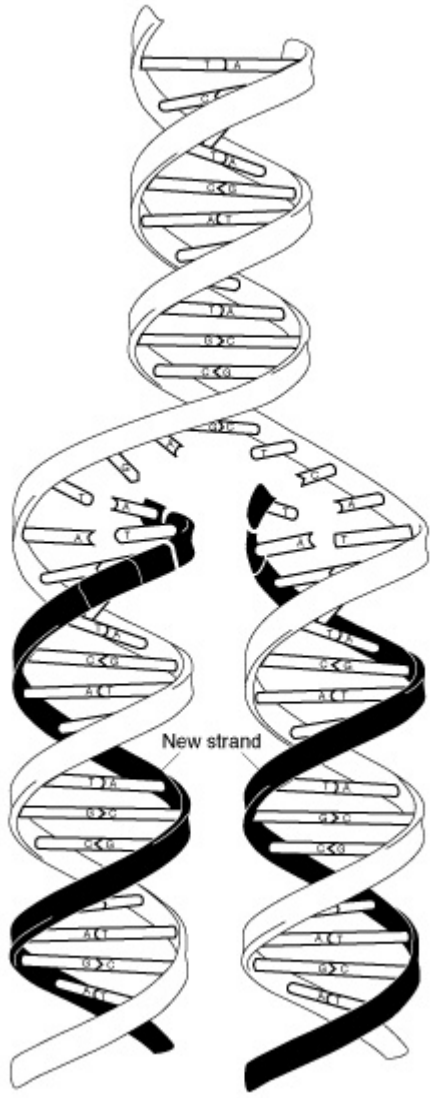
# Central Dogma



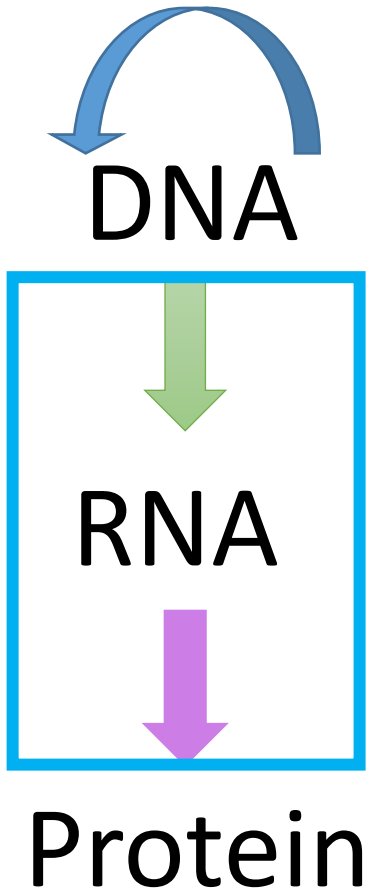



# Replication

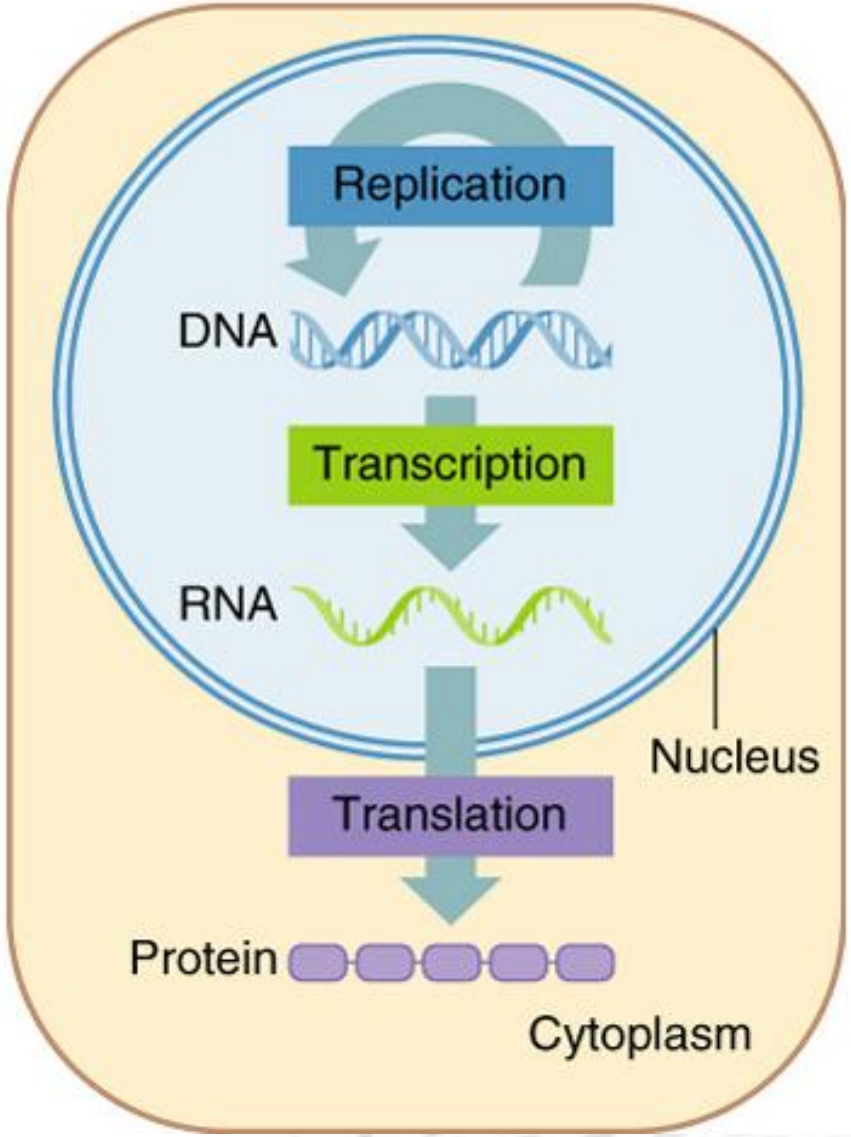
- DNA can make copies of itself
  - Before cell dividing
- Unzipping double helix
  - H-bonds break
  - Each original strand a template
- Adding new nucleotides
  - Complementary base-pairing
  - DNA polymerase



# Central Dogma

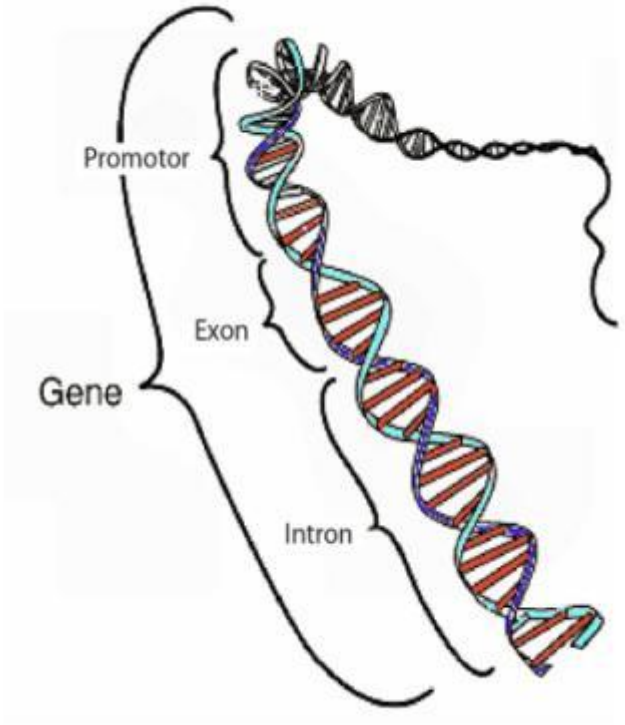



Gene Expression

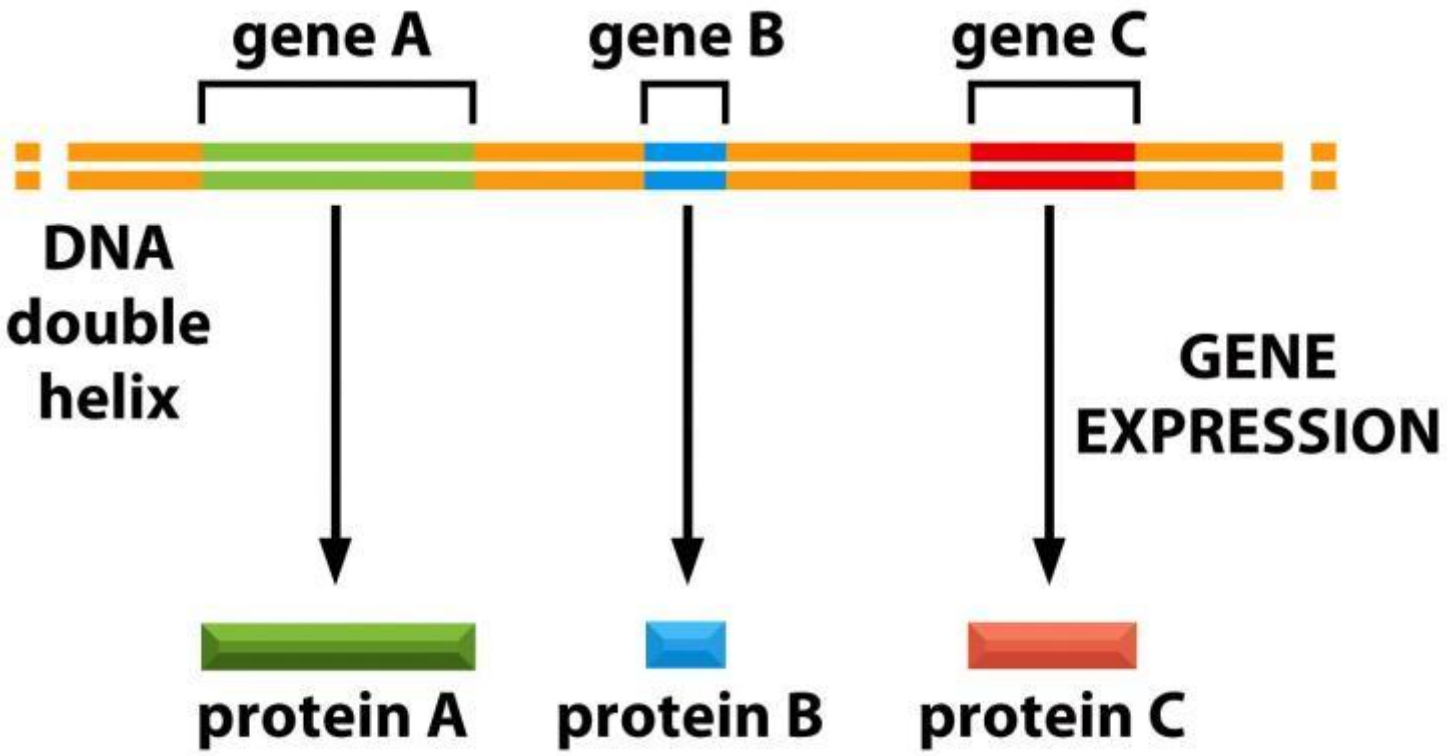


# What are Genes?

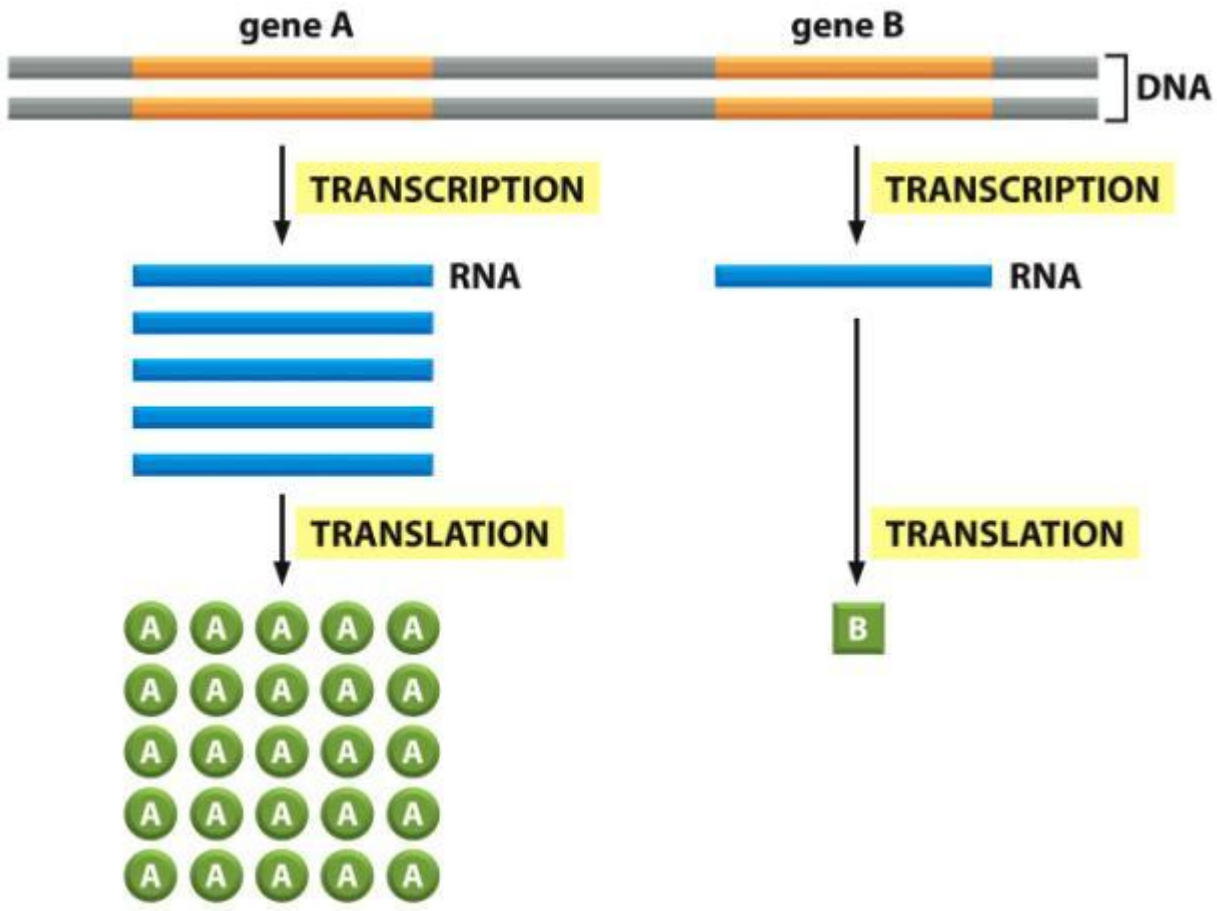
- Genes
  - the tiny sequences in DNA contain information to make proteins
- *Genome*
  - an organism's complete set of DNA, including all of its genes. (genetic material)
  - Each *genome* contains all of the information needed to build and maintain that organism.



# Gene expression



# Gene expression

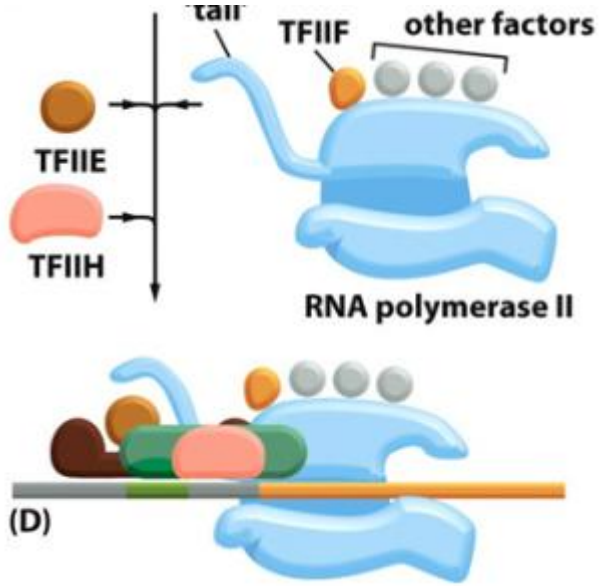


# There are THREE type of RNA

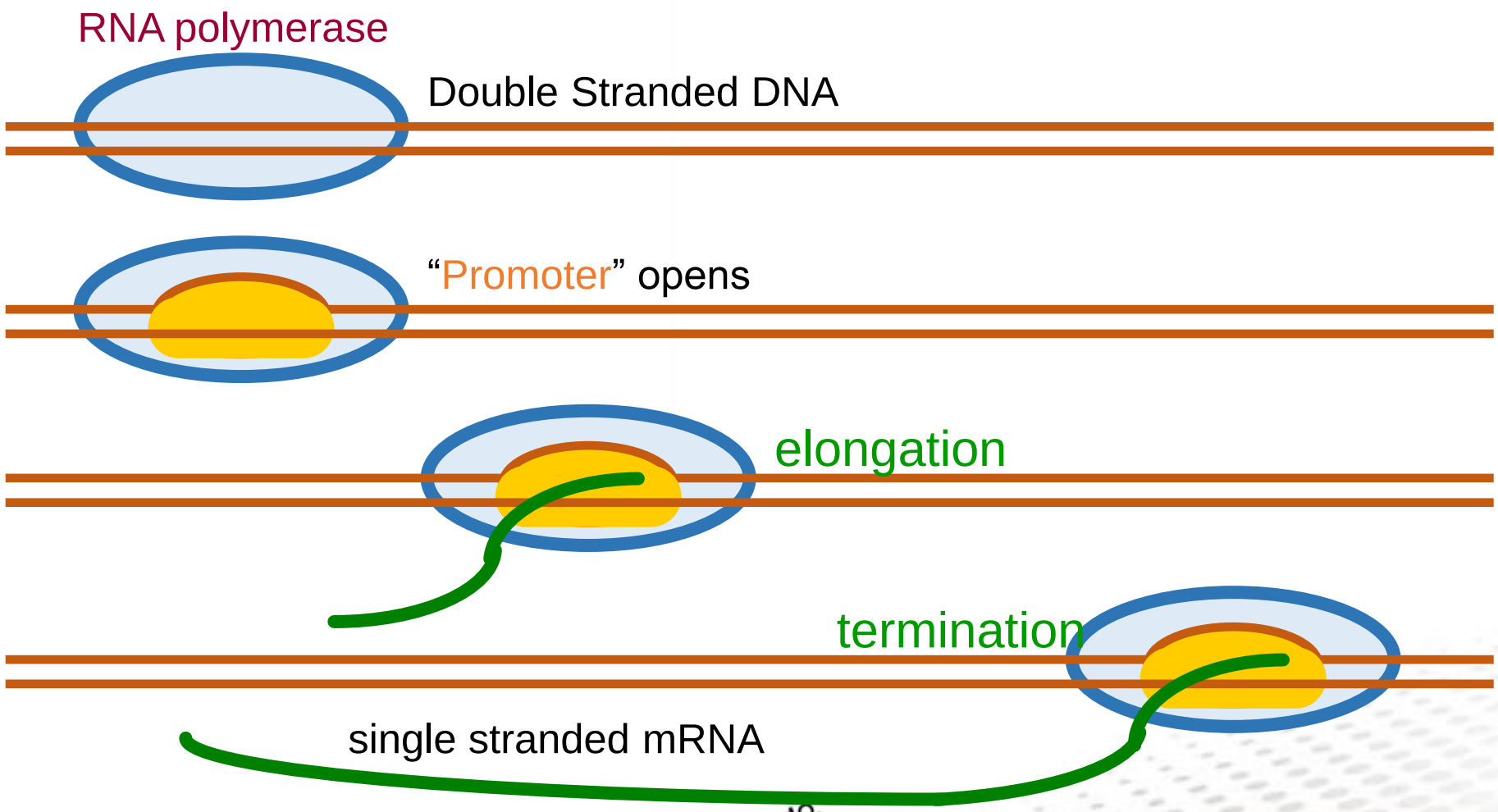
- Messenger RNA (mRNA)
  - Long strands of RNA nucleotides that are formed complementary to one strand of DNA
- Ribosomal RNA (rRNA)
  - Associates with proteins to form ribosomes in the cytoplasm
- Transfer RNA (tRNA)
  - Smaller segments of RNA nucleotides that transport amino acids to the ribosome where proteins are made by adding 1 a.a. at a time

# Transcription (Important Players)

- Promoter
  - DNA site that promotes RNA polymerase to bind
- RNA Polymerase
  - Enzyme that completes process of transcript
- Transcription Factors
  - proteins that attract the RNA polymerase and regulate
- Repressor
  - molecule that binds to DNA to block transcription

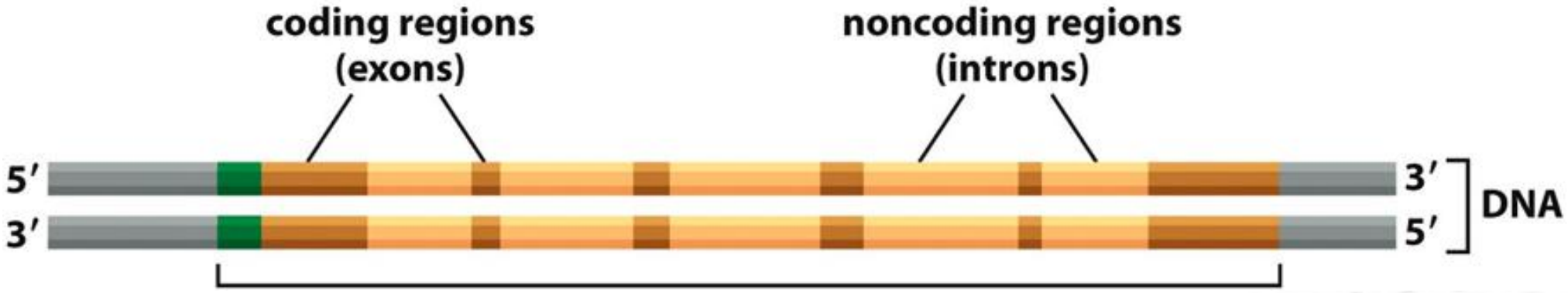


# Transcription

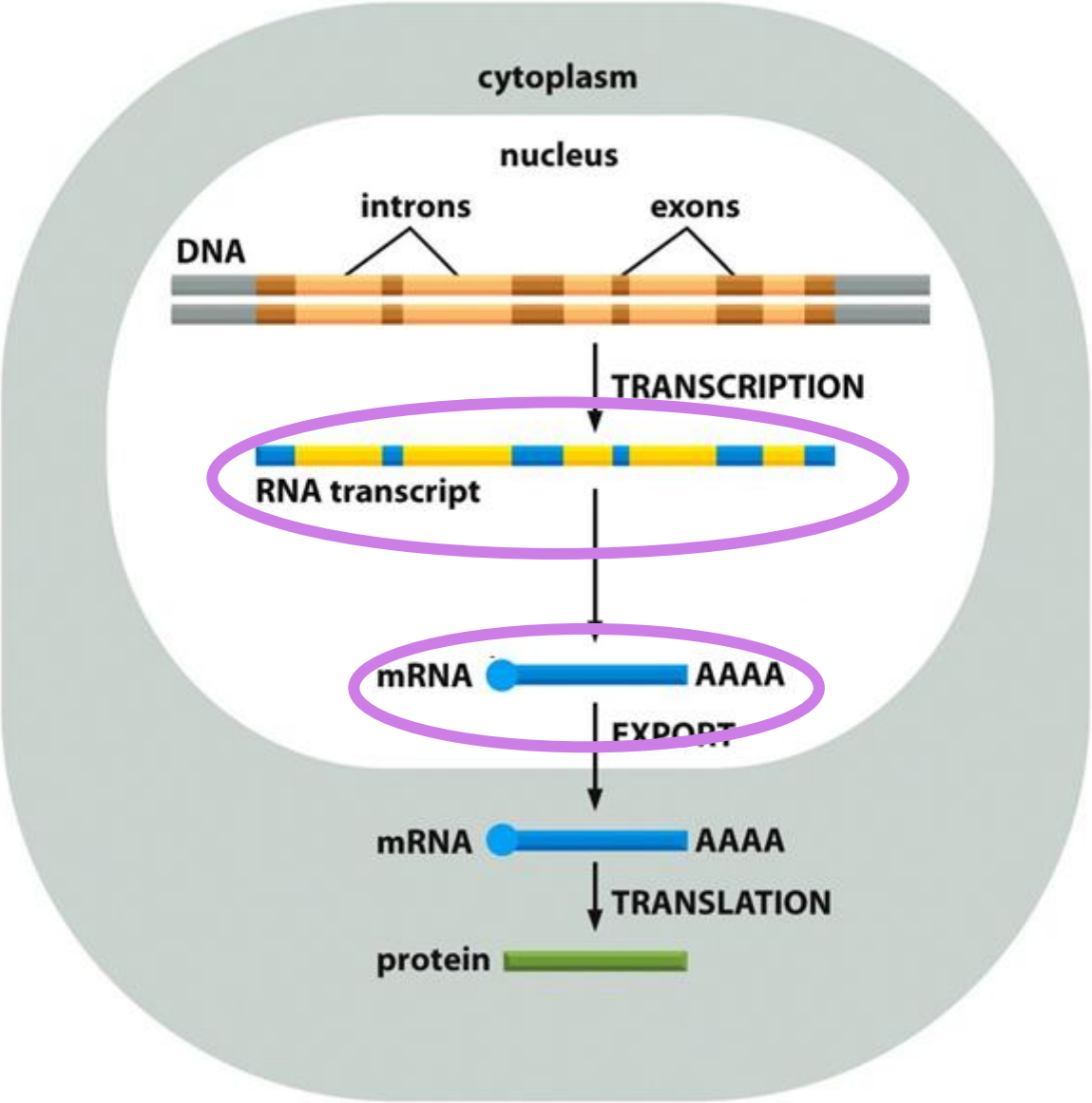


# Processing mRNA

- Splicing out of introns
  - Introns are removed at splice sites
  - Leaving only exons for translation



# mRNA Splicing



# Translation

...AGAGCGGAATGGCAGAGTGGCTAAGCATGTCGTGATCGAATAAA...

↓

AGAGCGGA.AUG.GCA.GAG.UGG.CUA.AGC.AUG.UCG.UGA.UCGAAUAAA

↓

M.A.G.T.L.S.M.S.STOP

| 4 Nucleotides | 20 amino acids |
|---|---|

1 base codon - $4^1 = 4$ possible amino acids

2 base codon - $4^2 = 16$ possible amino acids

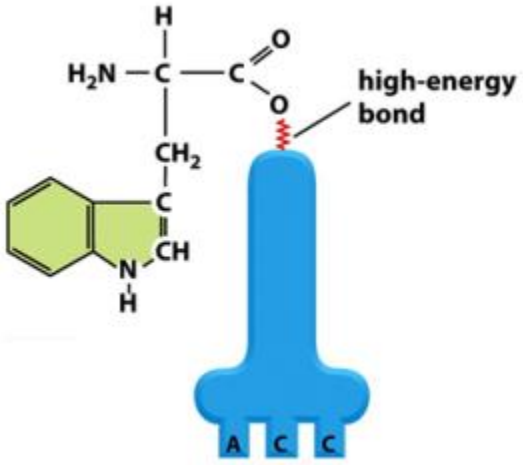
3 base codon - $4^3 = 64$ possible amino acids

# The Genetic code

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

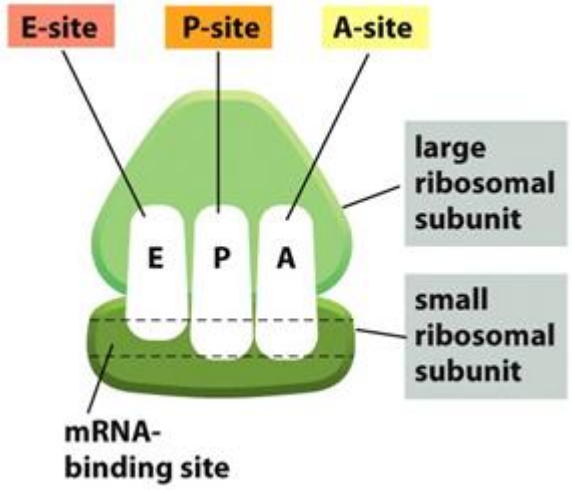
# Translation (Important Players)

- tRNA (transfer RNA)
  - Binds codon on one side and amino acid on the otherside
- Ribosome
  - enzyme that gathers the correct tRNA and makes the peptide bond between two amino acids
- Stop codons
  - stop translation

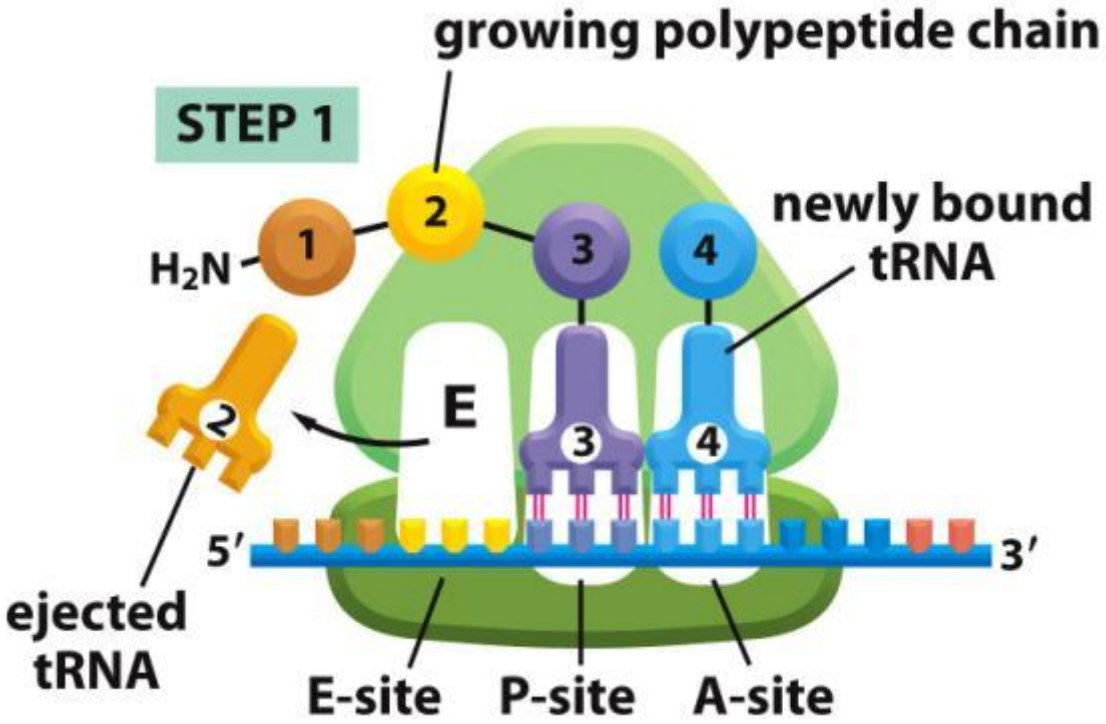


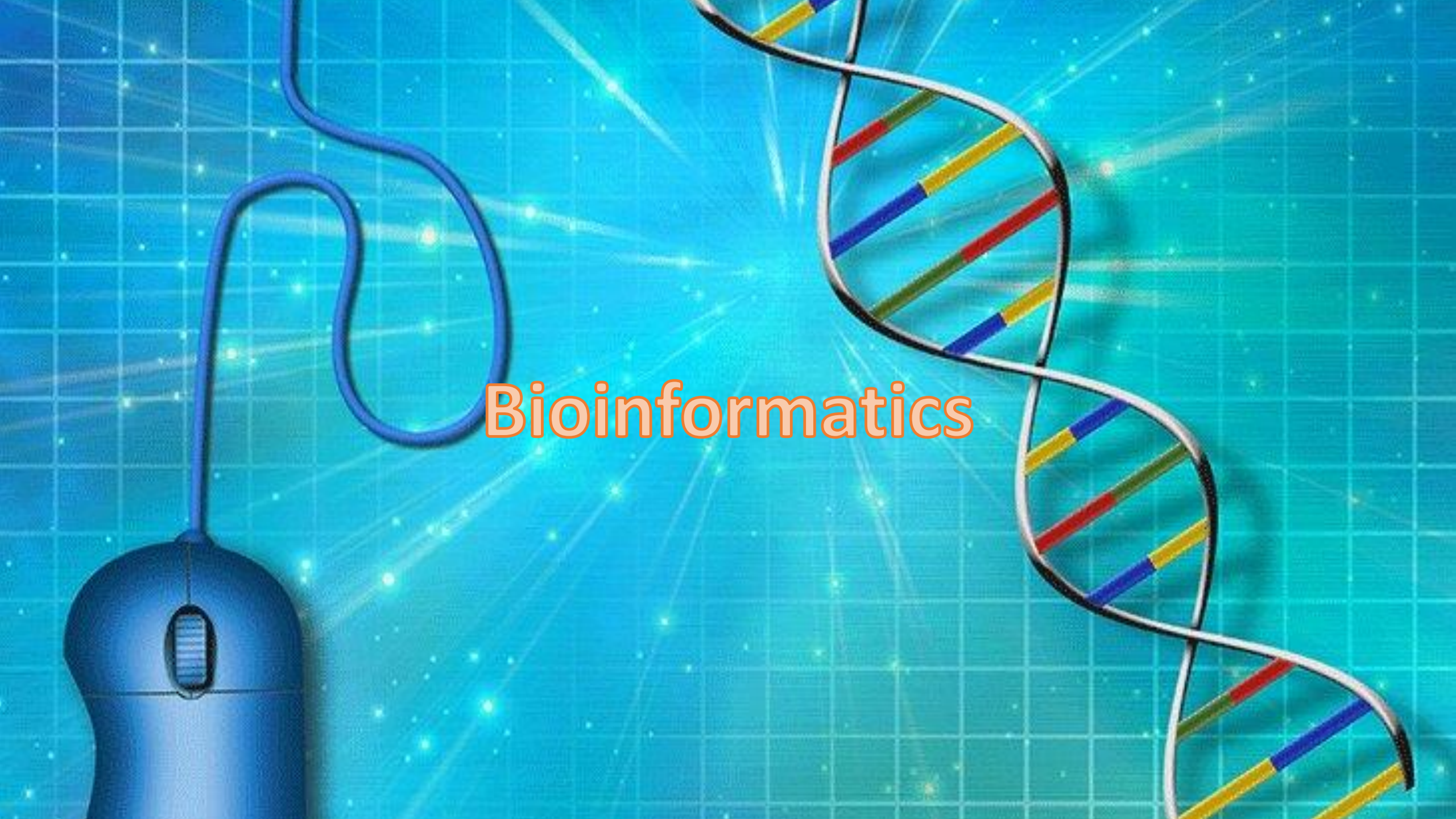

UAA
UAG
UGA

stop



# Protein synthesis




Bioinformatics

# A bioinformatics problem

- *Sequence Alignment*
  - identify regions of similarity between biological sequences (protein or nucleic acid)
  - similarity may indicate relationships
    - functional
    - structural
    - evolutionary

# Sequence alignment is important for:

* prediction of function

* database searching

* gene finding

* sequence assembly

# Problem Definition

- The problem of finding a maximal level of identity between two sequences by lining them up.

  - The sequences are padded with **gaps** (dashes) so that wherever possible, columns contain **identical characters** from the sequences involved

***DNA-sequence-1***

```
tcctctgcctctgccatcat---caaccccaaagt
|||| ||| ||||| |||||   ||||||||||||
tcctgtgcatctgcaatcatgggcaaccccaaagt
```

***DNA-sequence-2***

# Alignment vs. LCS

- Longest Common subsequence (LCS)
  - A classic problem in CS
- Alignment
  - An old problem in Bioinformatics
  - Needleman and Wunsch (1970)
- Difference:
  - Scoring is biologically inspired in Alignment