Introduction to Bioinformatics – COMP 402

BY DR. MOHAMMAD HASHIM

Lect. 1

Course Description

Course Code/Title:	COMP 402 Introduction to Bioinformatics
Level/Semester	**4 / Spring (2024-2025)**
• #credit hours • #hours/Lecture	**3** **3 (Sunday 8:00 - 11:00)**
Lecturer	Dr. Mohammad Fakhry & Dr. Mohammad Hashim
Contact	m_abdelrahman@sci.asu.edu.eg

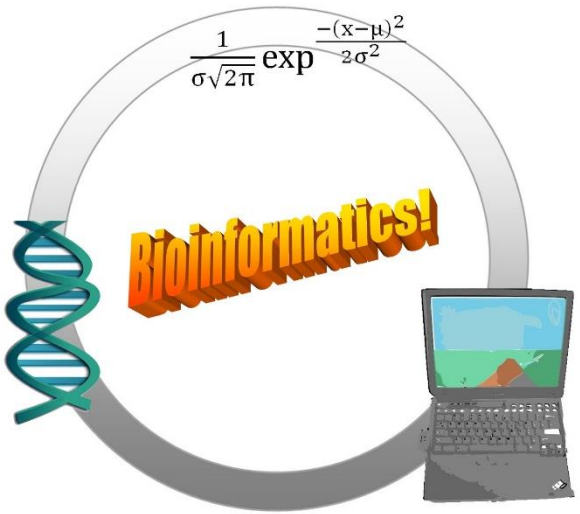
Grading System
• Final exam
• Term -Work
• Oral
 	
• Grades Distribution
  • Midterm exam
  • Assignments
	Total is 150
	105
	37
	8
	
	45 (in total) will be distributed as follows:
	• ?? Marks (at the 9th Week)
	• ?? Marks

Final Exam Time	**3 Hours**
All Lectures, book, assignments, and sheets will be on this link	https://sciasuedu-my.sharepoint.com/:f:/g/personal/m_abdelrahman_sci_asu_edu_eg/Eoa6uXwhjw1IiS5G4JZJAagB5ho-adXUN2d-Ffsjb_RlxQ?e=331ZF6

Outlines

1. What is Bioinformatics?
2. Biological Background

What is Bioinformatics?

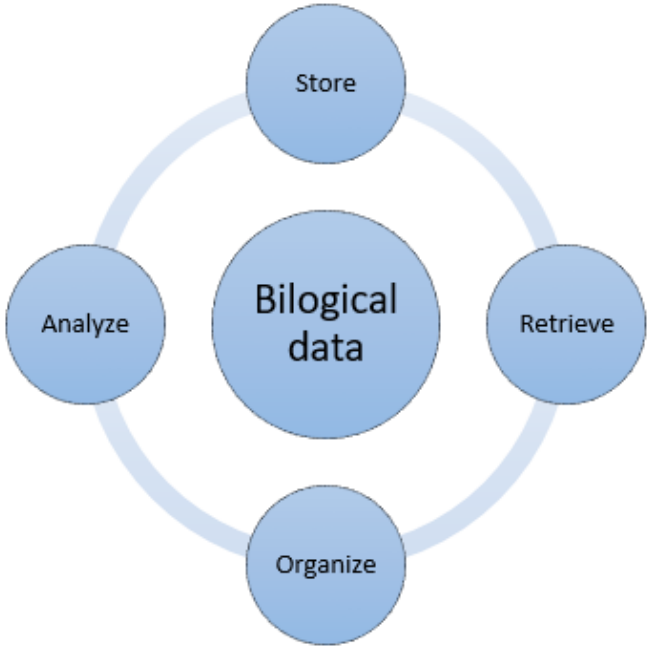


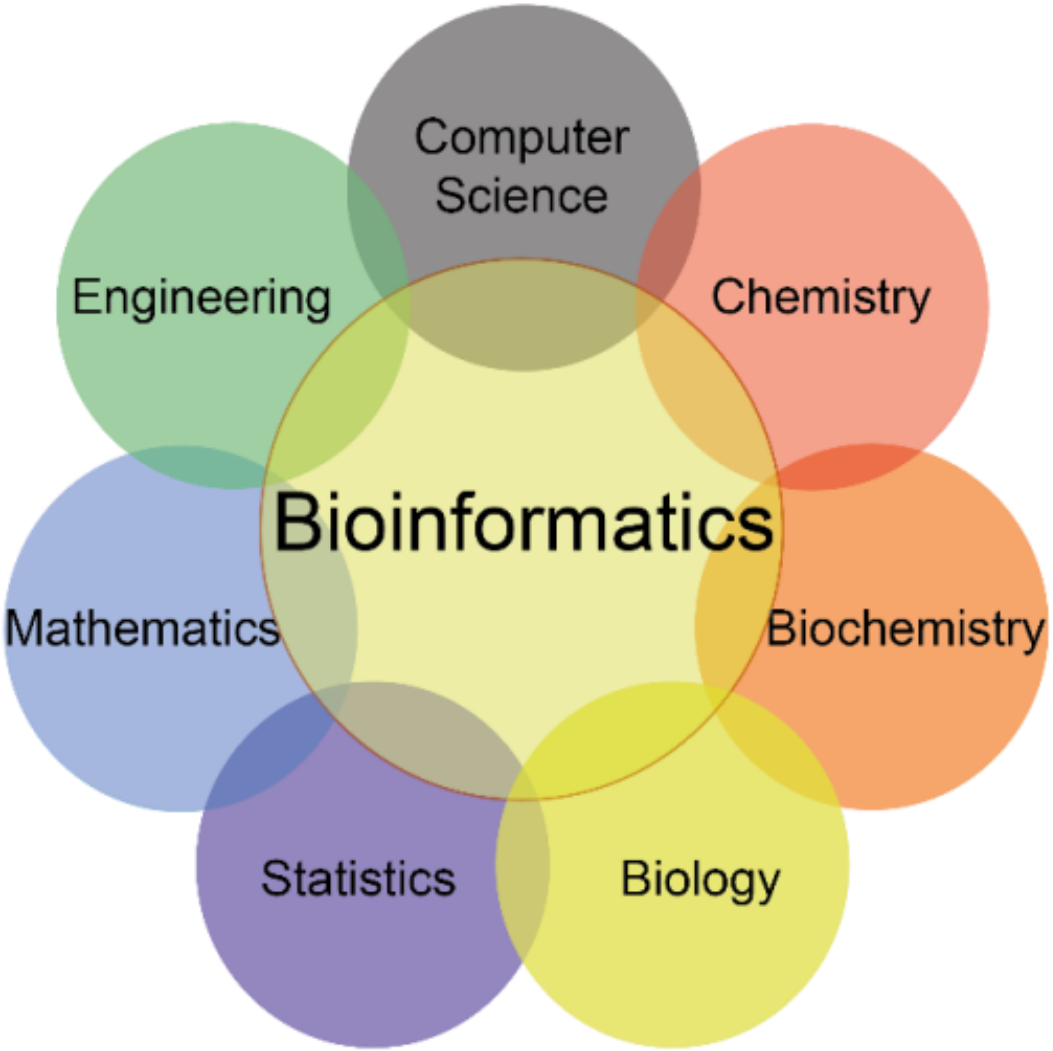
The US National Institutes of Health (**NIH**) defined it as follows:

Bioinformatics : Research, development, or application of computational tools and approaches for expanding the use of biological, medical, behavioral or health data, including those to acquire, store, organize, archive, analyze, or visualize such data.

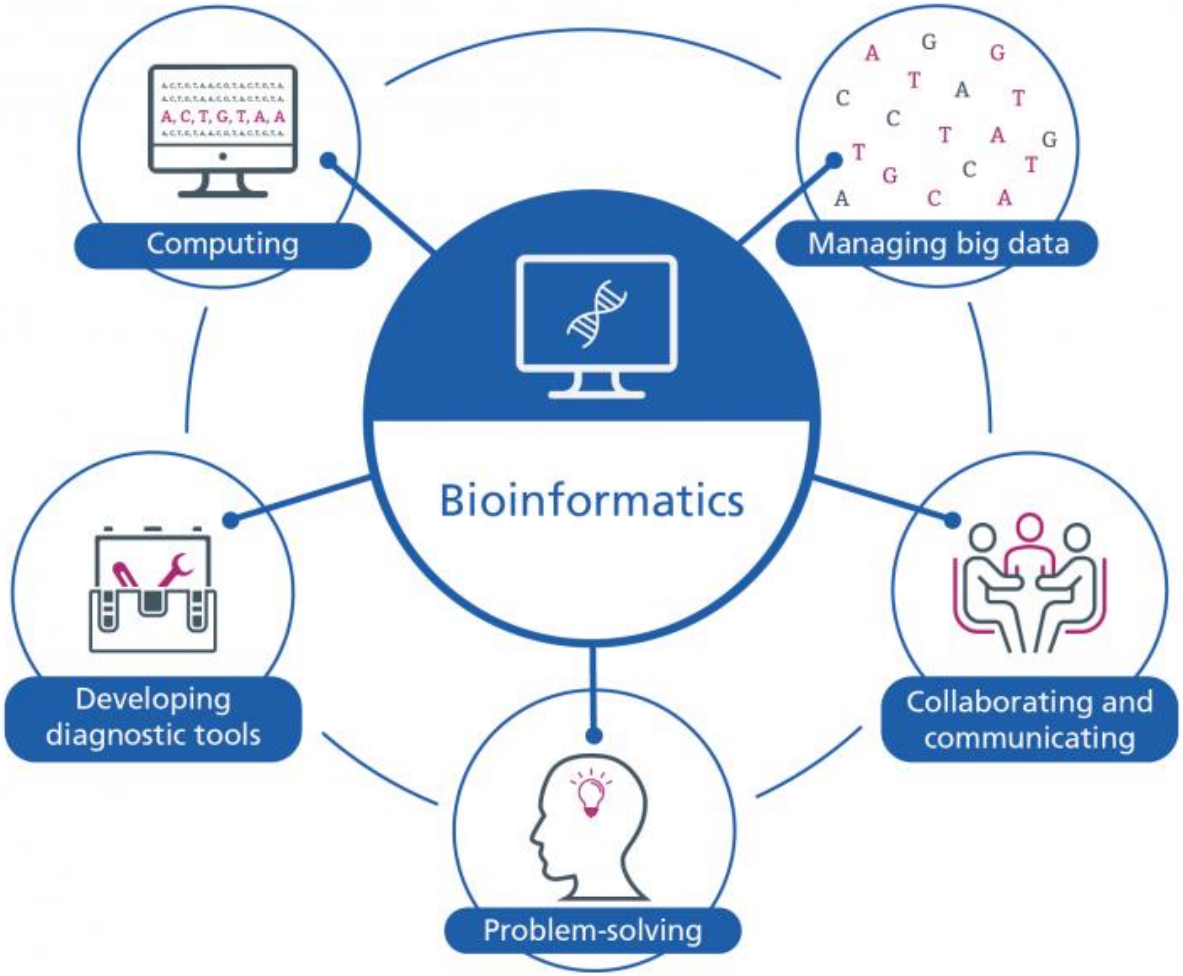
What is Bioinformatics?

Simply, Bioinformatics is any activity that deals with biological data using computational tools.





Bioinformatics is a Multidisciplinary Field



Needs for Bioinformatics

Needs for Bioinformatics

Data provide the foundation on which scientific knowledge is constructed.

Usual concern voiced by scientists is that they have too few data.

Data to be useful, they must be in a form that researchers can work with and make sense of, and this can become harder to do as the amount of data grows.

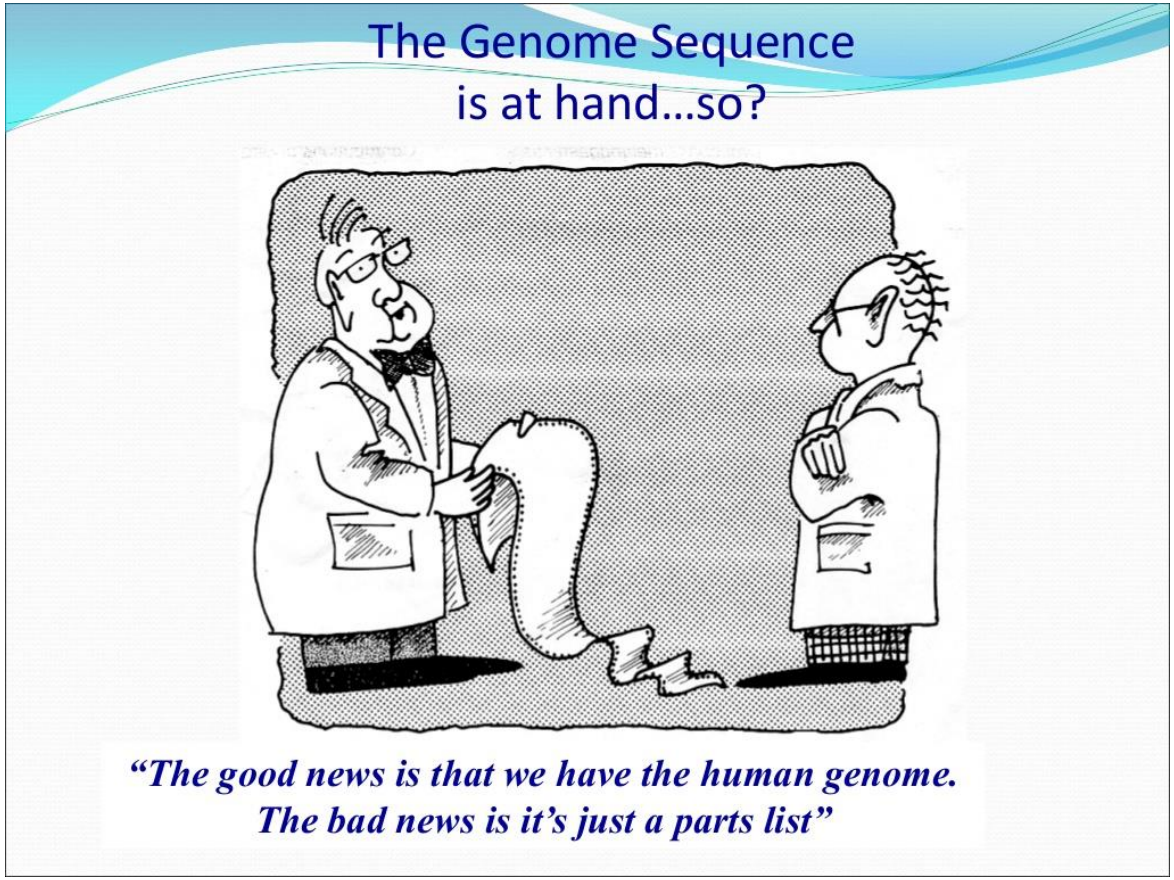
Needs for Bioinformatics

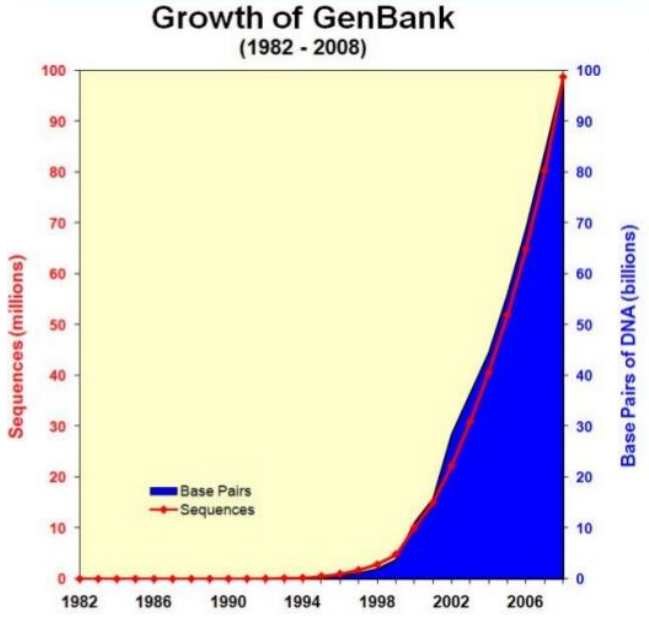
Efficient biotechnological methods for gathering biological data makes the amount of available data currently grows in a very fast rate.

"***An experimental laboratory can produce over 100 gigabytes of data a day with ease***".

"***Biological data is doubled every 15 month***".

Needs for Bioinformatics



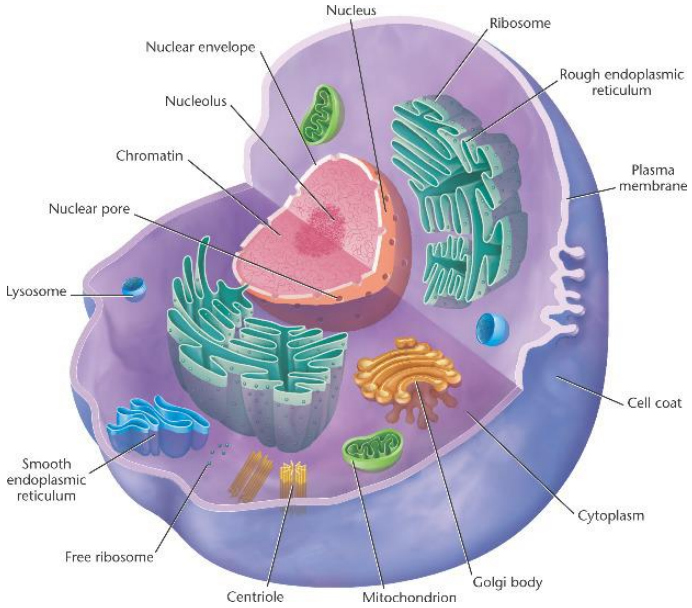


Goals of Bioinformatics

Understand a living cell and how it functions at molecular level.

Develop **databases** and **computational tools**.

Tools are used to **mine** (analyze) **databases** to generate knowledge to better understand the living systems.



Expected Benefits of Bioinformatics

- Molecular Medicine:
  - Improve diagnosis of disease
  - Detect genetic tendency to disease
  - Create drug based on molecular information
  - Use gene therapy and control systems as drugs
  - Design “custom drugs” (pharmacogenomics) based on individual genetic profiles

Expected Benefits of Bioinformatics

- DNA Identification (Forensics)
  - Identify potential suspects whose DNA may matc evidence left at crime scenes.
  - Exonerate persons wrongly accused of crimes.
  - Establish paternity and other family relationships.
  - Match organ donors with recipients in transplant programs.

Expected Benefits of Bioinformatics

- Agriculture, Livestock Breeding and Bioprocessing
  - Grow diseases-, insect- and drought-resistant crops
  - Breed healthier, more productive, disease-resistant farm animals.
  - Grow more nutritious products.
  - Develop biopesticides.
  - Incorporate edible vaccines into food products.

Steps to achive the goals of Bioinformatics

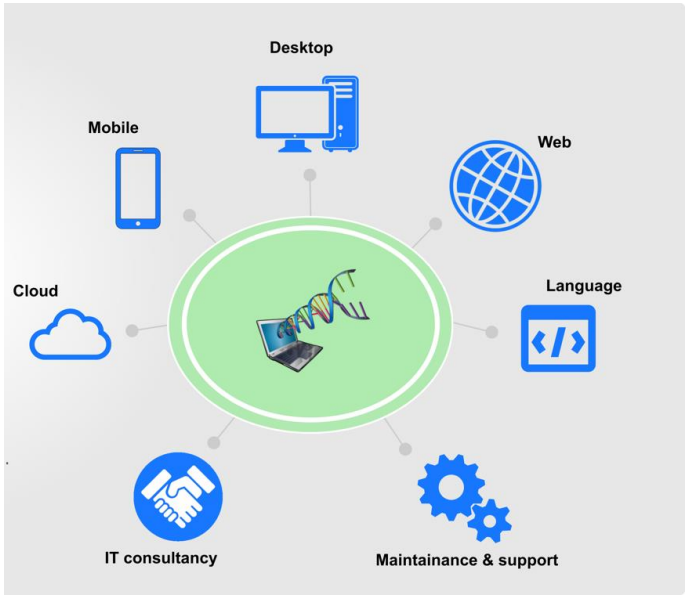
The complete understanding of an organism given its genomic data

Organizes data to be easily accessible (GeneBank).

Develops tools aid in the analysis of data (BLAST).

Interpret the results in a biologically meaningful manner.

CS Trends in Bioinformatics



Algorithmic

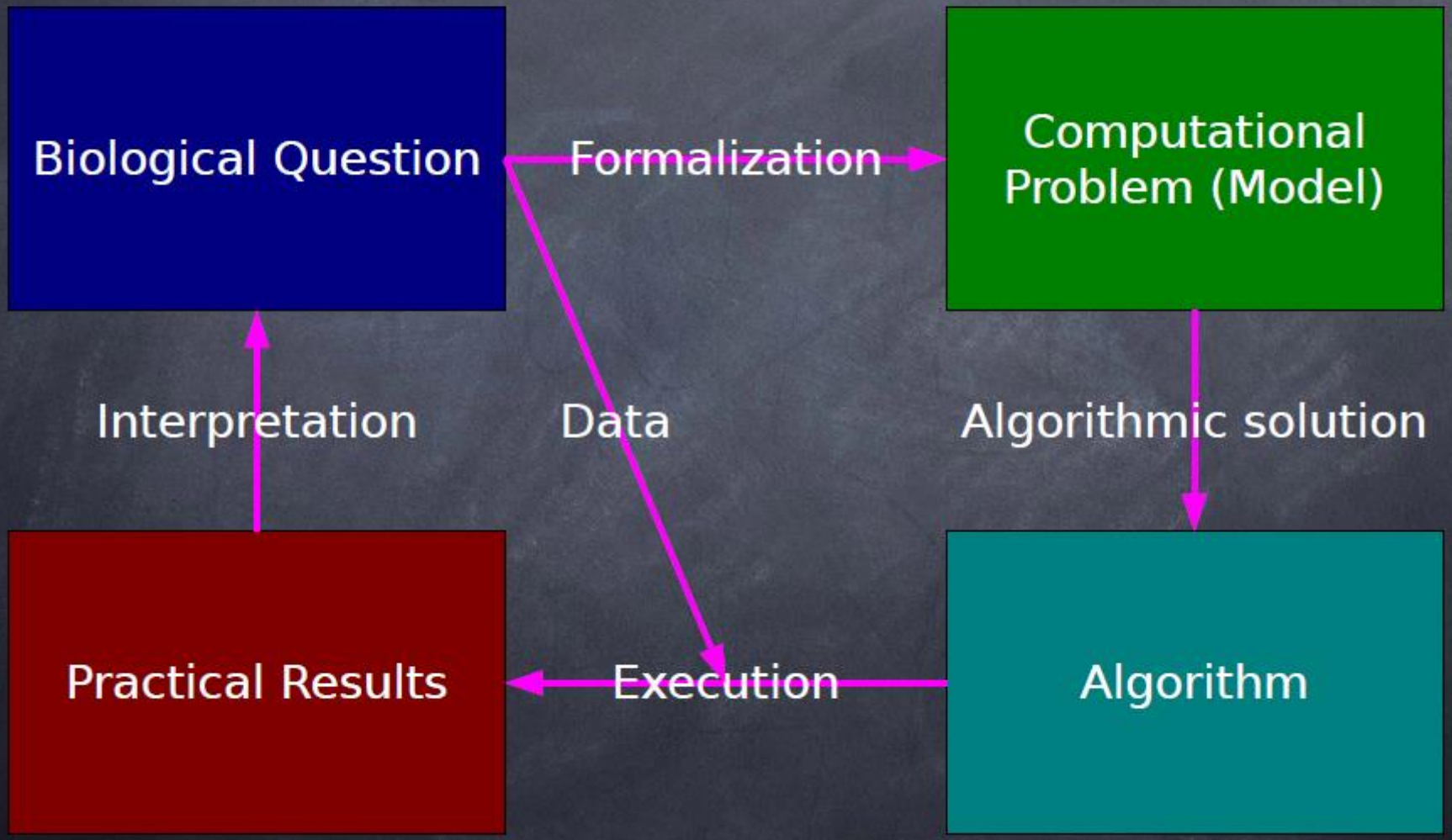
AI Techniques

Database and Data Mining

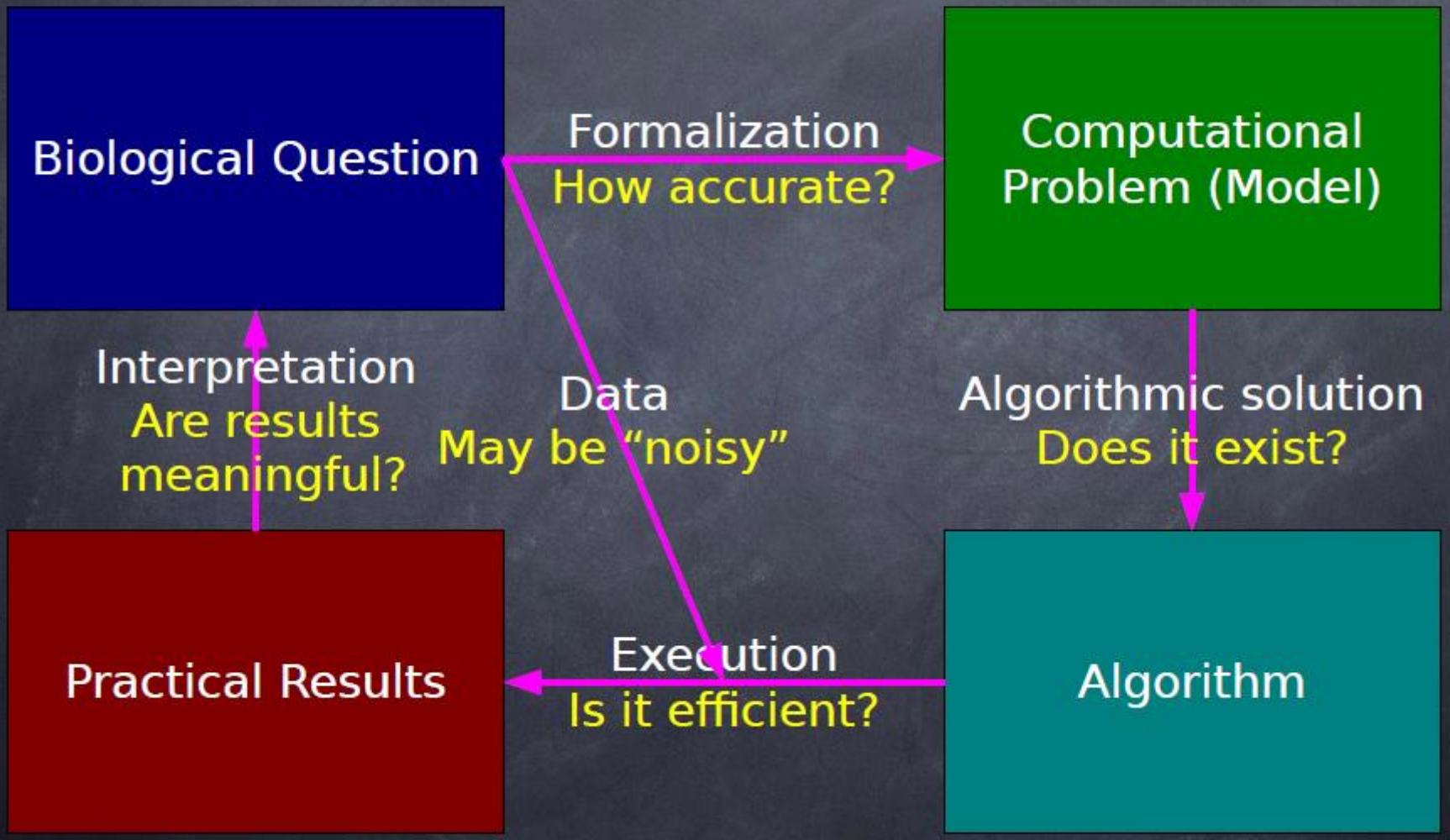
Computer Graphics

Mathematical Modeling …

Bioinformatics Workflow



Bioinformatics Bottlenecks



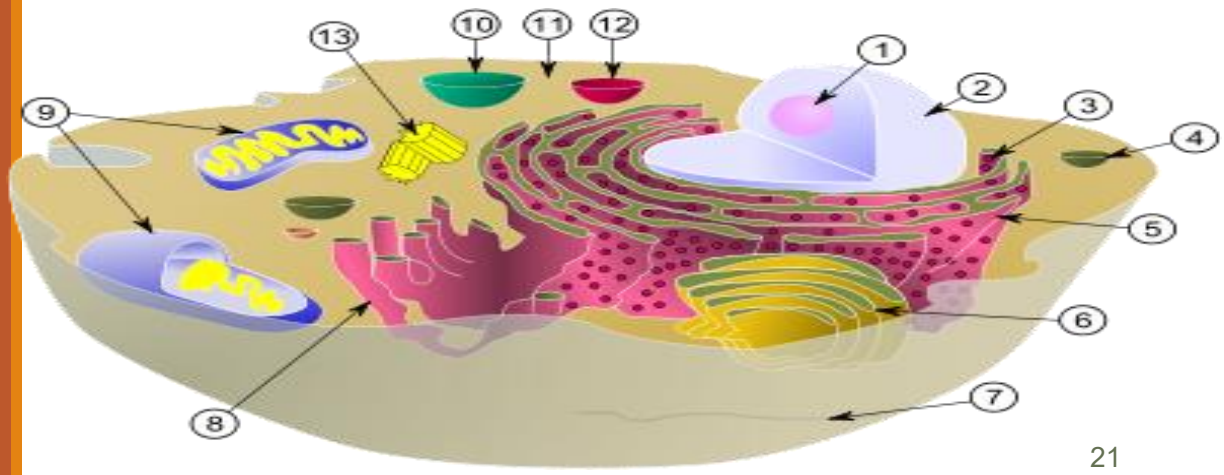
Biological Background

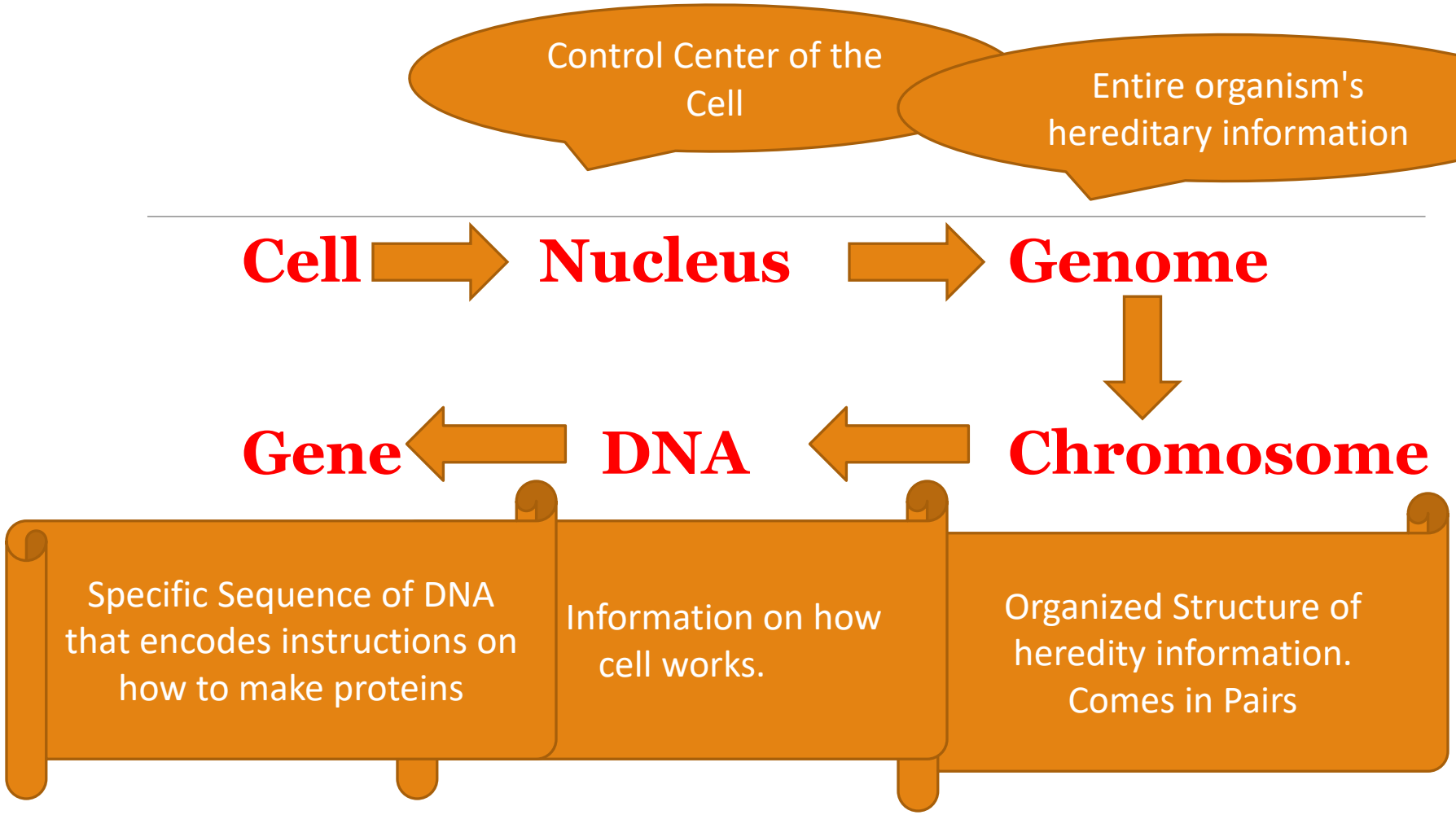
The most detailed model of a human cell to date, obtained using x-rays, nuclear magnetic resonance, and cryoelectron microscopy data sets

Biological Terms

Cell: is the functional unit of all known living organisms. It is the smallest unit of life that is classified as a living thing, and is often called the building block of life.

A typical cell size is 10 μm (10^{-6} meter) and a typical cell mass is 1 nanogram (10^{-9} gram).

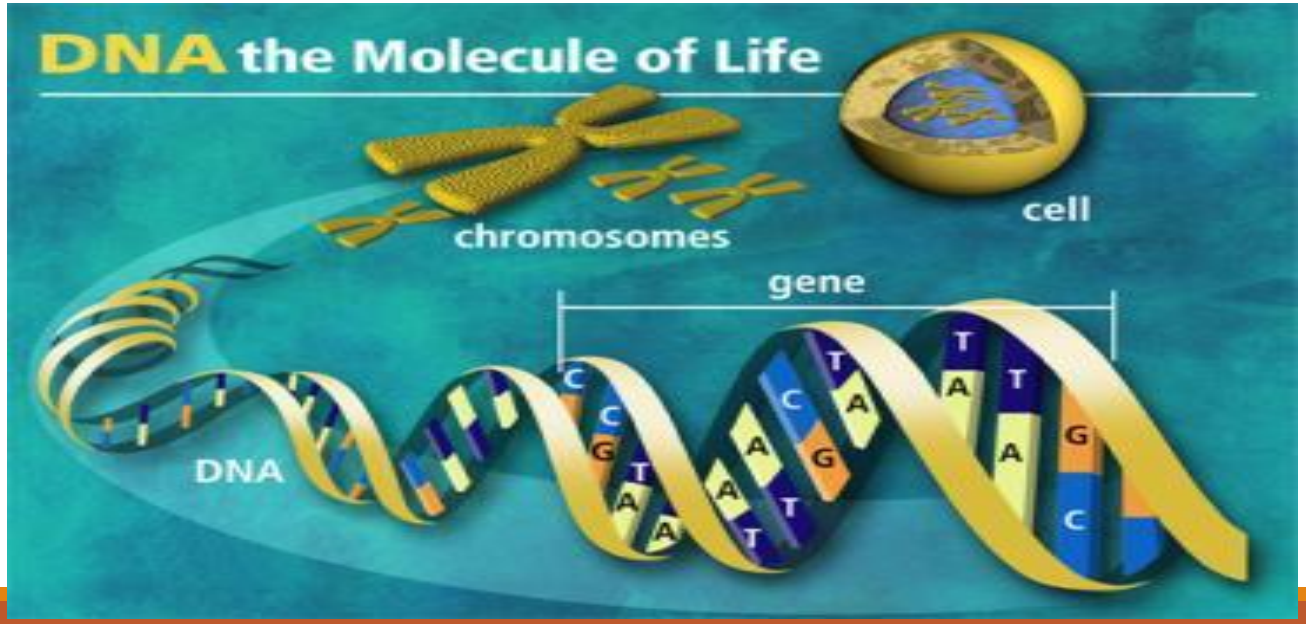



Control Center of the Cell
Entire organism's hereditary information
Cell
Nucleus
Genome
Chromosome
DNA
Gene
Organized Structure of heredity information. Comes in Pairs
Information on how cell works.
Specific Sequence of DNA that encodes instructions on how to make proteins

Cell → Nucleus → Genome

↓

Chromosome → DNA → Gene



Overview of organizations of life

Nucleus = library

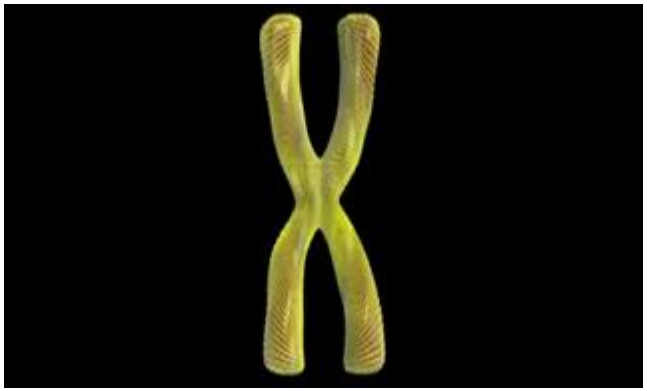
Chromosomes = bookshelves

Genes = books

Almost every cell in an organism contains the same libraries and the same sets of books.

Books represent all the information (DNA) that every cell in the body needs so it can grow and carry out its various functions.

From Chromosome to DNA



The length of DNA in one cell ≈ 2 meters

Total length of DNA in a human ≈ $2*10^{13}$m
≈ 70 trips from the earth to the sun and back.

DNA

DNA: Deoxyribonucleic acid is a nucleic acid that contains the genetic instructions used in the development and functioning of all known living organisms

DNA is a very long polymer.

The basic shape is like a twisted ladder called a double helix.

The **genome** is an organism's complete set of DNA.

- a bacteria contains about 600,000 DNA base pairs
- human and mouse genomes have about 3 billion.

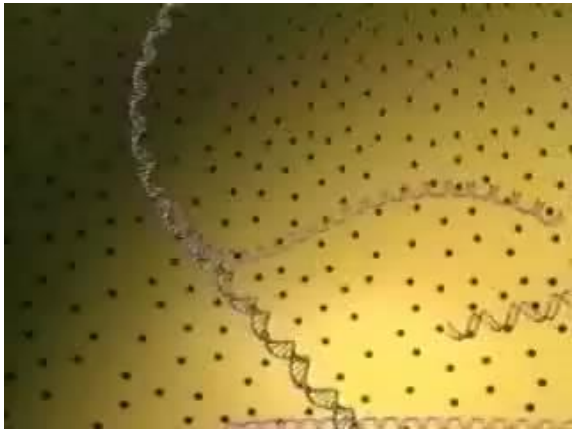
human genome has 23 distinct chromosome pairs.

- Each chromosome contains many **genes**.

Gene

- basic physical and functional units of heredity.
- specific sequences of DNA bases that encode instructions on how to make **proteins**.

DNA Structure



The DNA is a very long string composed of 4 nucleotides (letters)

A (Adenine) ⟺ **T (Thymine)**

C (Cytosine) ⟺ **G (Guanine)**

RNA

RNA: Ribonucleic acid is one of the three major macromolecules that are essential for life.

Looks like DNA, but it is usually a single stranded.

Transfer short pieces of information to different parts of the cell.

Provide templates to synthesize into proteins.

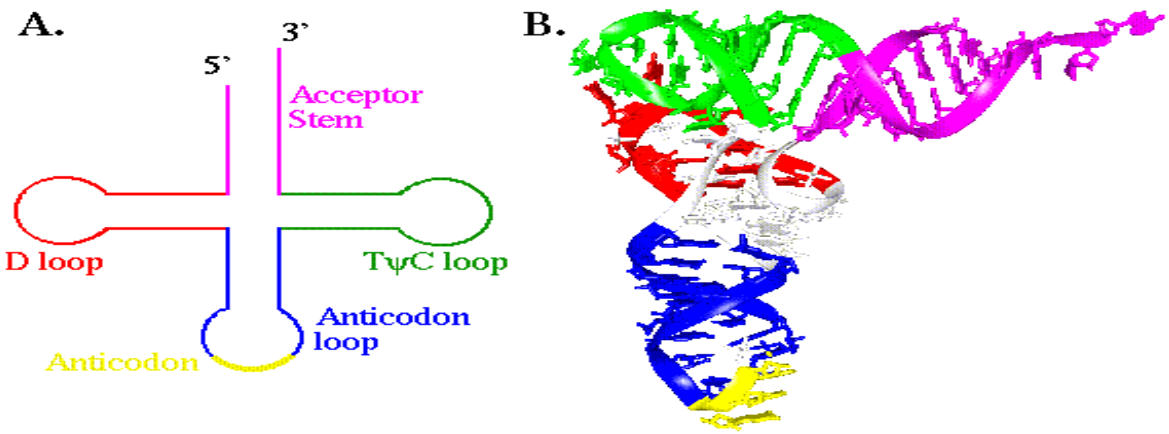
3 types of RNA:

mRNA **tRNA** **rRNA**

RNA structure

- RNA is similar to DNA chemically.
- It is usually only a single strand.
- T(hyamine) is replaced by U(racil)

Some forms of RNA can form secondary structures by "pairing up" with itself. This can have change its properties dramatically.



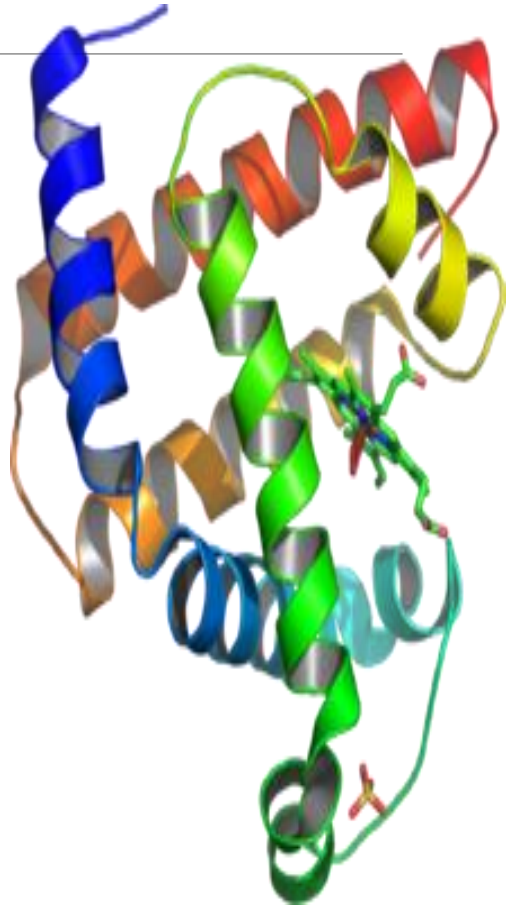
tRNA linear and 3D view: http://www.cgl.ucsf.edu/home/glasfeld/tutorial/trna/trna.gif

Proteins

They are the building blocks and functional components of a cell.

They are the most essential component of life.

Examples for protein are enzymes, hormones, receptors, and structure proteins that form the connective tissue.



Proteins

- Proteins are complex organic molecules made up of amino acid subunits
- There are 20* different kinds of amino acids. Each has a 1 and 3 letter abbreviation.
- Proteins are often enzymes that catalyze reactions.
- Also called “poly-peptides”

*Some other amino acids exist but not in humans.

Protein Synthesis

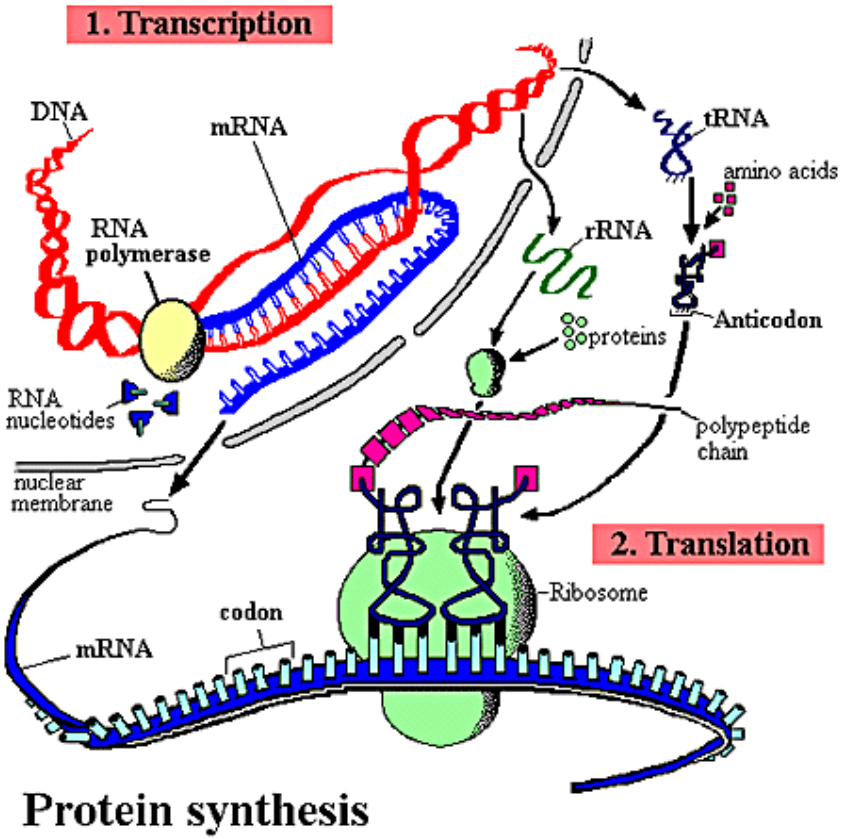
There are twenty amino acids, each coded by three- base-sequences in DNA, called "codons"

- This code is degenerate

The **central dogma** describes how proteins derive from DNA

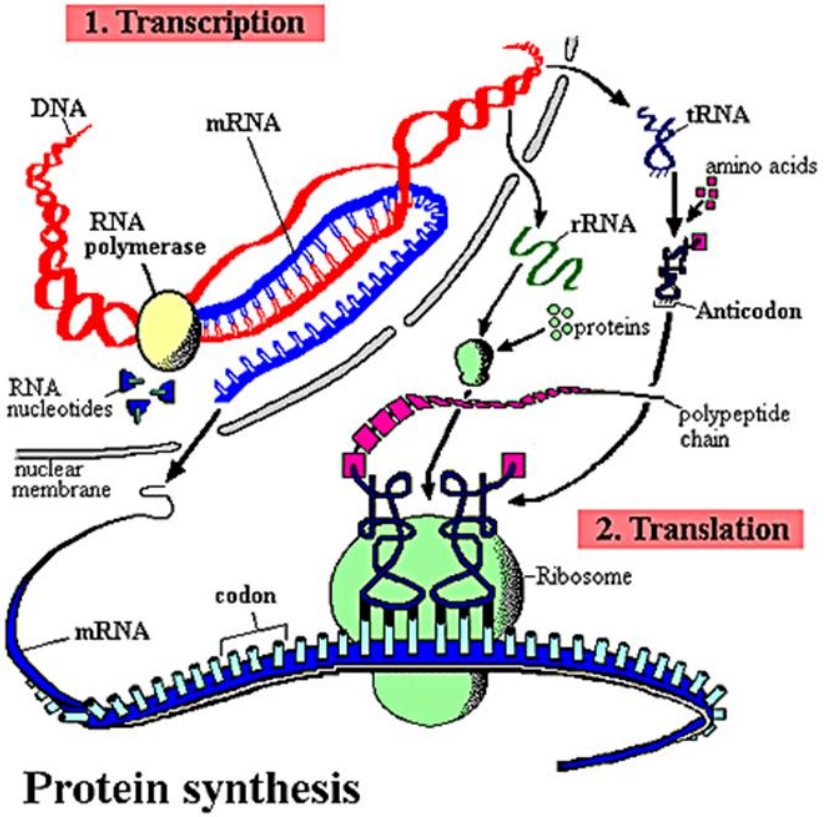
- DNA → mRNA → protein

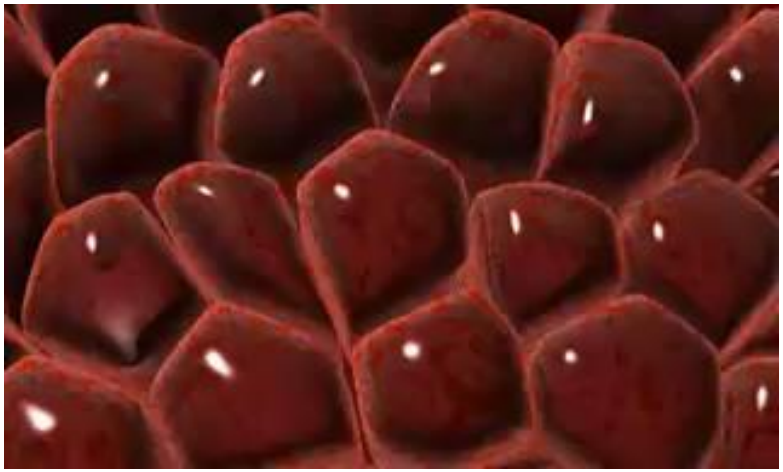
The protein adopts a 3D structure specific to it's amino acid arrangement and function

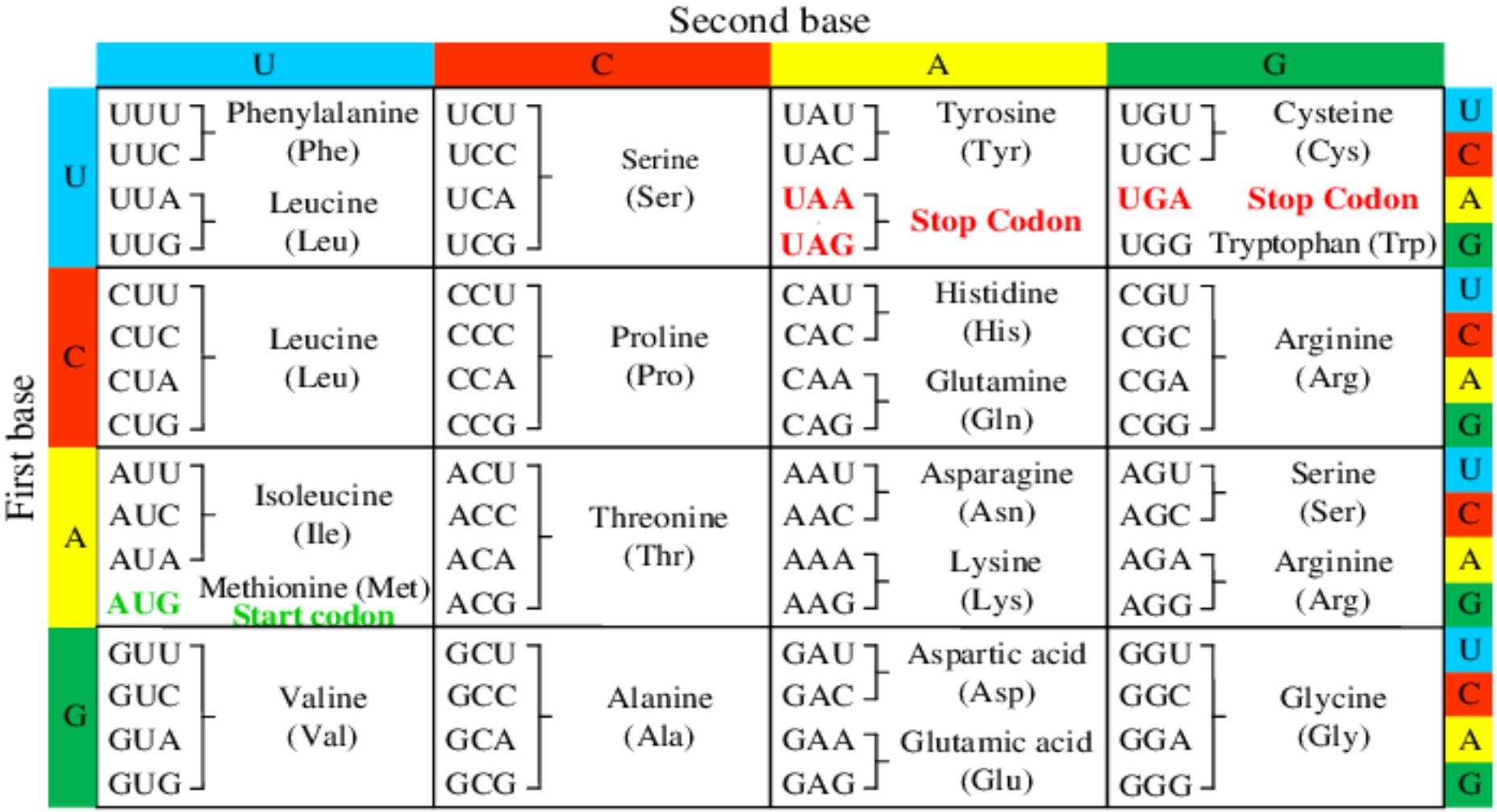


Protein synthesis

Gene Expression

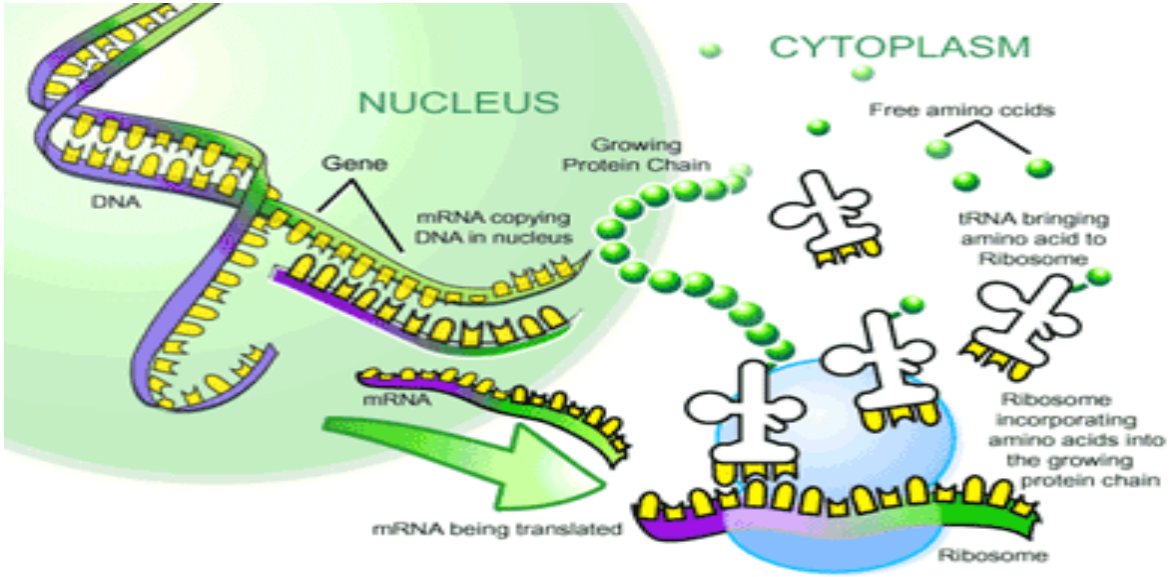






Complete list of Amino Acids chemical structures and abbreviations.

Gene Expression

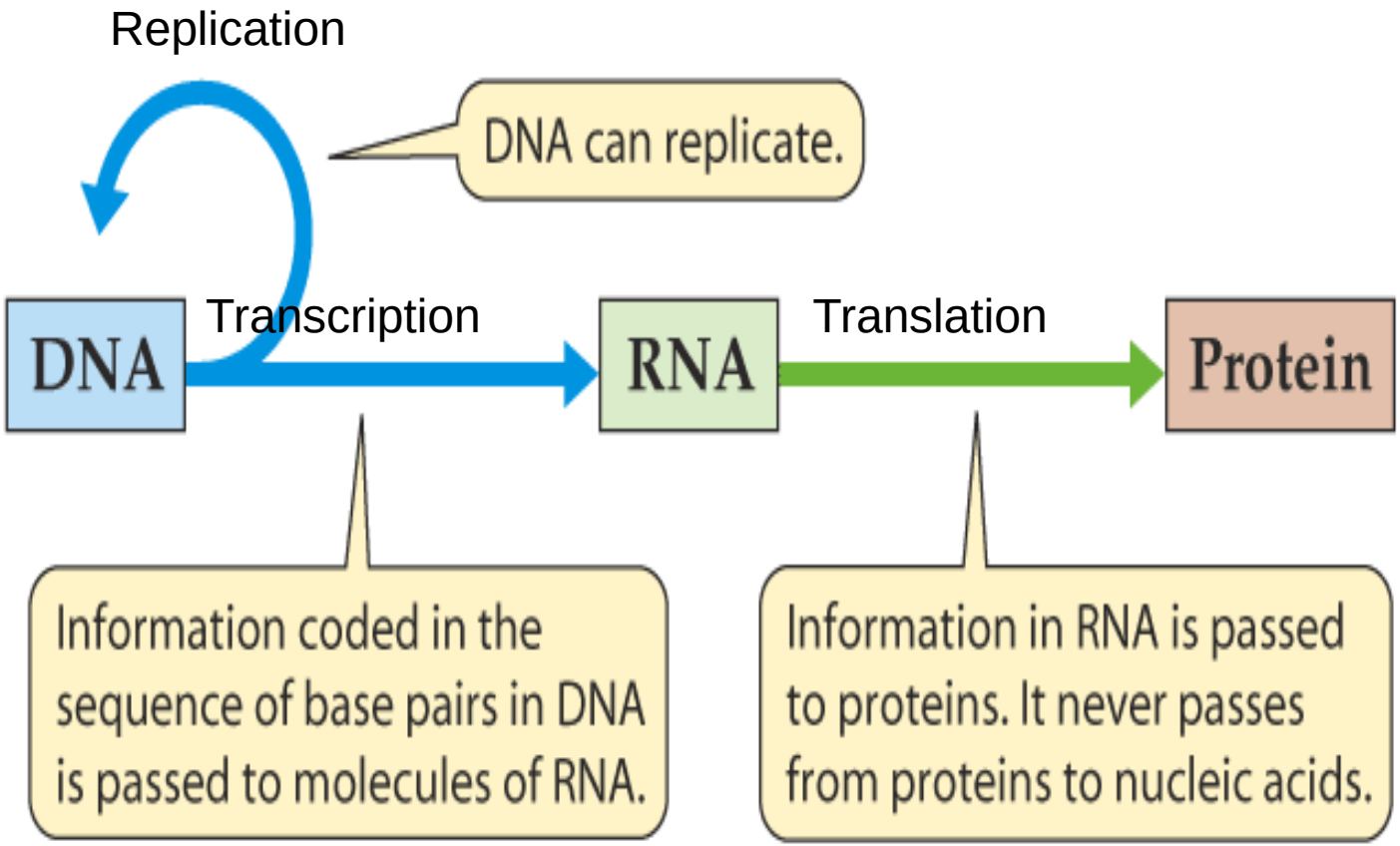


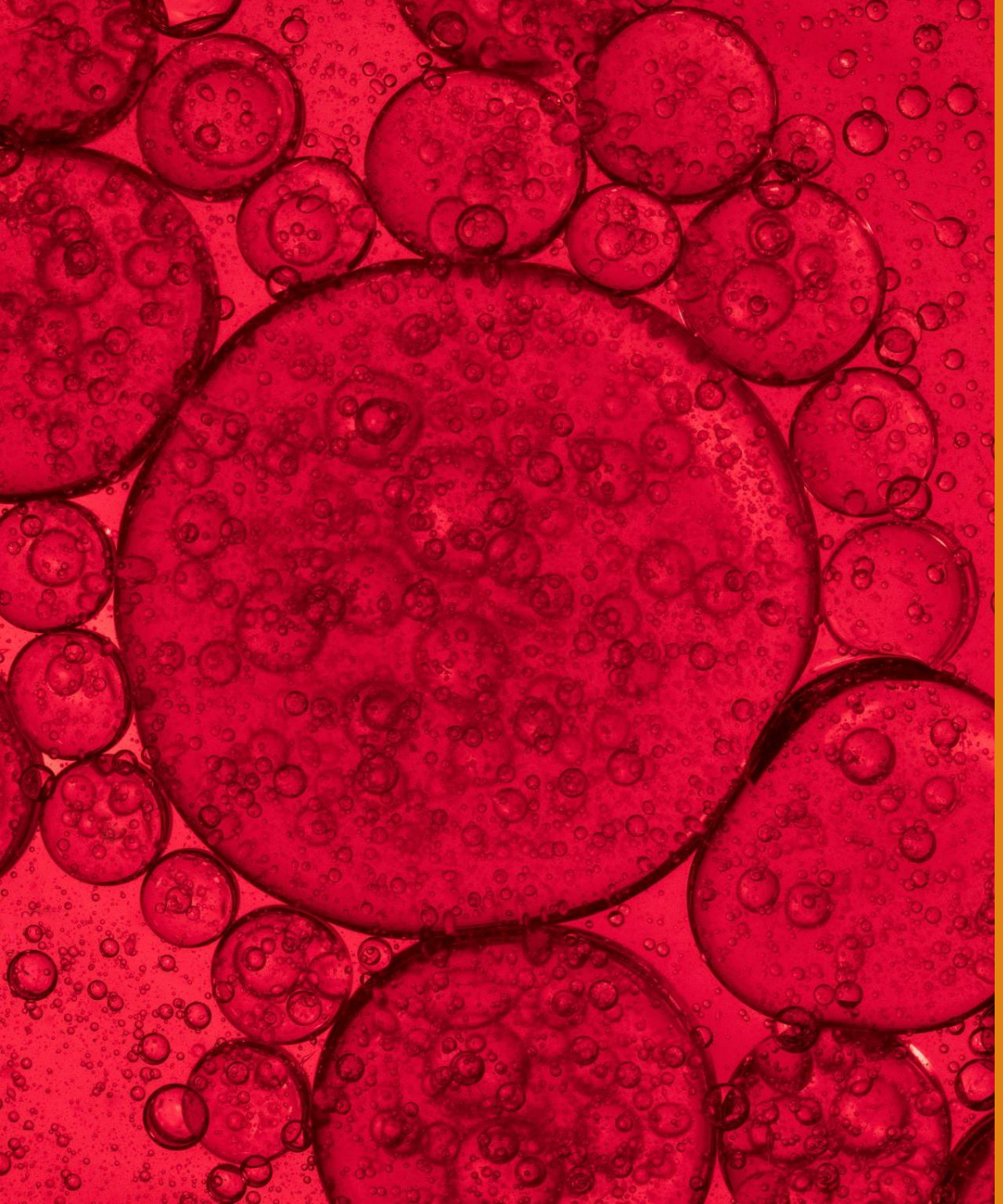
The Central Dogma of Molecular Biology

DNA —transcription→ RNA —translation→ Proteins

DNA —replication→ DNA

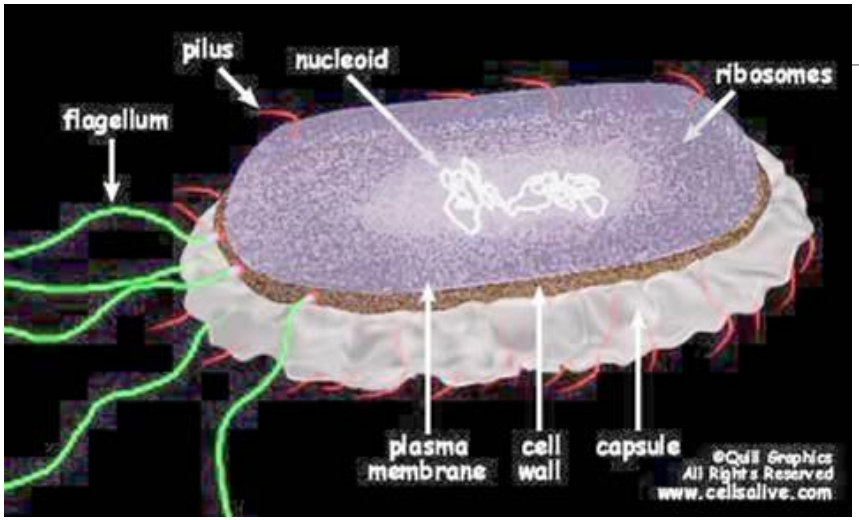
Central Dogma of Molecular Biology

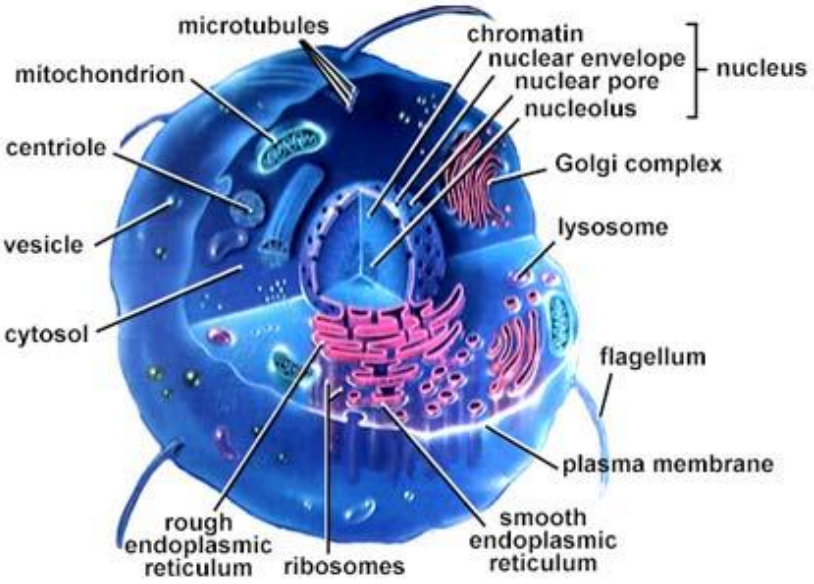




More Detailed Biological Background

2 types of cells: Prokaryotes vs. Eukaryotes





Prokaryotes and Eukaryotes, continued

Prokaryotes	**Eukaryotes**
Single cell	Single or multi cell
No nucleus	Nucleus
No organelles	Organelles
One piece of circular DNA	Chromosomes
No mRNA post transcriptional modification	Exons/Introns splicing

Cells Information and Machinery

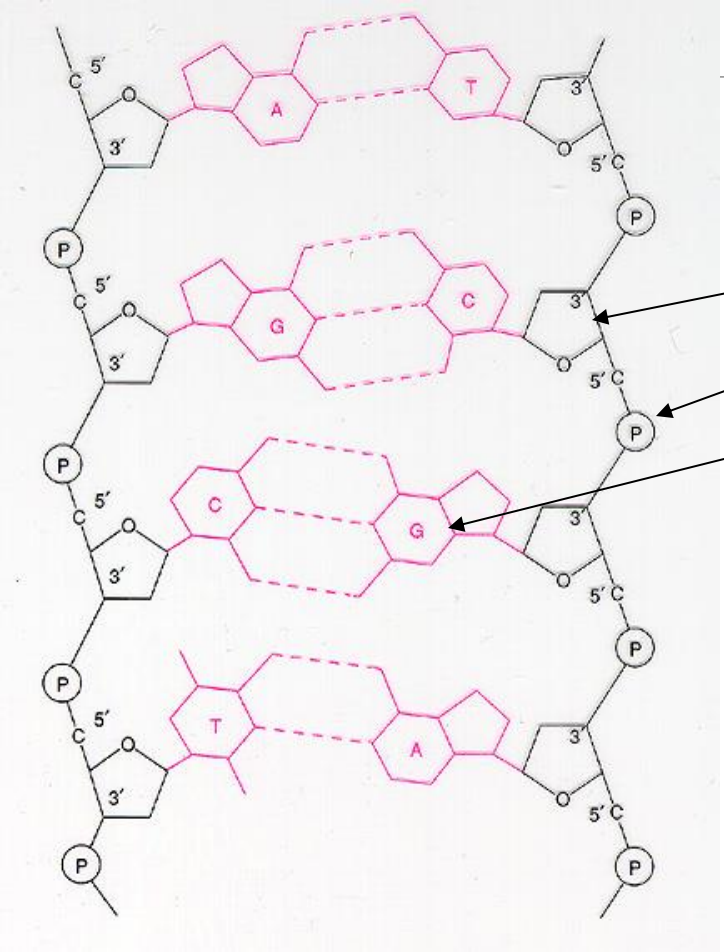
Cells store all information to replicate itself

- Human genome is around 3 billions base pair long
  - Base A, C, G, T
  - Base pair A—T, C—G,G—C,T—A.
- Almost every **cell** in human body contains same set of genes
- But **not all genes** are used or expressed by those cells

Machinery:

- Collect and manufacture components
- Carry out replication
- Kick-start its new offspring

DNA



DNA has a double helix structure which composed of

- sugar molecule
- phosphate group
- and a base (A,C,G,T)

DNA always reads from 5’ end to 3’ end for transcription replication

5’ ATTTAGGCC 3’

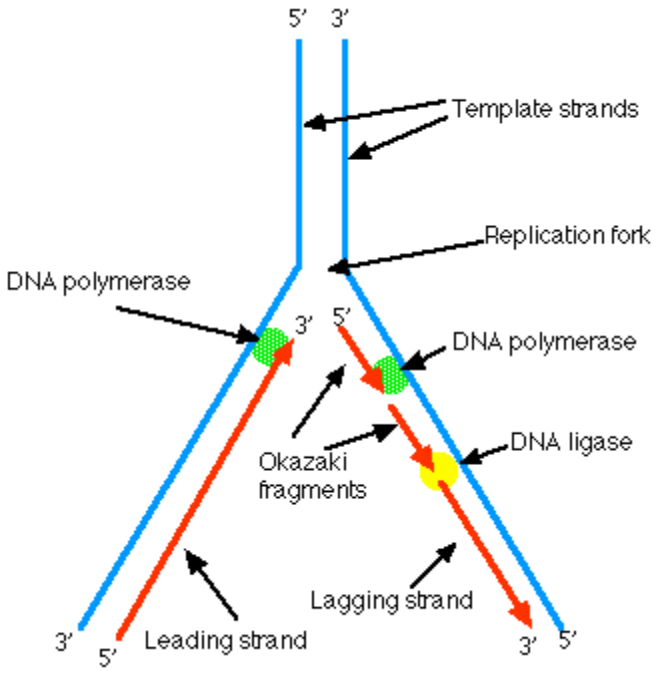
3’ TAAATCCGG 5’

DNA - replication

DNA can replicate by

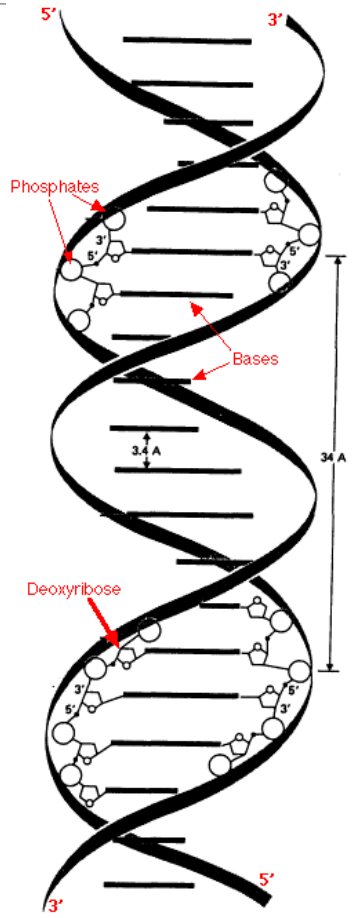
- splitting, and rebuilding each strand.

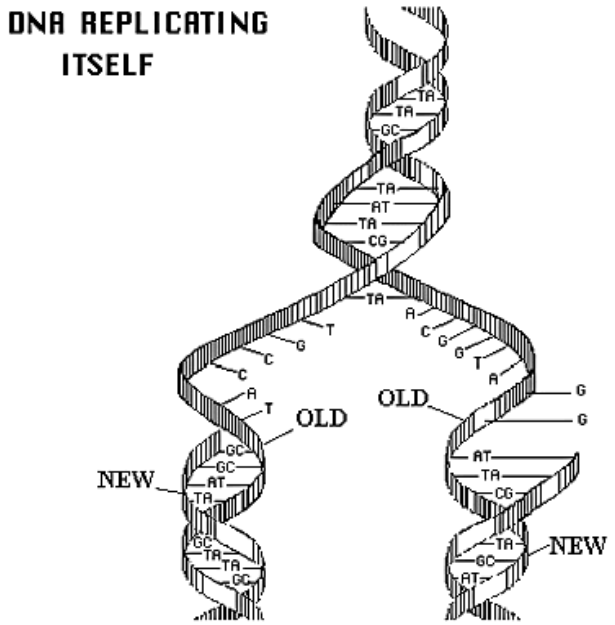
Note that the rebuilding of each strand uses slightly different mechanisms due to the 5’ 3’ asymmetry, but each daughter strand is an exact replica of the original strand.



http://users.rcn.com/jkimball.ma.ultranet/BiologyPages/D/DNAReplication.html

DNA - replication





Mutations

The DNA can be thought of as a sequence of the nucleotides: C,A,G, or T.

What happens to genes when the DNA sequence is mutated?

Normal DNA sequence: ATCTAG

Mutated DNA sequence: ATCGAG

The Good, the Bad, and the Silent

Mutations can serve the organism in three ways:

The Good : A mutation can cause a trait that enhances the organism's function:

Mutation in the sickle cell gene provides resistance to malaria.

The Bad : A mutation can cause a trait that is harmful, sometimes fatal to the organism:

Huntington's disease, a symptom of a gene mutation, is a degenerative disease of the nervous system.

The Silent: A mutation can simply cause no difference in the function of the organism.