Introduction to Molecular Biology

Homayoun Valafar

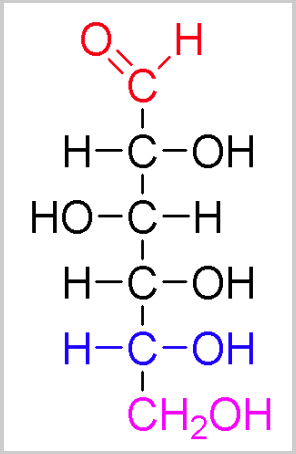
Department of Computer Science and Engineering, USC

Organic Macromolecules

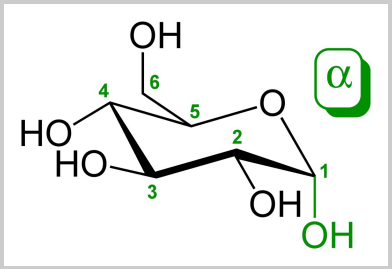
- Four major classes of biologically relevant macromolecules:
  - Polysaccharides/Carbohydrates
    - Chemical compounds that act as the primary biological means of storing or consuming energy
    - Most diverse class of macromolecules (~500,000 structures for a trimer)
  - Lipids
    - Comprises a diverse range of molecules
    - Cholesterols, fatty acids, steroids and was are some examples
  - Nucleic Acids
    - DNA (Deoxyribonucleic acid)
    - RNA (Ribonucleic acid)
  - Proteins

Polysaccharides/Carbohydrates

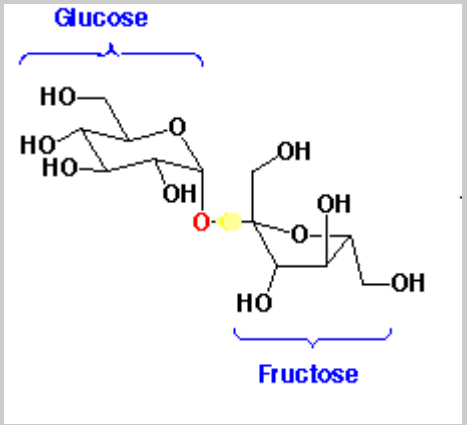
- Made of fundamental units of monosaccharides
- Monosaccharides are primarily made of C, O and H: $(CH_2O)_n$
  - Glucose
  - Can be in linear or ring formation
  - Can be in boat or chair conformation

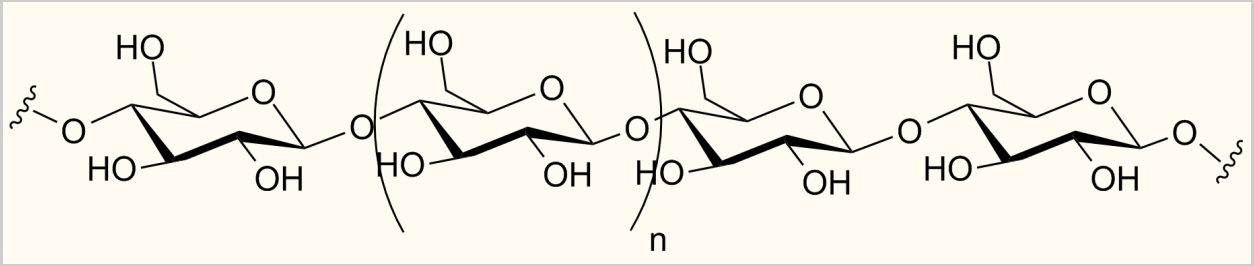


Glucose (monosaccharide)



Sucrose (disaccharide)

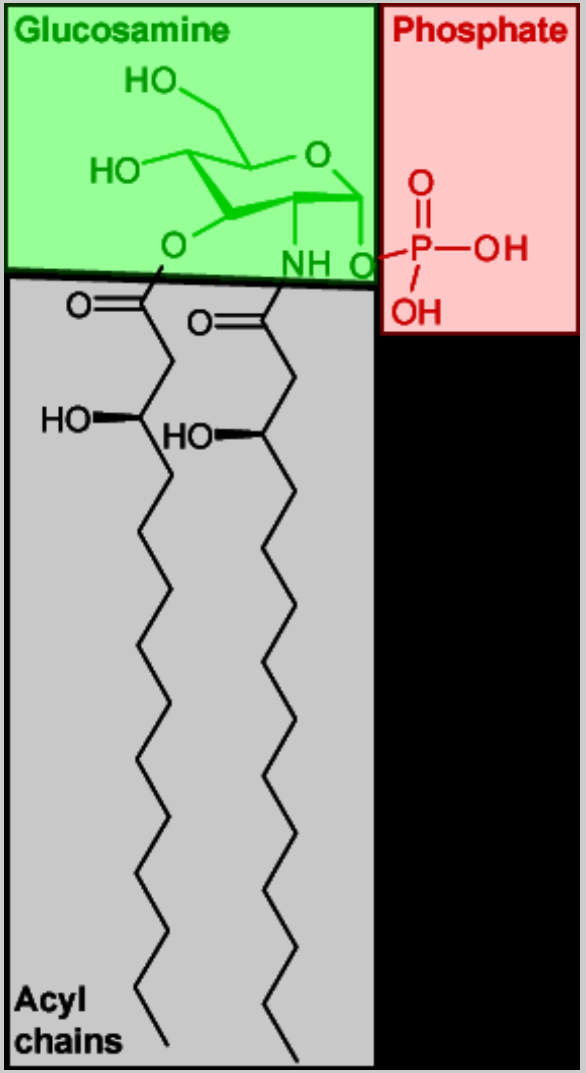




Amylose (polysaccharide)

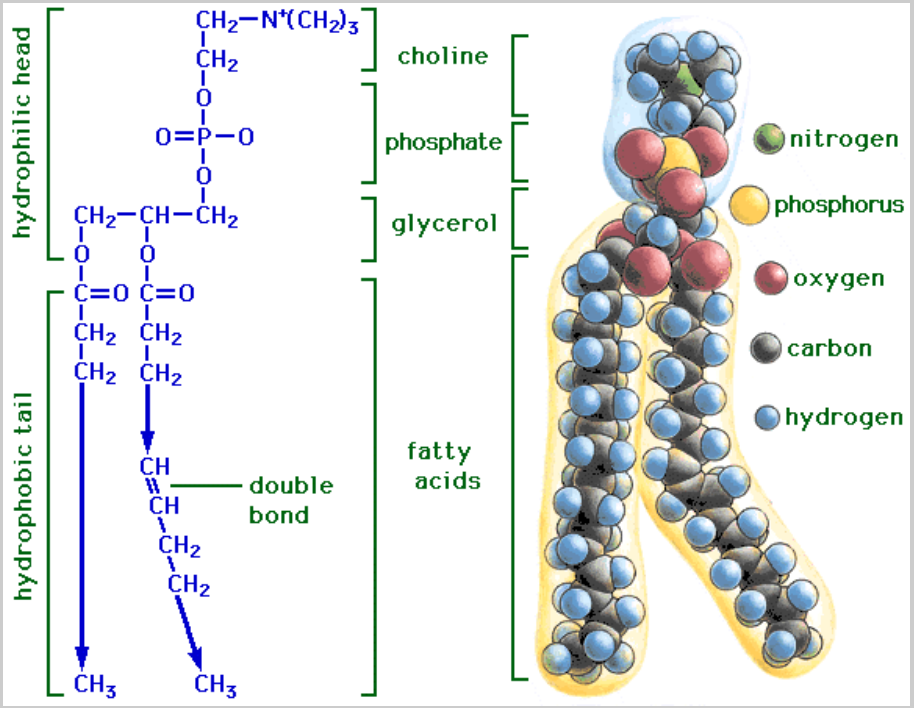
Lipids

- Any fat-soluble (lipophilic)
- Fats, oils, waxes, cholesterol, sterols, fat-soluble vitamins or phospholipids
- The main biological functions:
  - Energy storage
  - Structural components of cell membranes
  - Participating as important signaling molecules

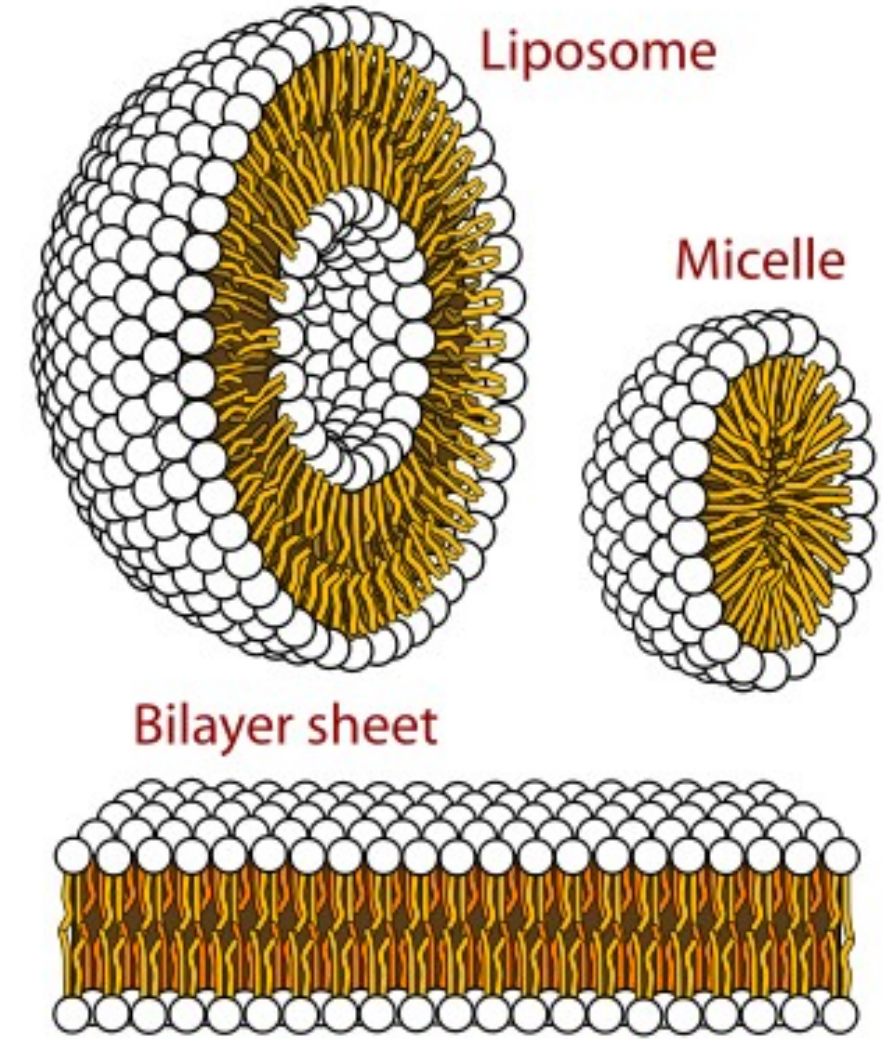


Saturated/Unsaturated Lipids

- Unsaturated lipids:
  - Contain double bonds in the Acyl chain
  - Have a lower melting point, hence increasing fluidity of the cell membranes
- Saturated lipids:
  - Contain no double bonds in the Acyl chain
  - Have a higher melting temperature, and are more problematic in relation to health

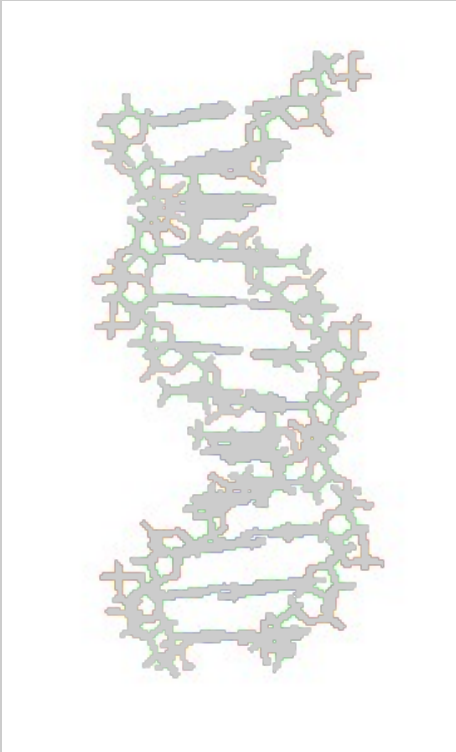


Lipids

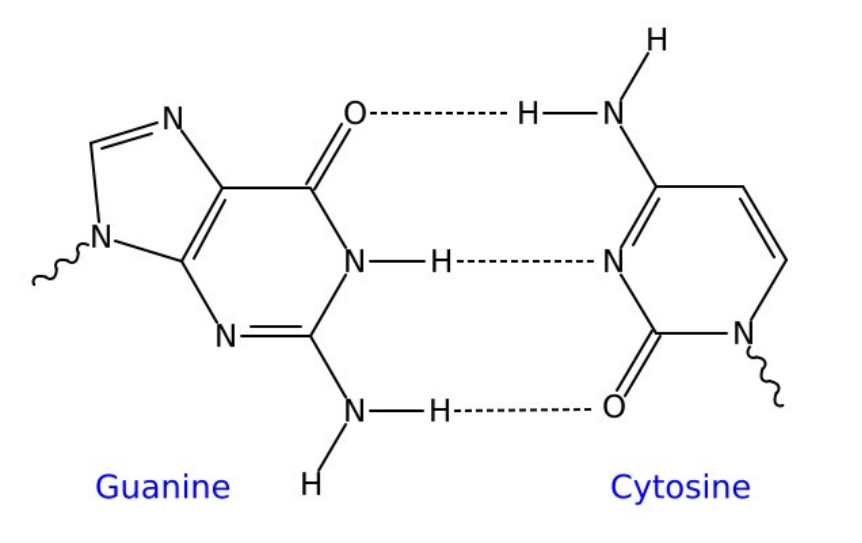


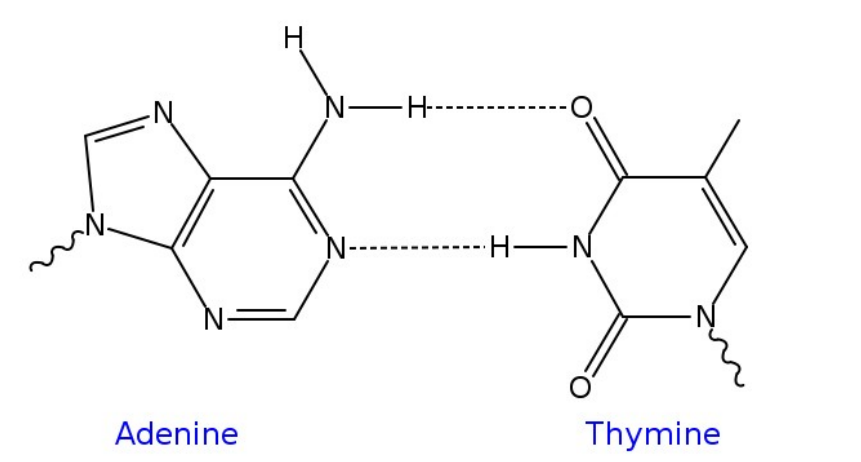
Deoxyribonucleic acid (DNA)

- First structure by Crick and Watson (1953)
- Linear sequence of four fundamental units of nucleotides
  - Adenine, Guanine, Cytosine, Thymine (A, G, C and T)
  - Hydrogen bond pairing: A = T, G ≡ C
- Double helix of life encapsulated in cell nucleus
  - Also referred to as the "molecule of heredity"
  - It is inherited and used to propagate traits
  - Double stranded, stability and complexity
  - Constitutes the genotype of an organism



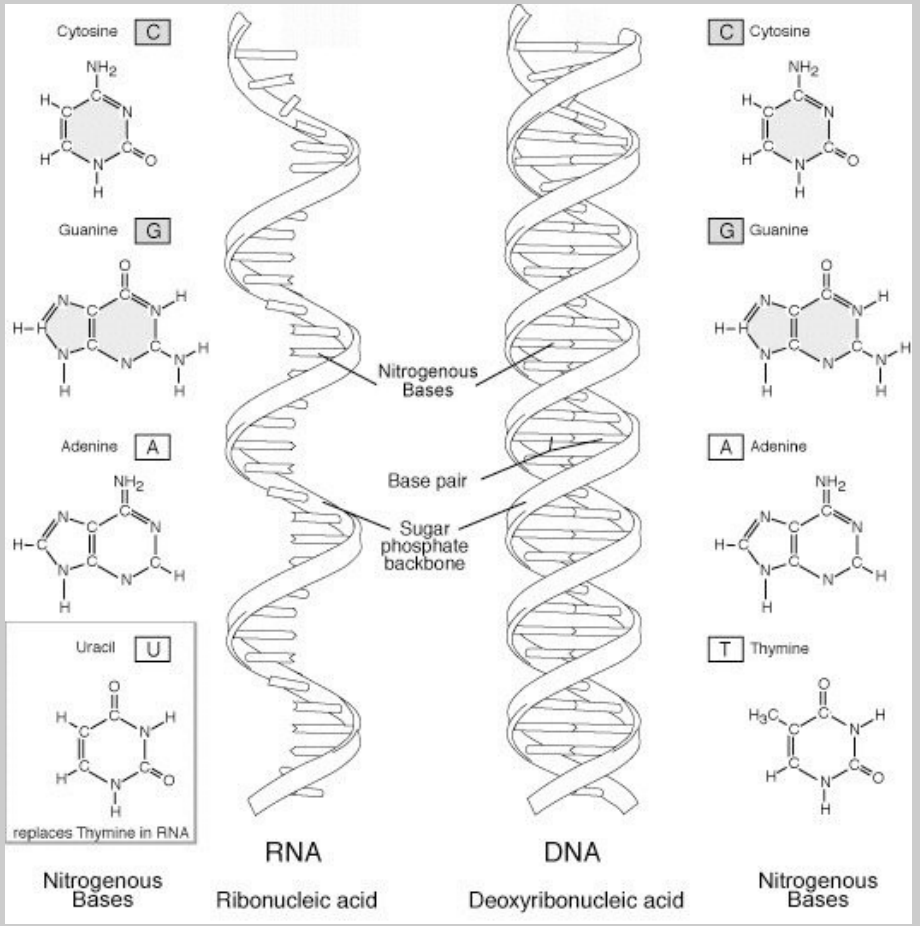
Hydrogen Bonding of Nucleic Acids



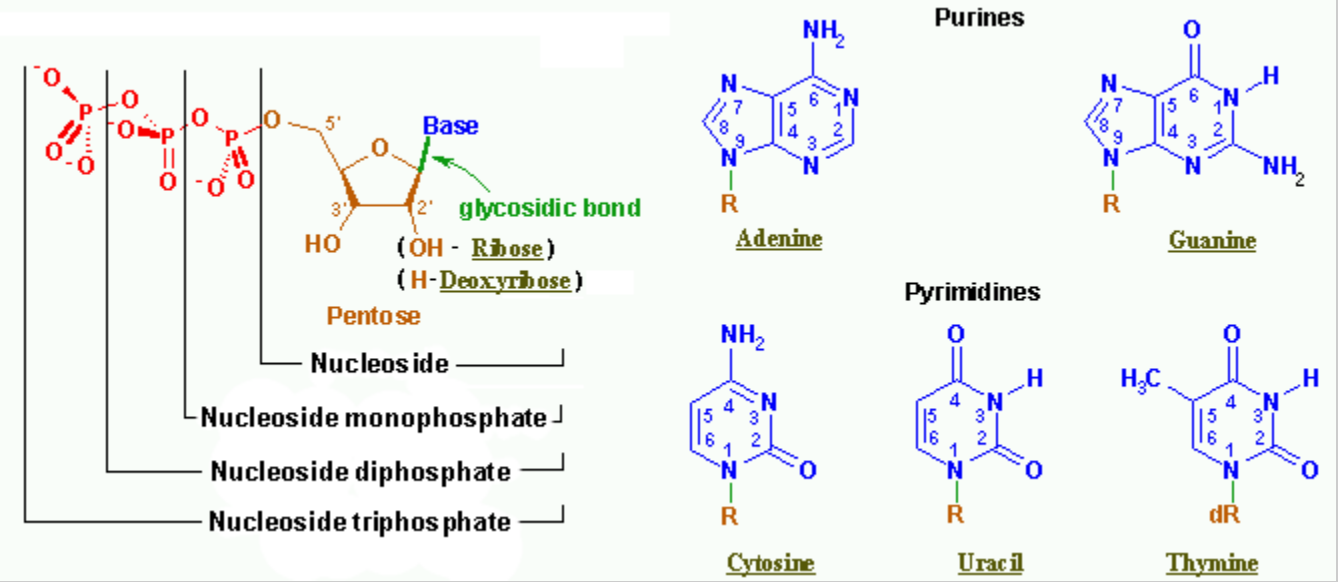


Ribonucleic acid (RNA)

- Also a linear sequence of four fundamental units of nucleotides
  - Adenine, Guanine, Cytosine, Uracil (A, G, C and U instead of T).
  - Complementary base pairing of A-U and G-C
- Single stranded and therefore unstable.
- Three types of RNA
  - mRNA (messenger RNA)
  - tRNA (transfer RNA)
  - rRNA (ribosomal RNA)

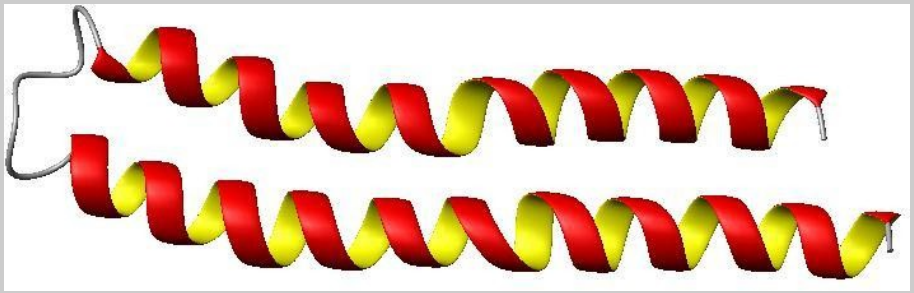


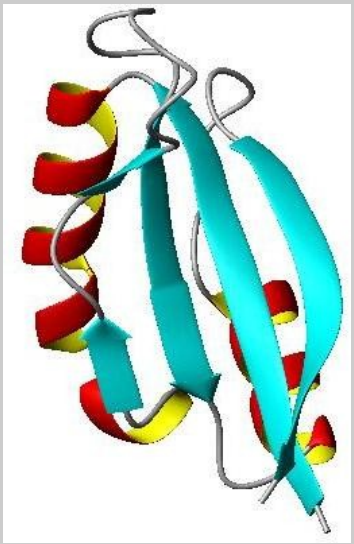
Nucleotides



Protein

- Proteins are functional units of cell
- Proteins are made of 20 “ amino acid” subunits.
- More complex than DNA
- Assumes 3D structures





Common Terms

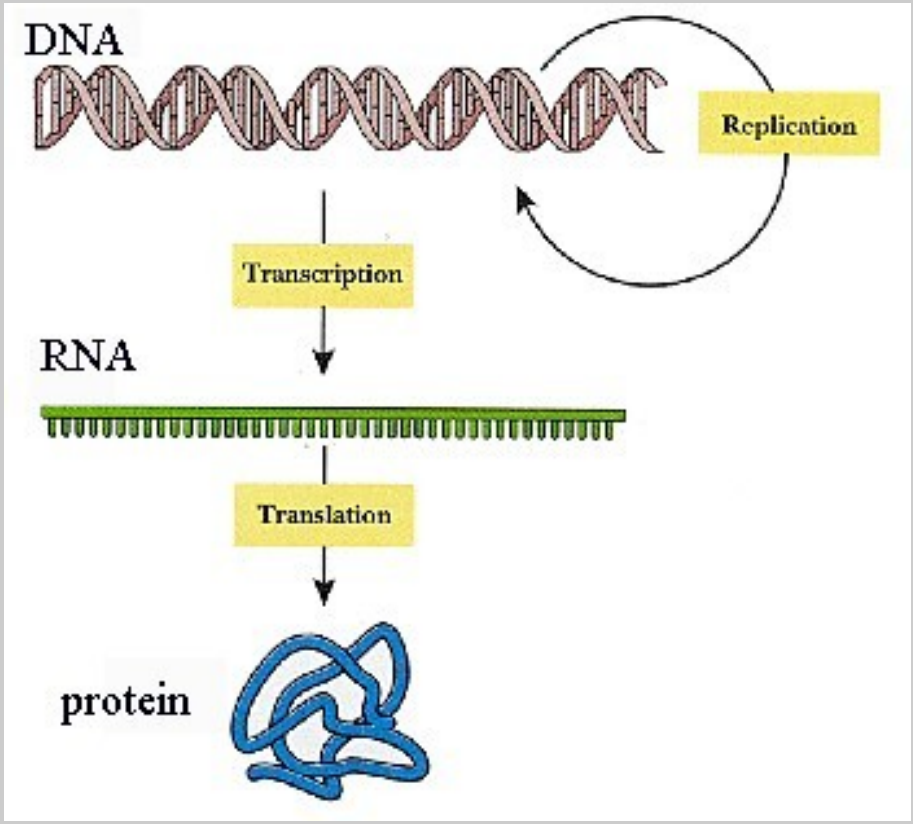
- Chromosome: A chromosome is a very long piece of DNA, which contains many genes
- Gene: Portion of a chromosome that encodes for a trait
- Allele: any one of a number of alternative forms of the same gene occupying a given locus
- Diploid/Haploid: living cells may have one (haploid) or two (diploid) copies of a chromosome
- Autosomal/Sex gene: if a gene is located on the 23rd pair of chromosomes it is a sex gene otherwise autosomal gene
- Dominant/Recessive gene: a dominant allele/ an allele that will be present only if it is present by itself
- Genotype: genetic makeup of an individual cell
- Phenotype: the overall effect of a gene
- Homozygote: a diploid cell that has two copies of the same allele
- Heterozygote : a diploid cell that has two different alleles

History of Modern Biology

- Prior to 1953 very few researchers cared much about DNA, protein characterization as early as 1838
- Linus Pauling, the world's leading structural chemist in 1950, believed that genes were made of protein
- In 1951, both Sir William Lawrence Bragg and Linus Pauling wanted to find the structure of the master molecule of life, the gene. But both were focused on proteins
- The only strong evidence against protein genes and in favor of DNA was a little-appreciated paper published in 1944 by Rockefeller Institute researcher Oswald Avery, who found that DNA, apparently by itself, could transfer new genetic traits between *Pneumococcus* bacteria
- "I knew the contention that DNA was the hereditary material, but I didn't accept it," Linus Pauling
- Watson & Crick (1953) "The Double Helix"

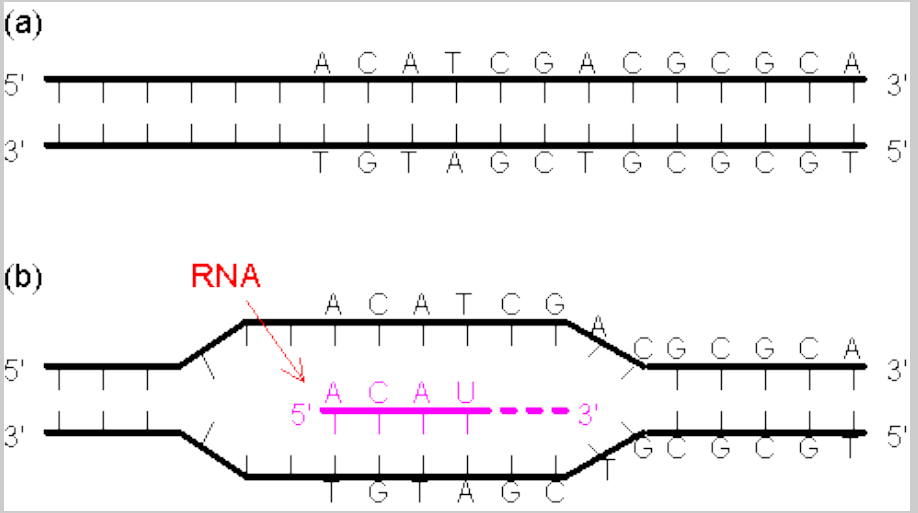
Central Dogma of Modern Biology

- Set forth by Francis Crick in 1958
- How do genes perform their function?
- How does genotype get translated to phenotype?
- The answer to these questions is the “central dogma of modern biology”
- Proteins are the functional units of cell



Transcription, Reference

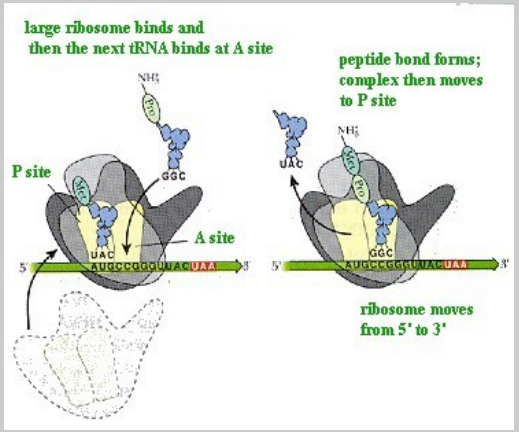
- The first process in gene expression
- DNA is copied to RNA.
- Performed by an enzyme called RNA polymerase (RNAP).
- Transcription is similar to DNA replication.
- 3'-5' strand (antisense strand) is transcribed into mRNA.
- Transcription starts from the promoter region and stops at the terminator sequence.
- A promoter is a DNA sequence that enables a gene to be transcribed.

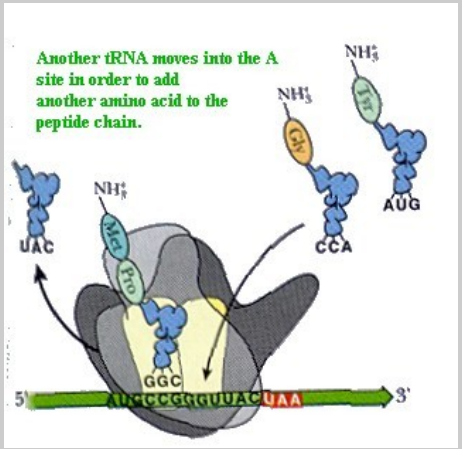


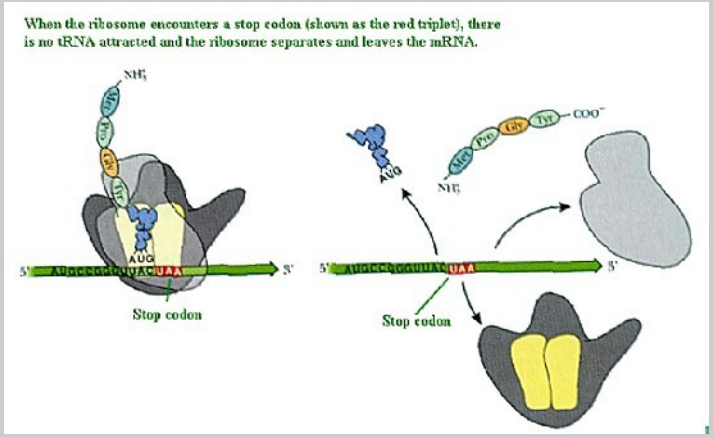
5' ACATCGACGCGCAGTTAATCCC..3' DNA coding strand (+)
3' TGTAGCTGCGCGTCAATTAGGG..5' DNA template strand (-)
5' ACAUCGACGCGCAGUUAAUCCC..3' RNA (+)

Translation, Reference

- mRNA is processed by ribosome to produce a protein through the process of translation
- Three parts to translation
- Appropriate amino acids are brought to ribosome by tRNAs
- 20 amino acids but only 4 nucleotides!!! How then??

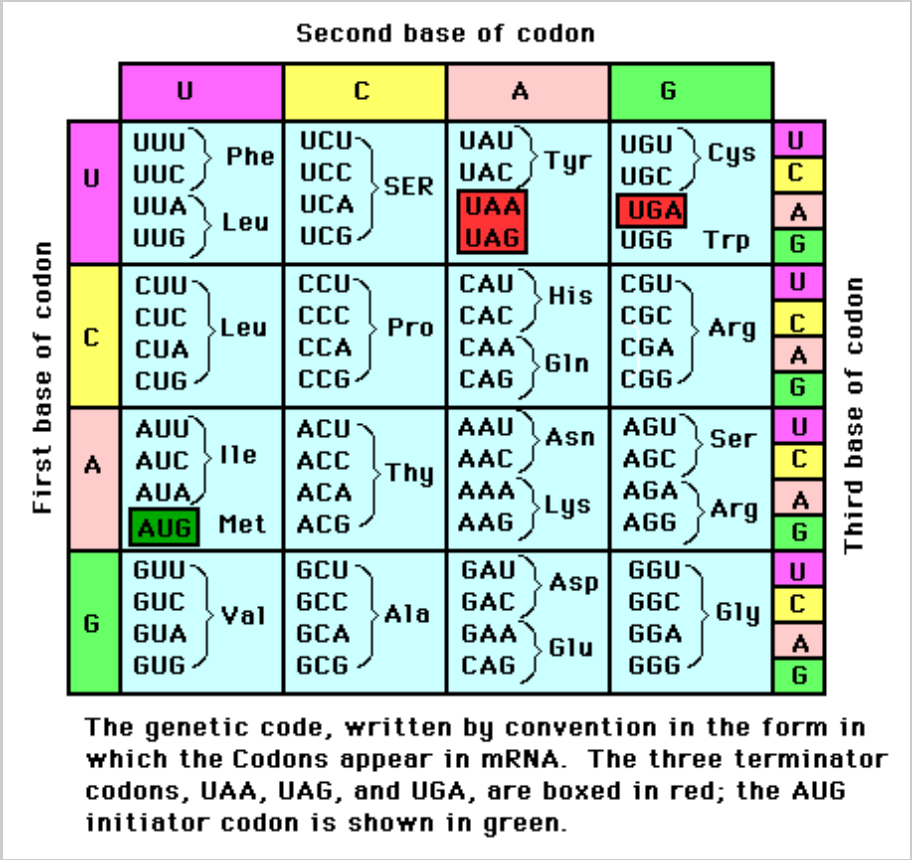






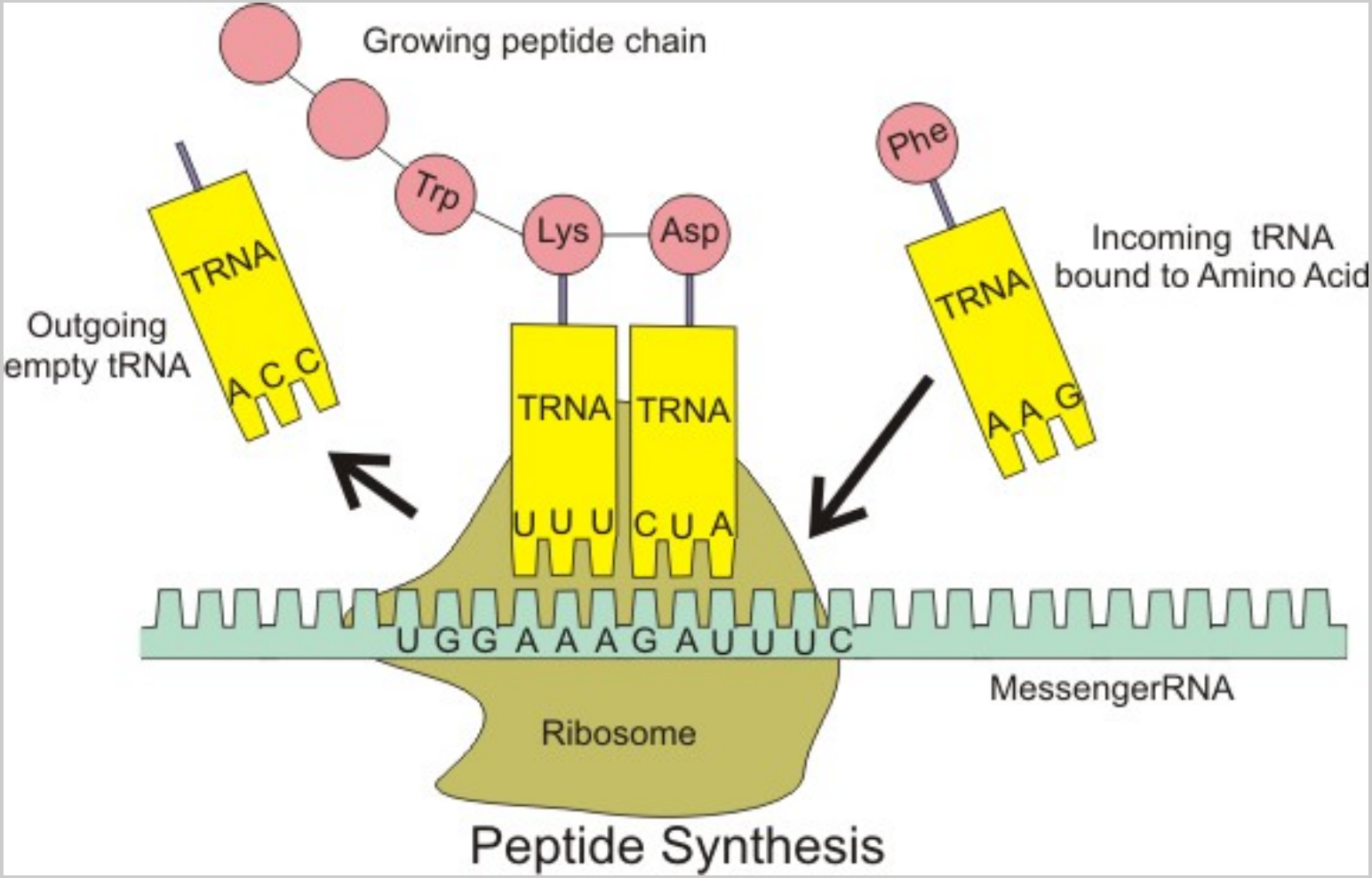
Translation Codon Table

- Every 3 nucleotides is called a codon
- Each codon can decipher for up to 4^3 or 64 different units
- Only 20 amino acids need to be indicated
- Always some redundancy in codon to amino acid mapping
- The redundancy could be considered room for growth (evolution?)



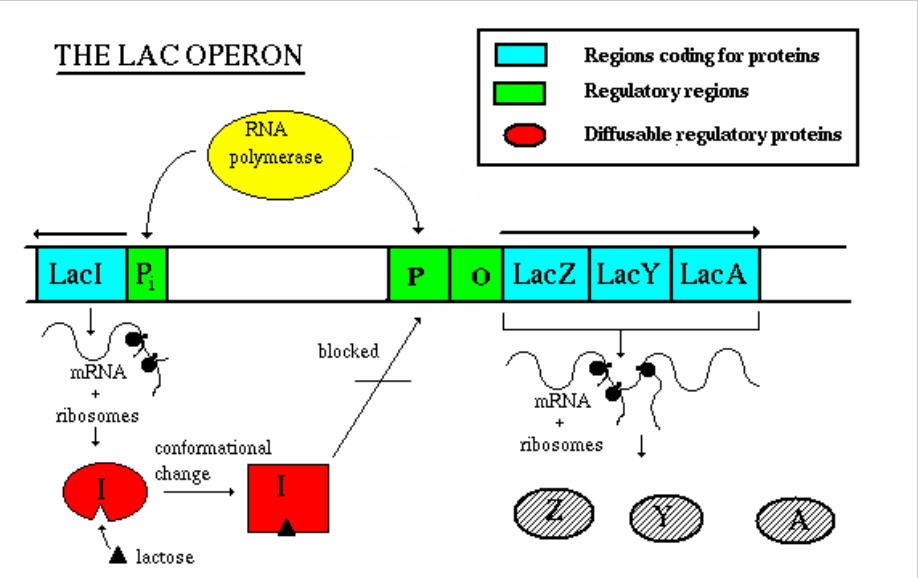
The genetic code, written by convention in the form in which the Codons appear in mRNA. The three terminator codons, UAA, UAG, and UGA, are boxed in red; the AUG initiator codon is shown in green.

Translation



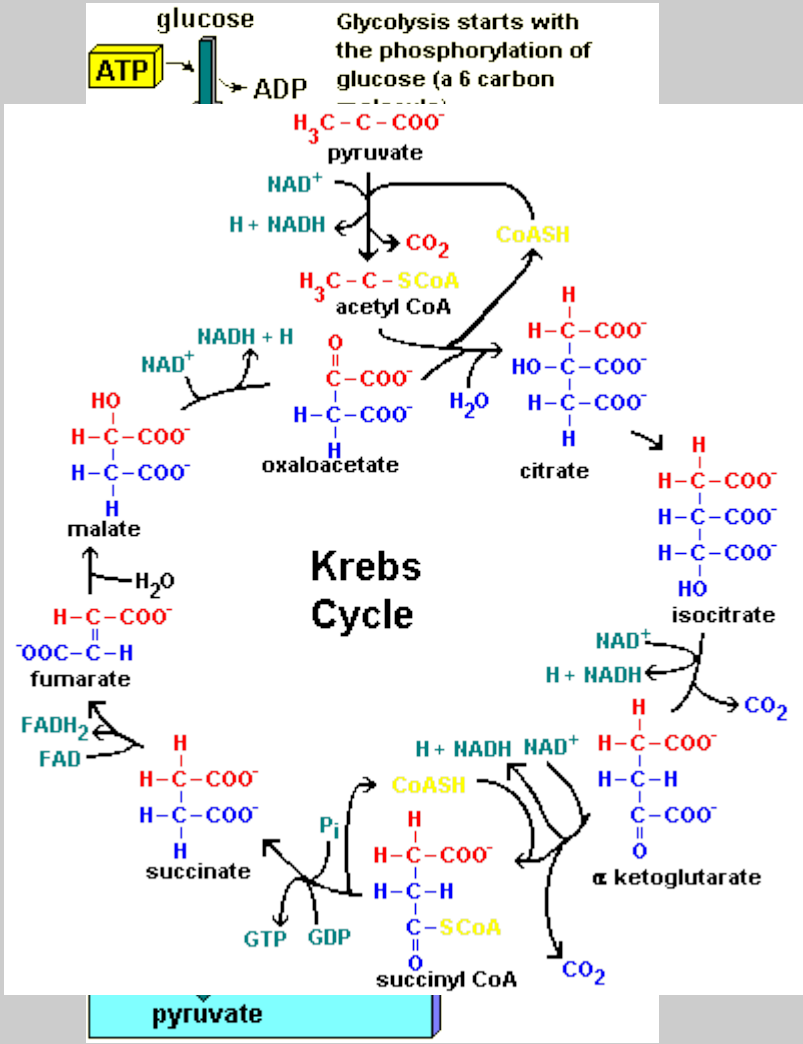
Regulation of Gene Expression

- Lactose Regulation in Bacteria (Lac Operon)



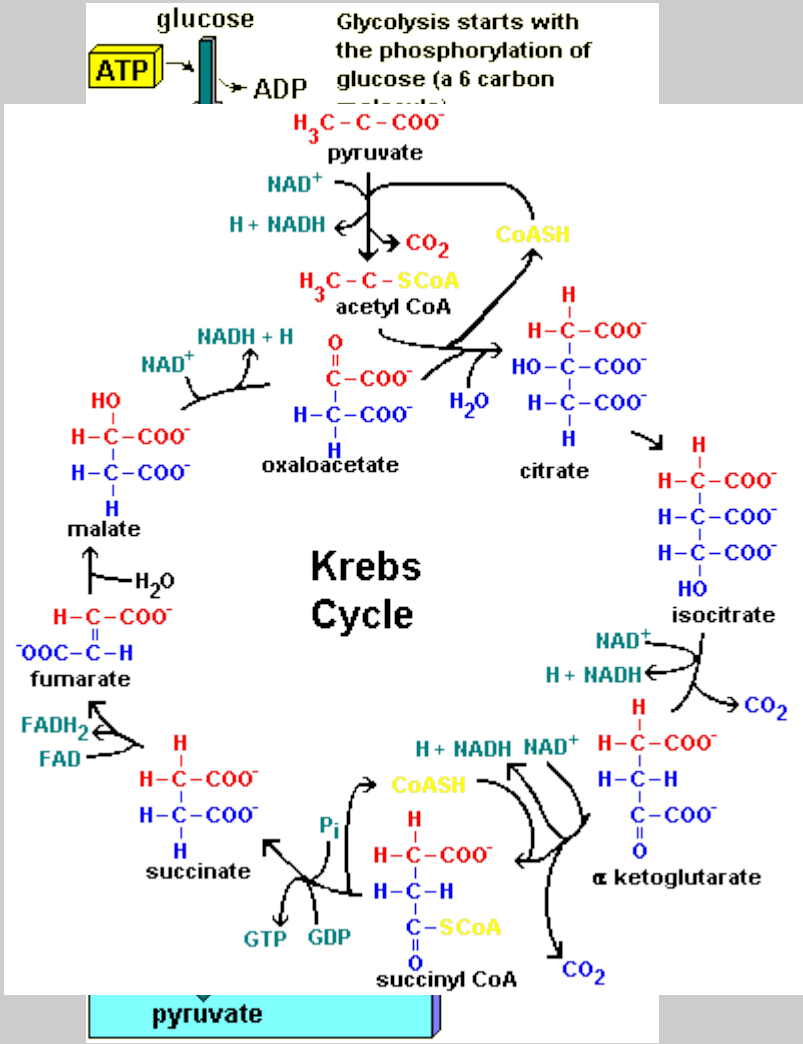
Example of Biochemical Pathway

- Glycolysis
  - The process of degrading Glucose
  - Takes place in the Cytoplasm
  - Requires no oxygen
  - Net yield of 2 ATPs
  - One glucose → 2 pyruvates
- Krebs cycle, Citric acid cycle
  - Aerobic process
  - Takes place in Mitochondria
  - Net yield of 36 ATPs



Example of Biochemical Pathway

- Glycolysis
  - The process of degrading Glucose
  - Takes place in the Cytoplasm
  - Requires no oxygen
  - Net yield of 2 ATPs
  - One glucose → 2 pyruvates
- Krebs cycle, Citric acid cycle
  - Aerobic process
  - Takes place in Mitochondria
  - Net yield of 36 ATPs



Genome Size Implies Complexity?

- Cow, worm and wheat have approximately the same size genome.
- They are very different in complexity. How is that possible?
- Total number of genes between higher organisms and lower organisms does not explain the difference in complexity.
- Multicellular organisms have cells with the same genome but different phenotypes and function. How is that possible?
- The entire human genome consists of ~3 billion bp and only 25000 genes.
- Does this mean that there are no more than 25000 proteins in humans?
- Alternate gene splicing provides the answer.

Same Gene/Different Function

- All cells in a given human being have identical genes
- Cells from different tissues have different functions
- How is that possible?
- Result of alternate gene expression and splicing
- Cell differentiation starts early in the development of the embryo
- Cell differentiation can be viewed as the start point of alternate splicing
- Differentiation is not reversible
- Most organs do not regenerate. Once gone, forever gone…
- Stem cell research

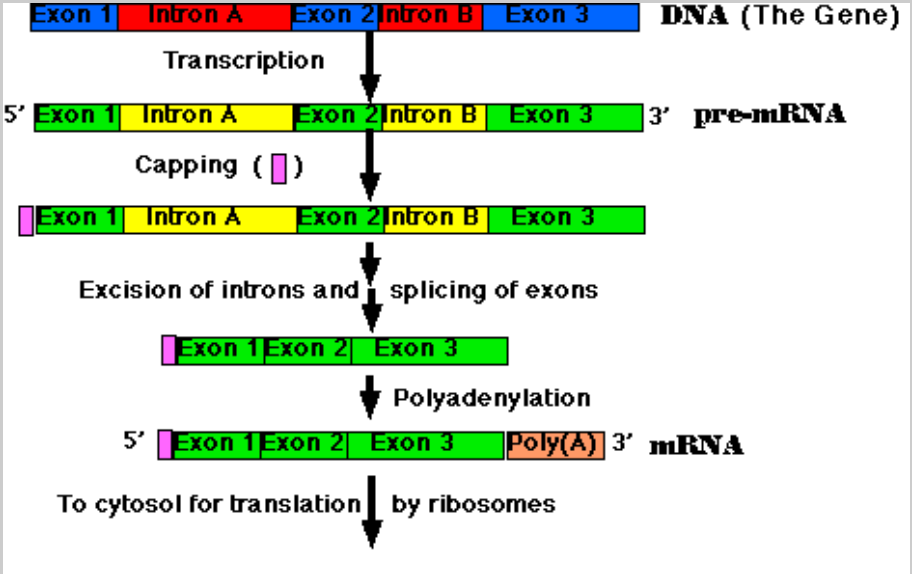
Hierarchy of Higher Organisms

- Cells similar in function form a tissue (Neurons).
- Tissues that are compatible in function form an organ (Brain).
- Organs that cooperate, form an organ system (Nervous system).
- Integrated systems produce a complex organism such as Humans.
- Total of 12 systems in humans.

- Circulatory system
- Digestive system
- Endocrine system
- Immune system
- Integumentary system
- Lymphatic system
- Muscular system
- Nervous system
- Reproductive system
- Respiratory system
- Skeletal system
- Urinary system

Alternate Splicing

- mRNA of eukaryotes have mixed regions named introns and exons
- Before mRNA exits the nucleus, the introns are deleted
- Different cell types have different criteria for what is intron/exon. This constitutes alternate splicing
- mRNA can become a completely different functional proteins in different cells due to alternate splicing
- Reference 1, 2, 3



Dynamic World of a Cell