# BINF 730
# Biological Sequence Analysis

**Saleet Jafri**
**Program in Bioinformatics and Computational Biology**

**George Mason University**

## Lecture 1

**Overview of Molecular and Cellular Biology**

## Biological References

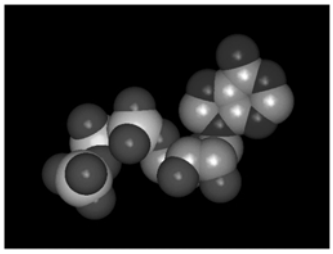
- ***Molecular Biology of the Cell***
  **by Bruce Alberts**
  **(1994 or newer edition)**

- ***Molecular Cell Biology***
  **by Darnell, Lodish, and Baltimore**
  **(1995 or newer edition)**

## Part I: Molecular Biology Review

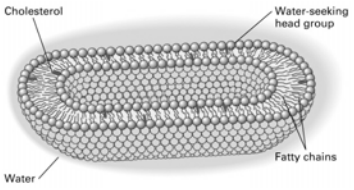
**Where do biological sequences come from?**

- **Life and evolution**
- **Proteins**
- **Nucleic Acids**
- **Central dogma**
- **Genetic code**
- **DNA structure**
- **Mitochondrial DNA**

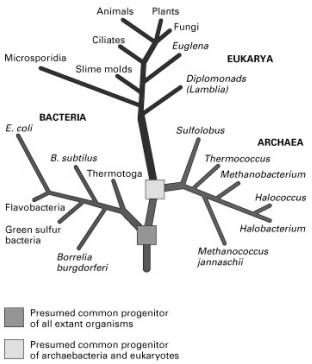


## Life

- **Evolved from common origin**
- **˜3.5 billion years ago**
- **All life shares similar biochemistry**
  - **Proteins: active elements**
  - **Nucleic acids: informational elements**
- **Molecular Biology: the study of structure and function of proteins and nucleic acids**




## Terrestrial Life



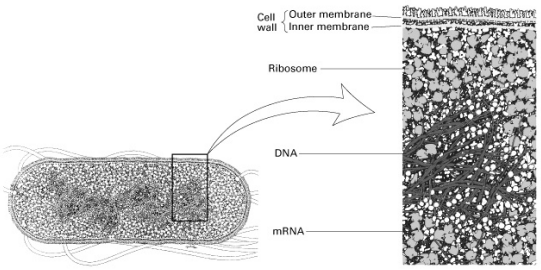
## Cell types

**Prokaryotes – no nuclear membrane, represented by cyanobacteria (blue-green algae) and common bacteria (*Escherichia coli*)**

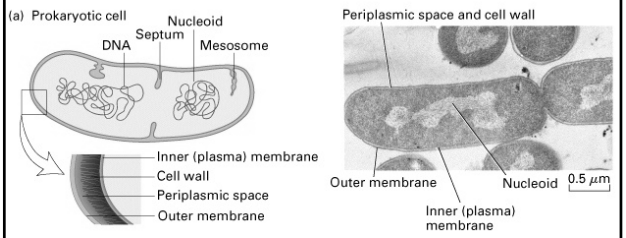
**Eukaryotes – unicellular organisms such as yeast and multicellular organisms**

**Archaebacteria – no nuclear membrane but similar to eukaryotes in transcription and translation mechanisms, discovered in deep sea thermal vents in 1982**

## Prokaryotic Cell



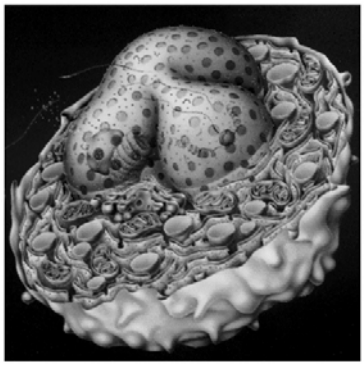

## Prokaryotic Cell



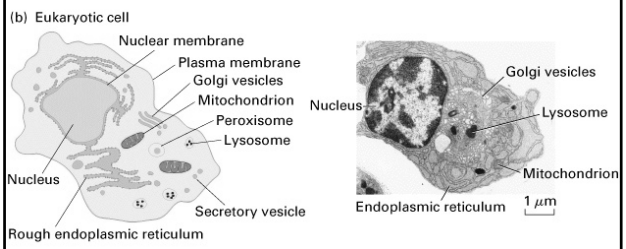

## Eukaryotes

- **In eukaryotes, transcription is complex:**
  - **Many genes contain alternating exons and introns**
  - **Introns are spliced out of mRNA**
  - **mRNA then leaves the nucleus to be translated by ribosomes**
- **Genomic DNA: entire gene including exons and introns**
  - **The same genomic DNA can produce different proteins by alternative splicing of exons**
- **Complementary DNA (cDNA): spliced sequence containing only exons**
  - **cDNA can be manufactured by capturing mRNA and performing reverse transcription**

## Eukaryotic Cell



## Eukaryotic Cell



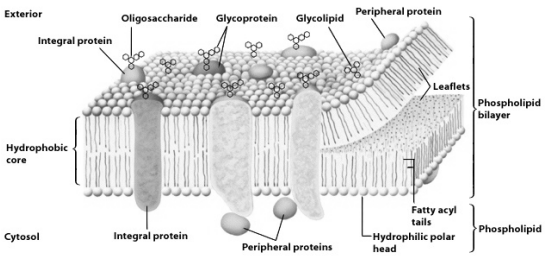
## Eukaryotic Cell Organelles

- Cell membrane
- Nucleus
- Cytoplasm
- Endoplasmic Reticulum – rough and smooth
- Golgi Apparatus – received newly formed proteins from the ER and modifies them and directs them to final destination
- Mitochondria – respiratory centers, have own circular DNA, bacterial origin

## Eukaryotic Cell Organelles

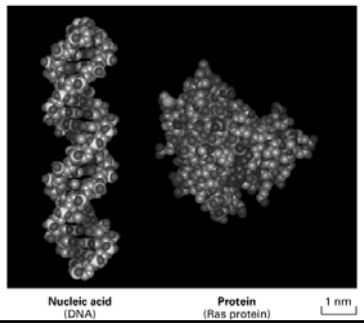
- Chromosomes – chromatin, histones, centromeres and arms (2 pairs in eukaryotes)
- Lysosomes – contain acid hydrolases – nucleases, proteases, glycodidases, lipases, phosphatases, sulfatases, phospholipases
- Peroxisomes – use oxygen to remove hydrogen from substrates forming $H_2O_2$, abundant in kidney and liver – detoxification
- Cytoskeleton

## Eukaryotic Membrane

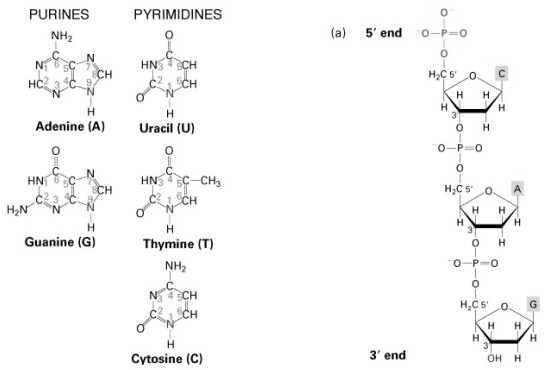



## Nucleic Acids

- Two kinds:
  - RNA: ribonucleic acid
  - DNA: deoxyribonucleic acid




## Nucleic Acid Structure

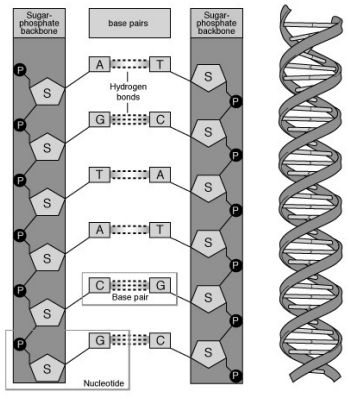



## DNA

- Double stranded
- Four bases: adenine (A), guanine (G), cytosine (C) and thymine (T)
- A and G are purines
- C and T are pyrimidines
- A always paired with T (complementary)
- C always paired with G (complementary)

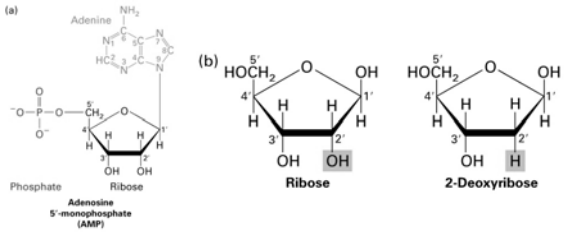
=> Watson-Crick base pairs (bp)

- DNA may consist of hundreds of millions bp
- A short sequence (<100) is called an oligonucleotide



## RNA

- Different sugar (ribose instead of 2'-deoxyribose)
- Uracil (U) instead of thymine (U binds with A)
- RNA does not form a double helix
- RNA may have a complex three-dimensional structure




## Central Dogma

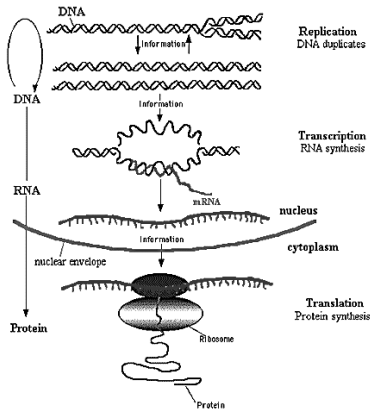
# DNA → RNA → Protein

**DNA = Deoxyribonucleic Acid**

**RNA = Ribonucleic Acid**

**Protein = Functional and Structural units of cells**

**Flow of Information is unidirectional**

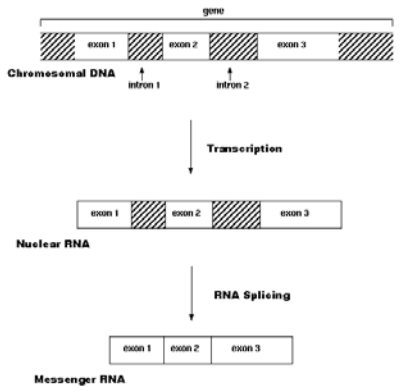



**The Central Dogma of Molecular Biology**

## Gene Transcription or DNA Transcription

- RNA molecules synthesized by *RNA polymerase*
- RNA polymerase found in free and bound form
- RNA polymerase binds very tightly to *promoter* region on DNA
- Promoter region contains *start site*
- Transcription ends at *termination signal* site.
- Primary transcript – direct coding of RNA from DNA
- RNA splicing – introns removed to make the mRNA
- mRNA – contains the sequence of *codons* that code for a protein
- uracil replaces thymine
- splicing and alternative splicing

## Transcription of DNA to Messenger RNA




## Translation

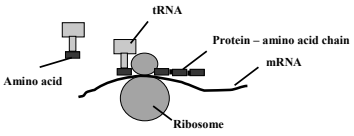
- Ribosomes made of protein and ribosomal RNA (rRNA)
- Transfer RNA (tRNA) make connection between specific codons in mRNA and amino acids
  - As tRNA binds to the next codon in mRNA, its amino acid is bound to the last amino acid in the protein chain
- When a STOP codon is encountered, the ribosome releases the mRNA and synthesis ends

## Gene Translation

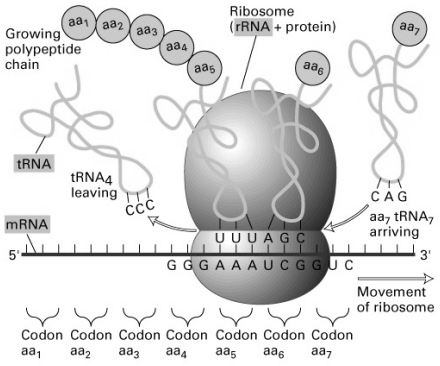
- ***tRNA*** **– links an amino acid to the codon on the mRNA via the *anit-codon***
- ***rRNA*** **– RNA found in ribosomes**
- ***ribosomes*** **– large and small subunit, made of protein and rRNA**
- ***initiator tRNA*** **always carries methionine**
- ***initiation factors*** **– proteins that catayze the start of transcription**
- **stop codon**
- **Endoplasmic Reticulum**
- **Posttranscriptional modification**

## Translation

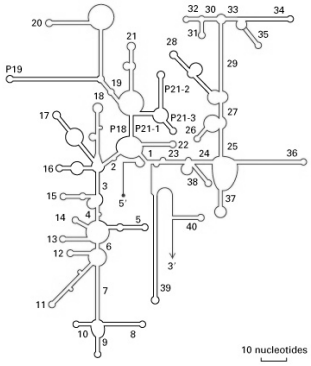
- **Involves ribosomes, and RNA**
- **Ribosomes made of protein and RNA**
- **Messenger RNA (mRNA) is the sequence transcribed from the DNA**
- **The mRNA is 'threaded' through the ribosomes.**
- **Transfer RNA (tRNA) brings the different amino acids to the ribosome complex so that the amino acids can be attached to the growing amino acid chain.**



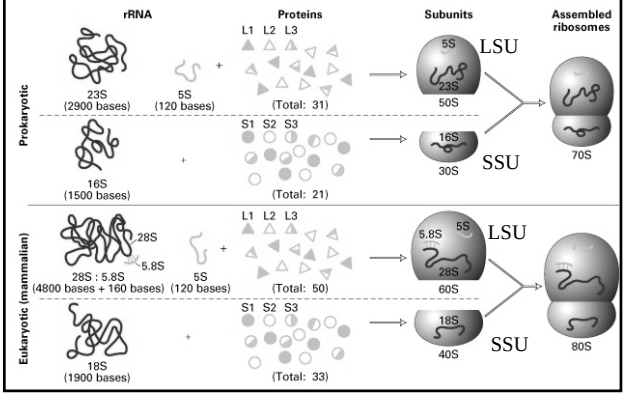

## Ribosome

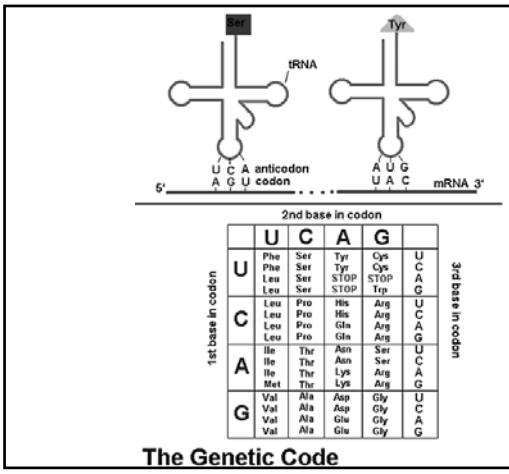



## Ribosomal RNA

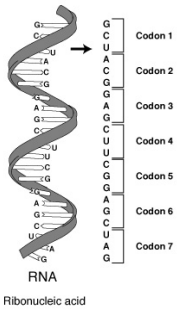



## Eukaryotic and Prokaryotic Ribosome Structure



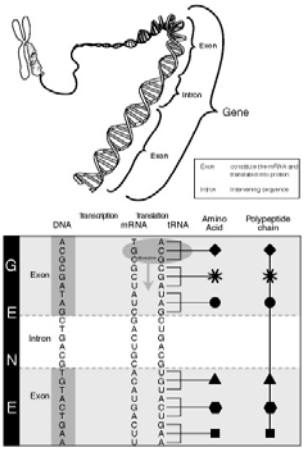




**The Genetic Code**




## DNA Structure

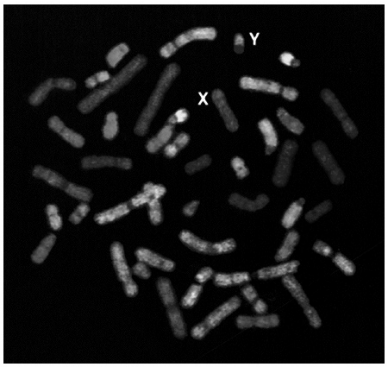
- DNA contains:
  - Promoters
  - Genes
  - Junk DNA
- Reading frames
- An open reading frames (ORF): a contiguous sequence of DNA starting at a start codon and ending at a STOP codon



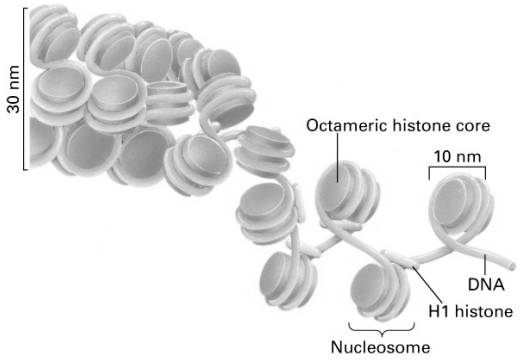

## Chromosomes

- A genome: a complete set of chromosomes within a cell
- Different species have different numbers of chromosomes in their genomes
- Prokaryotes usually have a single chromosome, often a circular DNA molecule
- Eukaryotic chromosomes appear in pairs (diploid), each inherited from one parent
  - Homologous chromosomes carry the same genes
  - Some genes are same in both parents
  - Some genes appear in different forms called alleles
    - E.g. human blood has three alleles: A, B, and O
- All genes are present in all cells, but a given cell types only expresses a small portion of the genes

## Chromosomes




## Chromatin Structure



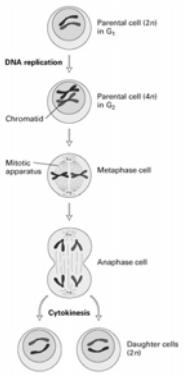
## Gene Coding and Replication

- Double helix
- Nitrogenous bases A,T,G,C
- Sugar-Phosphate backbone
- Nucleotide – sugar + base + phosphate group
- Nucleoside – sugar + base
- Purines – adenine, guanine
- Pyrimidines – cysteine, thymine
- A-T – 2 H bonds, G-C – 3 H bonds

## Gene Coding and Replication

- 5' end contains a phosphate group
- 3' end is free
- DNA extended from 5' to 3'
- Gene is a segment of DNA that codes for a specific protein
- Exons are coding regions of the DNA
- Introns are 'in between' regions, found in eukaryotes
- Codons
- Reading frame
- *Consensus sequences* are conserved regions found in a particular type of regulatory region

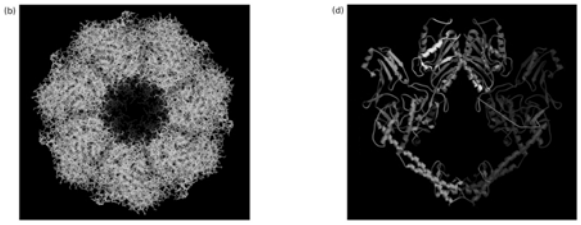
## Mitosis

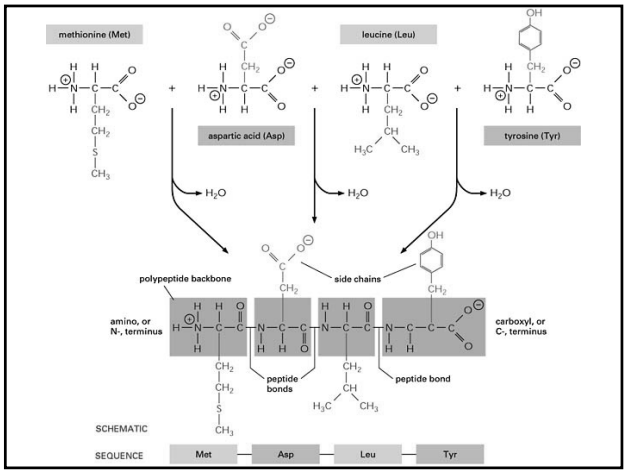



QuickTime™ and a Sorenson Video decompressor are needed to see this picture.

## Proteins

- Functions:
  - Structural proteins
  - Enzymes
  - Transport
  - Antibody defense
- Chains of amino acids
- Typical size ~300 residues




## Protein Folding

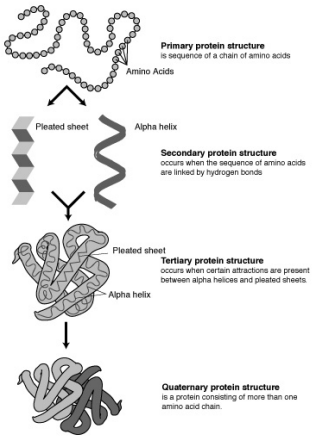
- Primary structure – amino acid sequence
- Secondary structure – local structure such as $\alpha$ helix and $\beta$ sheets
- Tertiary structure – 3-dimensional structure of a protein monomer
- Quarternary structure – 3-dimensional structure of a fully functional protein (protein complexes).

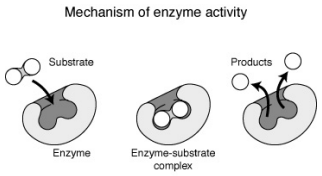
**Primary structure: residue sequence**

**Secondary structure:**
**local structures**
**(Helices, sheets, loops)**

**Tertiary structure:**
**position of each atom**

**Quaternary structure:**
**how groups of proteins pack together**




**The protein's 3-d shape determines what molecules it can bind to**

**Unsolved problem: predicting a protein's folding pattern based on its sequence**

## Cell Signaling and Biochemical Pathways

- **Surface receptors**
- **G-proteins, kinases, etc**
- **Transcription factors**
- **Other biochemical reactions – glycolysis, citric acid cycle, etc.**

## Molecular Biology Summary

- **Life and evolution**
- **Proteins**
- **Nucleic Acids**
- **Eukaryotes versus Prokaryotes**
- **Ribosome**
- **Translation**
- **Transcription**
- **Central dogma**
- **Genetic code**
- **DNA structure**
- **Chromosome**
- **Mitosis**