# Genetic Information: DNA Structure and Function

Umut Fahrioglu, PhD MSc

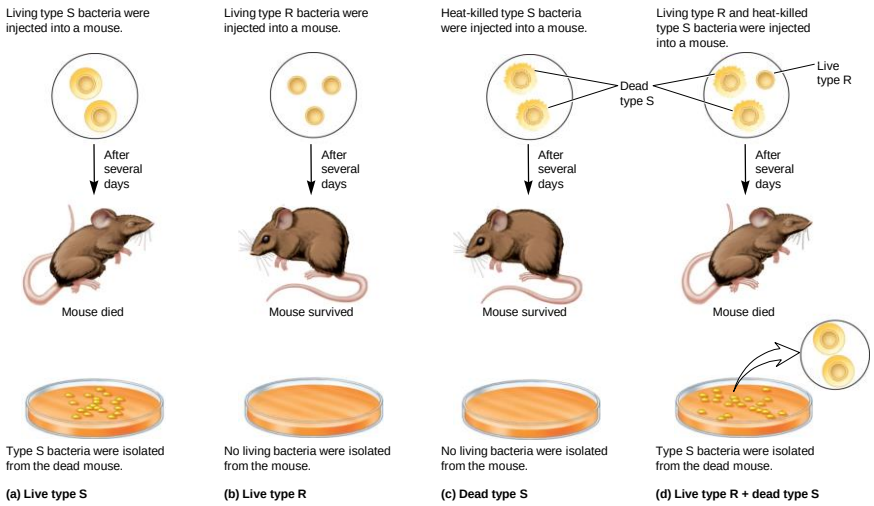
## Genetic material

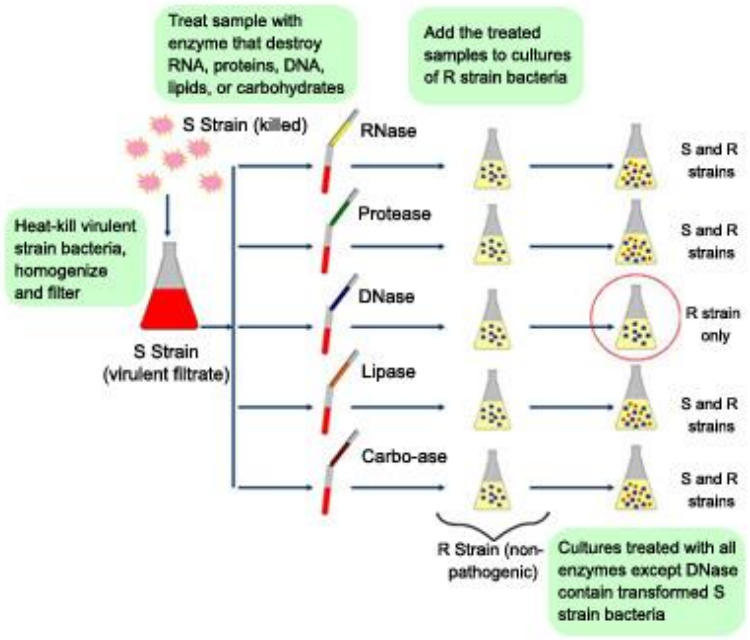
- There must be information stored in our cells such that when it is passed to new generation it influences the characteristic of each individual.
- This same information is also responsible for directing the many complex processes that lead an organism to an adult form. And obviously to keep the organism running properly.
- Until 1944, we were not clear on which chemical components of chromosomes made up the genes and counted as the genetic material. (It could have been proteins or nucleic acids since the chromosomes were known to have both. (Oswald, Avery, MacLeod and McCarty)
- Once the nucleic acid DNA was realized as the informational basis of heredity, we set out to determine its structure and unravel the mysteries that connect its structure to its function.
- In 1953, James Watson and Frances Crick put forth a hypothesis for the double helical nature of DNA.

- In 1928, Griffith conducted experiments using two strains of *S. pneumoniae*: type IIIS and type IIR
  - 1. Inject mouse with live type IIIS bacteria
    - Mouse died
    - Type IIIS bacteria recovered from the mouse's blood
  - 2. Inject mouse with live type IIR bacteria
    - Mouse survived
    - No living bacteria isolated from the mouse's blood
  - 3. Inject mouse with heat-killed type IIIS bacteria
    - Mouse survived
    - No living bacteria isolated from the mouse's blood
  - 4. Inject mouse with live type IIR + heat-killed type IIIS cells
    - Mouse died
    - Type IIIS bacteria recovered from the mouse's blood

9-6








# The Genetic Material: Four crucial characteristics

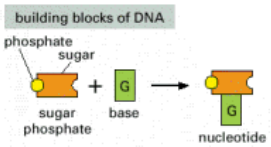
- REPLICATION: it is one of the most important aspects of the cell cycle and is therefore a fundamental property of all living organisms.
- STORAGE OF INFORMATION: this requires the molecule to act as a repository of genetic information regardless of whether it will be used in that cell.
- EXPRESSION OF INFORMATION: This is a complex process and it forms the basis for the information flow within the cell. The Central Dogma of Molecular Genetics.
- VARIATION BY MUTATION: Genetic material is a source of variability among organisms through the process of mutation. A mutation is a change in the chemical composition of DNA. It maybe passed to future generations.

## The Central Dogma of Molecular Biology


**Transcription** is carried out by **RNA polymerase**
**Translation** is performed on **ribosomes**
**Replication** is carried out by **DNA polymerase**
Reverse transcriptase copies RNA into DNA

## Nucleic Acids

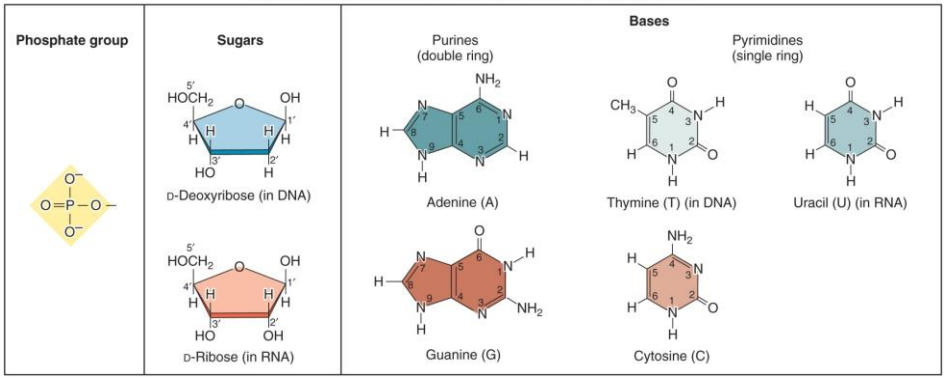
- First discovered in 1869 by Miescher.
- They were acid compounds found in the nuclei therefore they were named nucleic acids
- They contained C, N, O, and high amounts of P
- DNA is a nucleic acid and nucleotides are the building block of all nucleic acid molecules.
- A nucleotide is made up of three essential components: nitrogenous base, a pentose sugar and a phosphate group.



# Nucleic acids continued

- There are two kinds of nitrogenous bases
  - Nine-member double ring purines
  - Six member single ring pyrimidines
- Two types of purines and three types of pyrimidines are commonly found in nucleic acids
  - Purines are Adenine (A) and Guanine (G)
  - Pyrimidines are Cytosine (C), Thymine (T) and Uracil (U)
- Both DNA and RNA contain A, C and G but only DNA contains the base T and only RNA contains the base U.
- The pentose sugars found in nucleic acids give them their names
  - Ribonucleic acids (RNA) contain Ribose
  - Deoxyribonucleic acids (DNA) contain Deoxyribose
- If a molecule is composed of a base and a sugar it is called a **nucleoside.**
- If a phosphate group is added to the nucleoside, the molecule is now called a **nucleotide.**

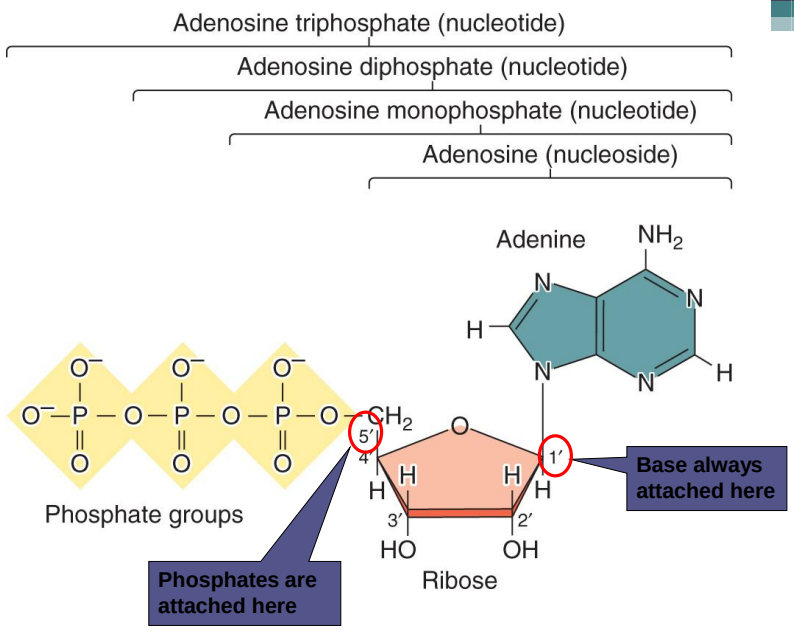




# Nomenclature

| BASE | NUCLEOSIDE | NUCLEOTIDE |
|---|---|---|
| | Deoxyribose sugar | Phosphate Added |
| PURINES:<br>Adenine<br>Guanine<br>Hypoxanthine | Adenosine<br>Guanosine<br>Inosine | Adenosine<br>Guanosine<br>Inosine |
| PYRIMIDINES:<br>Thymine<br>Cytosine | Thymidine<br>Cytidine | Thymidine<br>Cytidine |
| | Ribose sugar | |
| PYRIMIDINES:<br>Uracil | Uridine | Uridine |





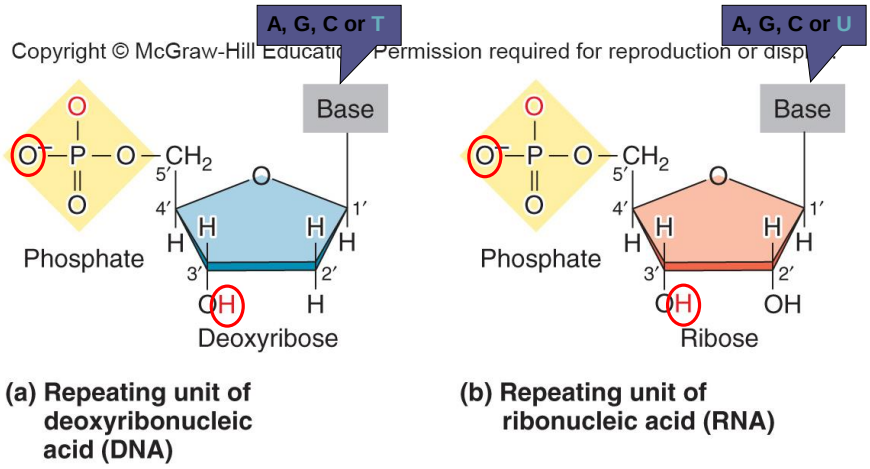
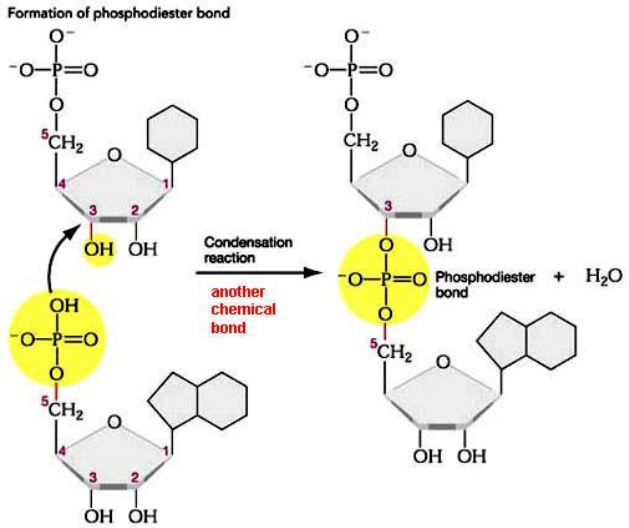




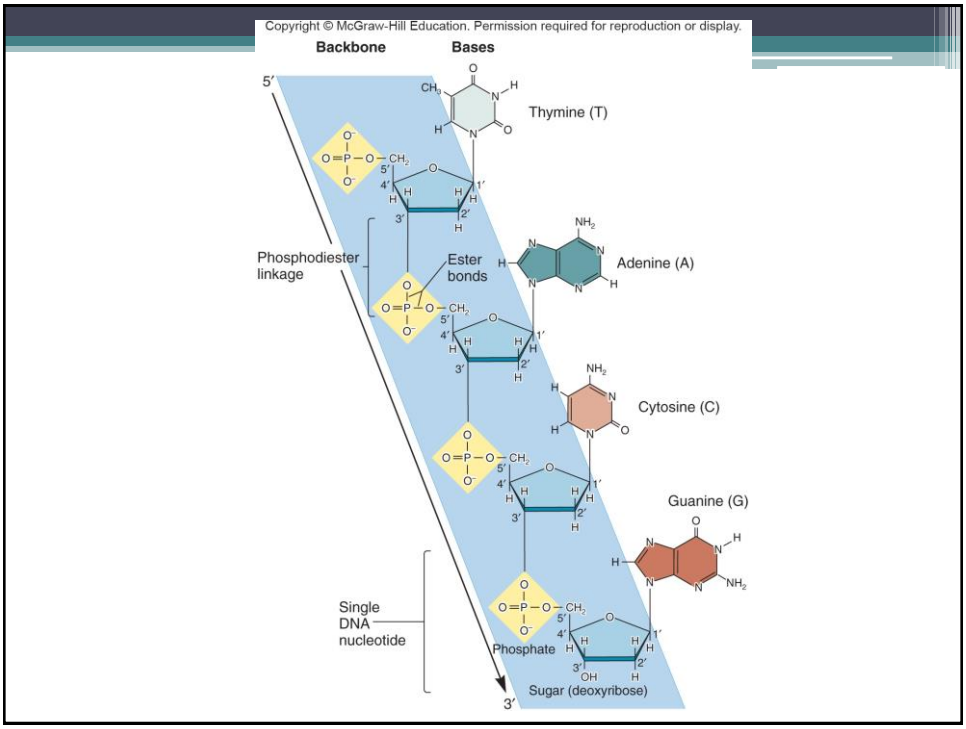

**Figure 9.8** **The structure of nucleotides found in (a) DNA and (b) RNA**

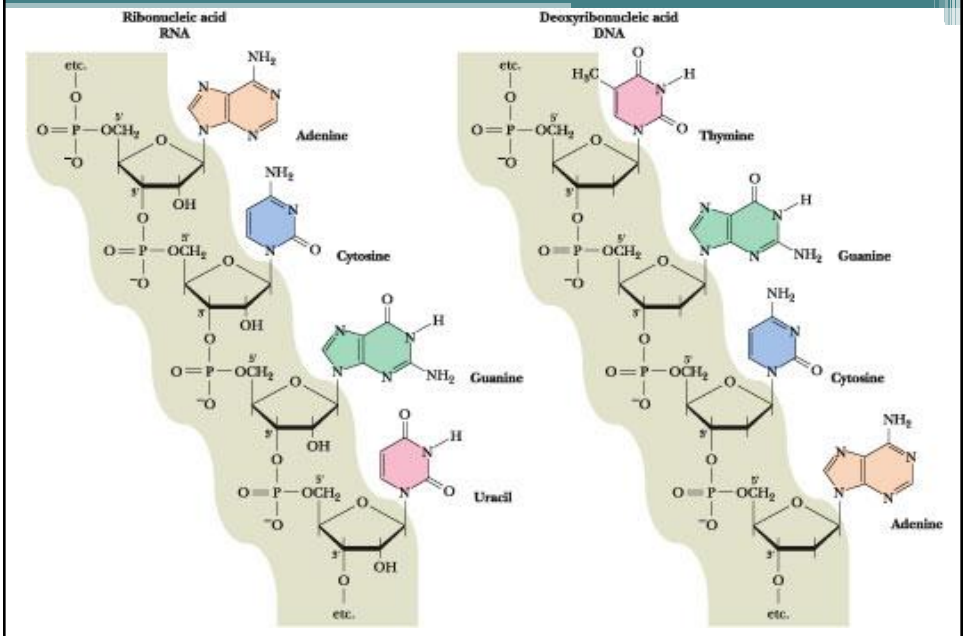
# Phosphodiester bond formation



# STRUCTURE OF A DNA STRAND

- Nucleotides are covalently linked together by phosphodiester bonds
  - A phosphate connects the 5’ carbon of one nucleotide to the 3’ carbon of another
- Therefore the strand has directionality
  - 5’ to 3’
  - In a strand, all sugar molecules are oriented in the same direction
- The phosphates and sugar molecules form the backbone of the nucleic acid strand
  - The bases project from the backbone






## Nucleases hydrolyze phosphodiester bonds

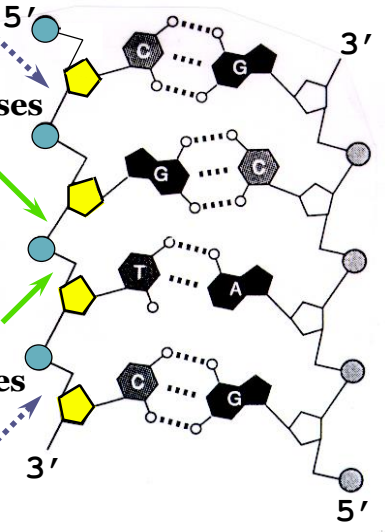
**Exonucleases cleave at terminal nucleotides.**

**e.g., proofreading exonucleases**

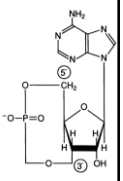
**Endonucleases cleave internally and can cut on either side of a phosphate leaving 5' phosphate or 3' phosphate ends depending on the particular endonuclease.**

**e.g., restriction endonucleases**

**Exonucleases cleave at terminal nucleotides.**



## Other functions of nucleotides



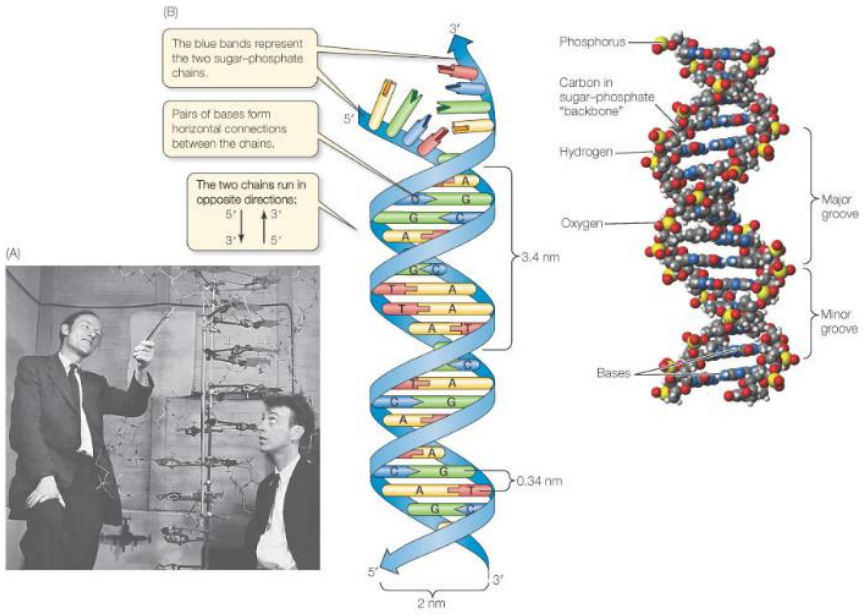
- Nucleotide 5'-triphosphates are carriers of energy (ATP)
- Bases serve as recognition units
- Cyclic nucleotides are signal molecules and regulators of cellular metabolism and reproduction
- ATP is central to energy metabolism
- GTP drives protein synthesis (responsible for binding of tRNA to the ribosome)
- CTP drives lipid synthesis (Glycerophospholipid syntheisis)
- UTP drives carbohydrate metabolism (UDP-glucose enters glycogen synthesis and UTP in metabolism of galactose)

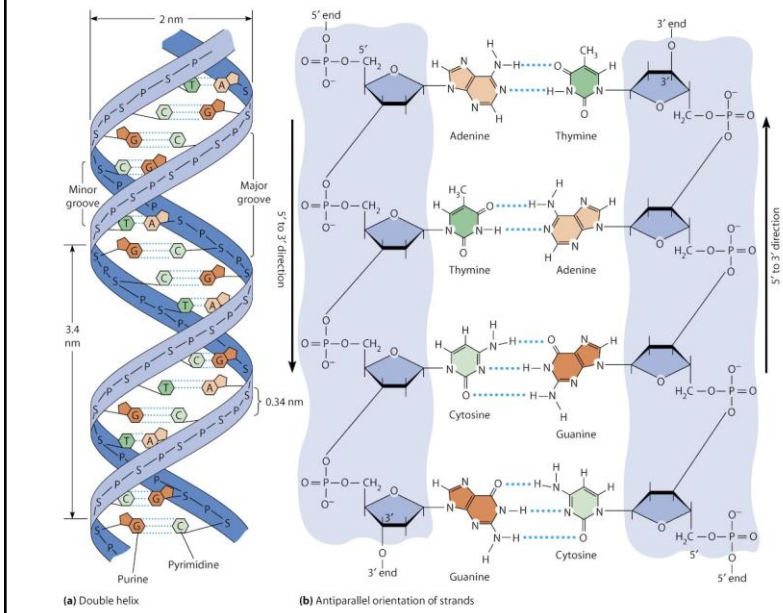
## Questions that came up about the DNA?

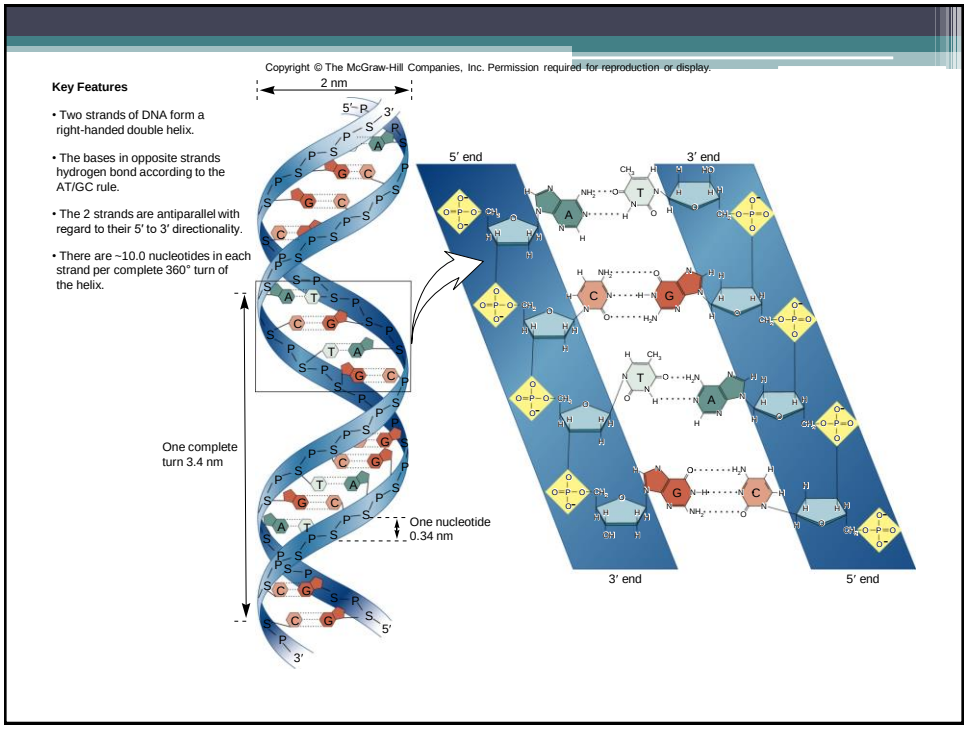
- How are the polynucleotides arranged into DNA?
- Is there one chain or more than one chain?
- If there is more than one chain, how do these chains relate to each other?
- Do the chains branch?
- How does the structure of the DNA relate to its various functions? (storage, replication, expression and mutation)
- How does the DNA serve as the genetic basis of life?
- The answer was believed to be in its chemical structure and organization

# The Watson and Crick Model of DNA

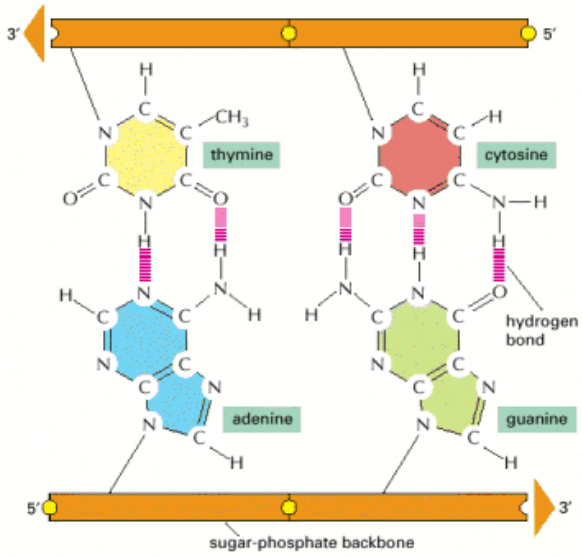
- Based on X-ray diffraction analysis and base-composition studies they came up with the following model
  1. Two long polynucleotide chains are coiled around a central axis, forming a right handed double helix.
  2. The chains are anti-parallel, that is their C-5' to C-3' orientations run in opposite directions.
  3. The bases of both chains are flat structures lying perpendicular to the axis: They are stacked on one another, 3.4 Å (0.34 nm) apart, on the inside of the double helix.
  4. The nitrogenous base of the opposite chains are paired as the result of the formation of hydrogen bonds; in DNA only A=T and G≡C pairs occur.
  5. Each complete turn of the helix is 34 Å (3.4 nm) long thus, each turn of the helix is the length of a series of 10 base pairs.
  6. A large major grove alternating with a smaller minor grove winds along the length of the molecule
  7. The double helix has a diameter of 20 Å (2.0 nm)

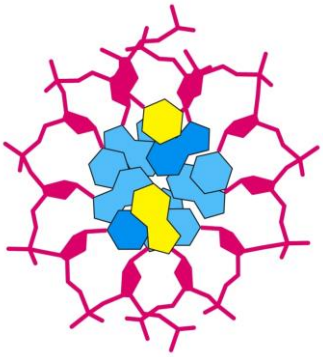


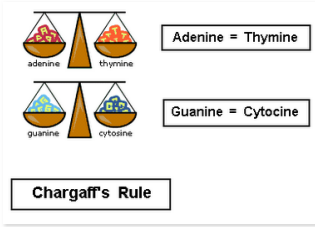

2 nm
Minor groove
Major groove
3.4 nm
0.34 nm
5′ to 3′ direction
Purine
Pyrimidine
5′ end
3′ end
Adenine
Thymine
Cytosine
Guanine
(a) Double helix
(b) Antiparallel orientation of strands
Copyright © 2009 Pearson Education, Inc.


Copyright © The McGraw-Hill Companies, Inc. Permission required for reproduction or display.
Key Features
• Two strands of DNA form a right-handed double helix.
• The bases in opposite strands hydrogen bond according to the AT/GC rule.
• The 2 strands are antiparallel with regard to their 5′ to 3′ directionality.
• There are ~10.0 nucleotides in each strand per complete 360° turn of the helix.
2 nm
One complete turn 3.4 nm
One nucleotide 0.34 nm
5′ end
3′ end

# Base pairing



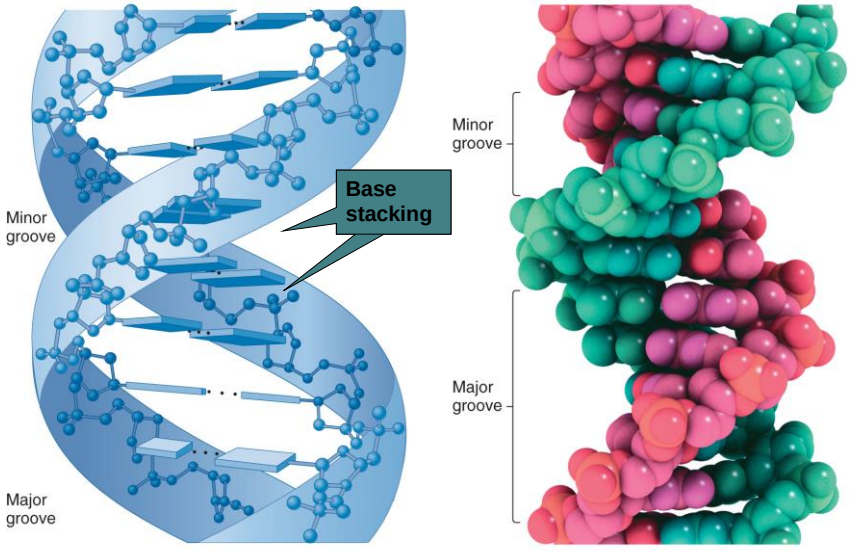



# Data that led to Chargaff's Rules

**Table 18-1** DNA Base Composition Data That Led to Chargaff's Rules

| | Number of Each Type of Nucleotide* | | | | Nucleotide Ratios** | | |
|---|---|---|---|---|---|---|---|
| Source of DNA | A | T | G | C | A/T | G/C | (A + T)/(G + C) |
| Bovine thymus | 28.4 | 28.4 | 21.1 | 22.1 | 1.00 | 0.95 | 1.31 |
| Bovine liver | 28.1 | 28.4 | 22.5 | 21.0 | 0.99 | 1.07 | 1.30 |
| Bovine kidney | 28.3 | 28.2 | 22.6 | 20.9 | 1.00 | 1.08 | 1.30 |
| Bovine brain | 28.0 | 28.1 | 22.3 | 21.6 | 1.00 | 1.03 | 1.28 |
| Human liver | 30.3 | 30.3 | 19.5 | 19.9 | 1.00 | 0.98 | 1.53 |
| Locust | 29.3 | 29.3 | 20.5 | 20.7 | 1.00 | 1.00 | 1.41 |
| Sea urchin | 32.8 | 32.1 | 17.7 | 17.3 | 1.02 | 1.02 | 1.85 |
| Wheat germ | 27.3 | 27.1 | 22.7 | 22.8 | 1.01 | 1.00 | 1.19 |
| Marine crab | 47.3 | 47.3 | 2.7 | 2.7 | 1.00 | 1.00 | 17.50 |
| *Aspergillus* (mold) | 25.0 | 24.9 | 25.1 | 25.0 | 1.00 | 1.00 | 1.00 |
| *Saccharomyces cerevisiae* (yeast) | 31.3 | 32.9 | 18.7 | 17.1 | 0.95 | 1.09 | 1.79 |
| *Clostridium* (bacterium) | 36.9 | 36.3 | 14.0 | 12.8 | 1.02 | 1.09 | 2.73 |

*The values in these four columns are the average number of each type of nucleotide found per 100 nucleotides in DNA.
**The A/T and G/C ratios are not all exactly 1.00 because of experimental error.



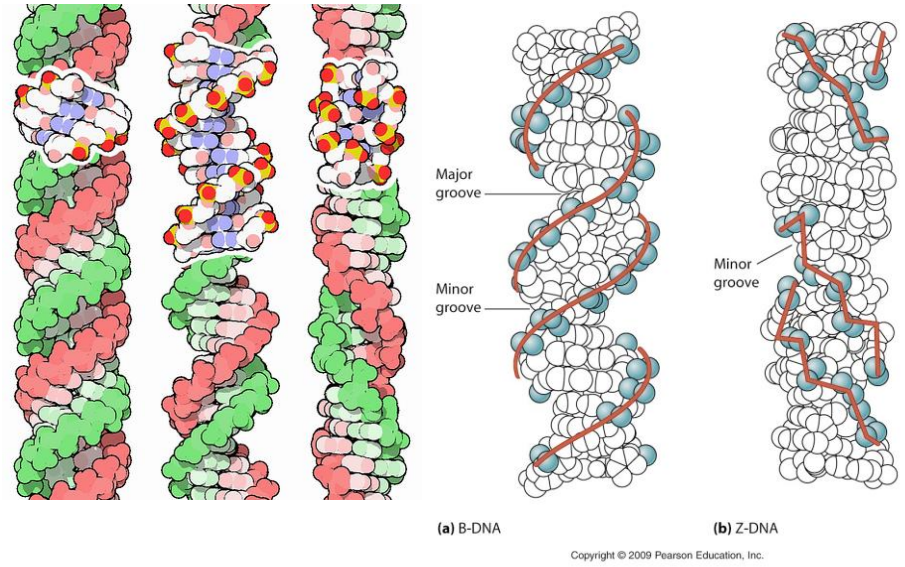

(a) Ball-and-stick model of DNA

(b) Space-filling model of DNA



# Alternative forms of DNA

- Under different conditions of isolation we can see different conformations of DNA. (initially A and B were known)
- B-DNA forms under aqueous, low salt conditions and is thought to be the biologically significant form.
- A-DNA is prevalent under high salt or dehydration conditions. It is slightly more compact. It is also right handed but the bases are tilted and displaced and it is probably not likely to be present *in vivo*.
- C-DNA forms under even higher dehydration conditions.
- D-DNA and E-DNA occur in helices lacking guanine in their base pair composition.
- P-DNA occurs when DNA is artificially stretched.
- Z-DNA is quite different, it is left-handed helix that is 18 Å in diameter with 9 base pairs per turn and has a zigzag configuration. Major grove is nearly eliminated in this form.
- Z-DNA occurs where there are alternating pyrimidines and purines (on one strand). The transition of B- to Z-DNA is facilitated by 5-methylcytosine.
- It is thought that these different forms might exist to accommodate different functions of the DNA

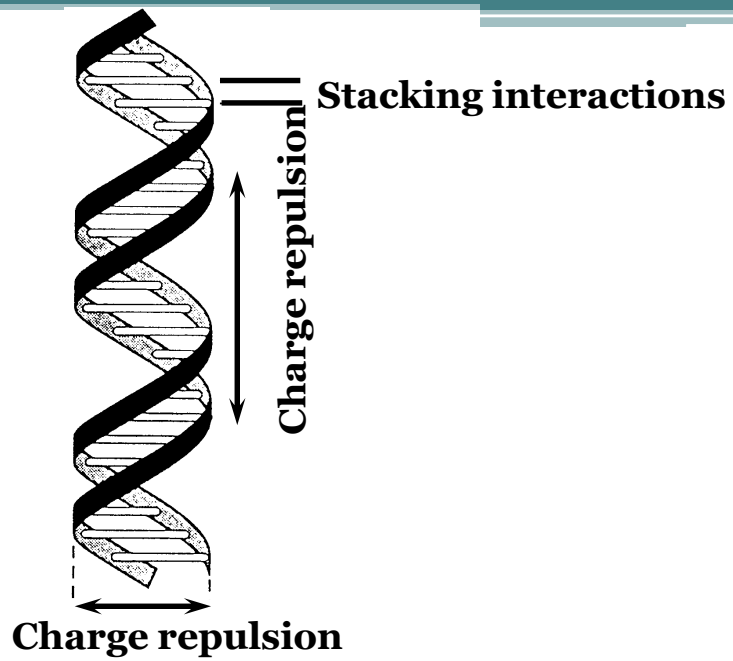



# Comparison of different DNA forms

| Form | Diameter | Bp/turn | Full Turn | Direction | Description |
|---|---|---|---|---|---|
| A | 2.2 nm | 11 | 2.5 nm | Right handed | Short and broad |
| B | 2.0 nm | 10 | 3.4 nm | Right handed | Longer and thinner |
| Z | 1.8 nm | 9 | 4.6 nm | Left Handed | Longest and thinnest |

## Forces affecting the stability of DNA

- hydrophobic interactions – stabilize
  - The hydrophobic environment inside with the bases and the hydrophilic environment outside with the sugar phosphate backbone
- stacking interactions – stabilize
  - relatively weak but additive van der Waals forces
- hydrogen bonding – stabilize
  - relatively weak but additive and facilitates the stacking of the bases
- electrostatic interactions – destabilize
  - contributed primarily by the (negative) phosphates
  - affect intrastrand and interstrand interactions
  - repulsion can be neutralized with positive charges (e.g., positively charged Na+ ions or proteins)



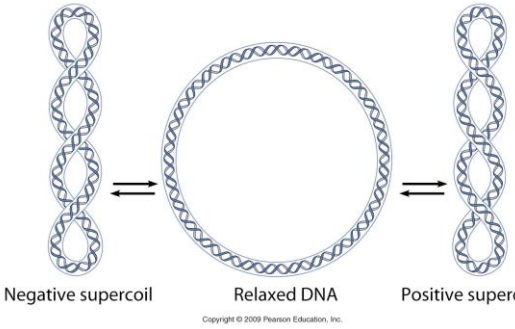
## DNA structure

- Primary (1°) Structure: Linear array of nucleotides
- Secondary (2°) Structure: the double helix
- Tertiary (3°) Structure: Super-coiling, stem-loop formation, cruciforms
- Quaternary (4°) Structure: Packaging into chromatin

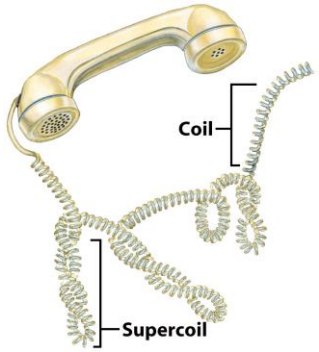
## Supercoiled DNA

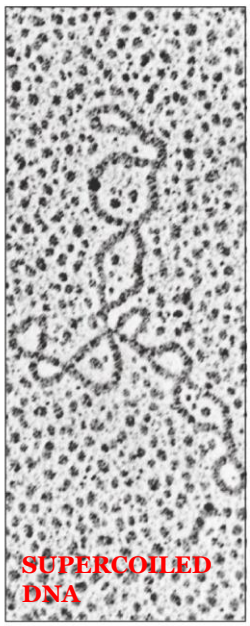
- In addition to helical configuration typical of all DNA molecules, a DNA can be twisted upon itself to form a new, higher-order helix giving rise to supercoiled DNA.
- In duplex DNA, ten bp per turn of helix (relaxed form)
- Over winding of DNA helix can be compensated by supercoiling.
- Supercoiling prevalent in circular DNA molecules and within local regions of long linear DNA strands.
- Positive supercoiling results from overwinding DNA and normally occurs during DNA replication.
- Negative supercoiling results from underwinding DNA and normally occurs in the nucleosome.
- Enzymes called topoisomerases or gyrases can introduce or remove supercoils
- In vivo most DNA is negatively supercoiled. Therefore, it is easy to unwind short regions of the molecule to allow access for enzymes
- Negative coiling can sometimes cause cruciform formation.

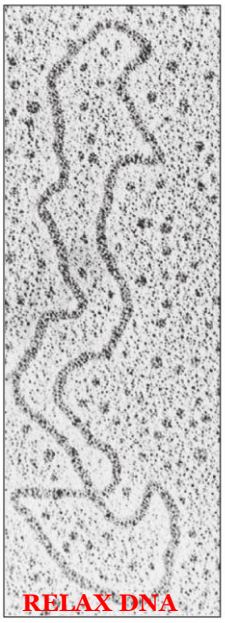
# Positive and Negative supercoil



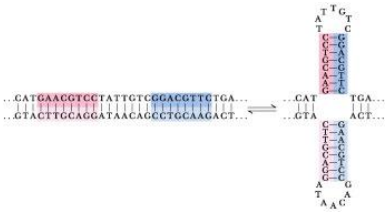




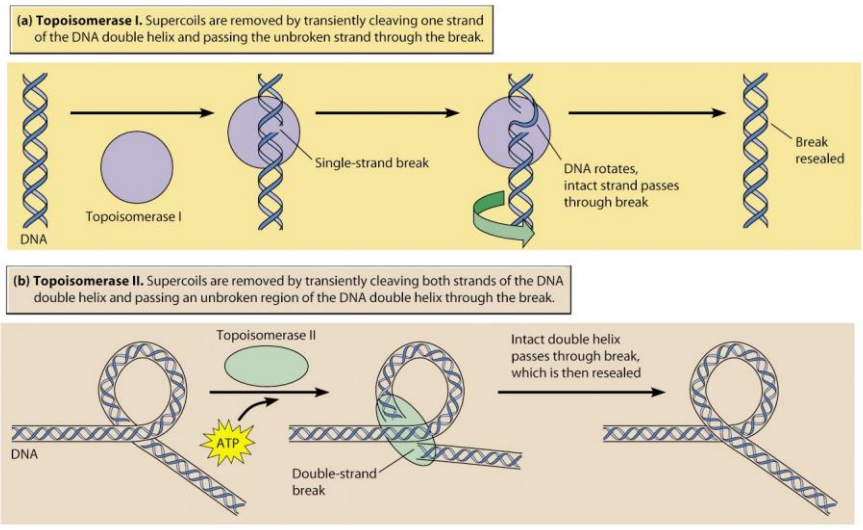





-Cruciform occur in palindromic regions of DNA.
-Can lead to base pairing within the same chain
-Promoted by negative supercoiling

# Topoisomerase I and topoisomerase II









# Base pairing during replication

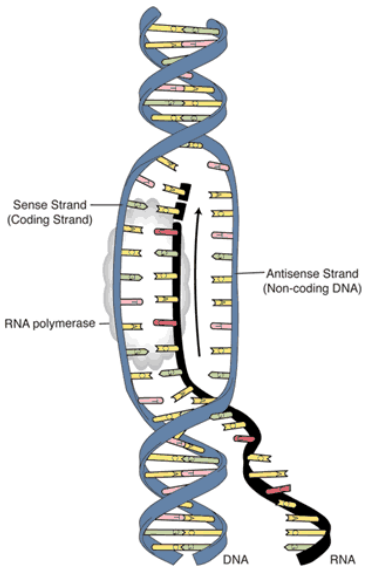
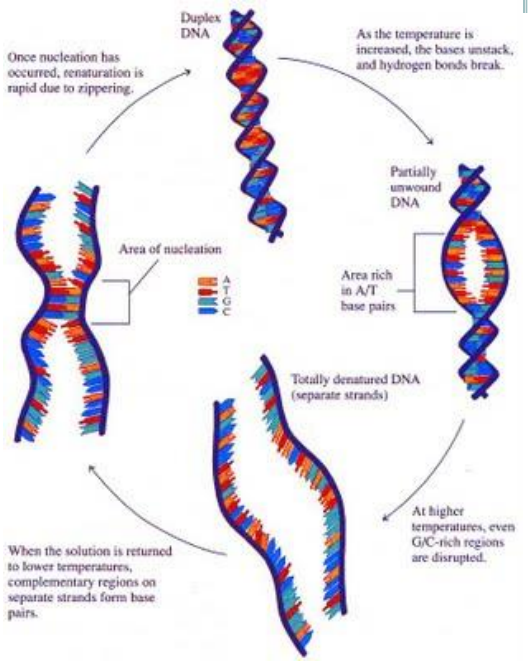


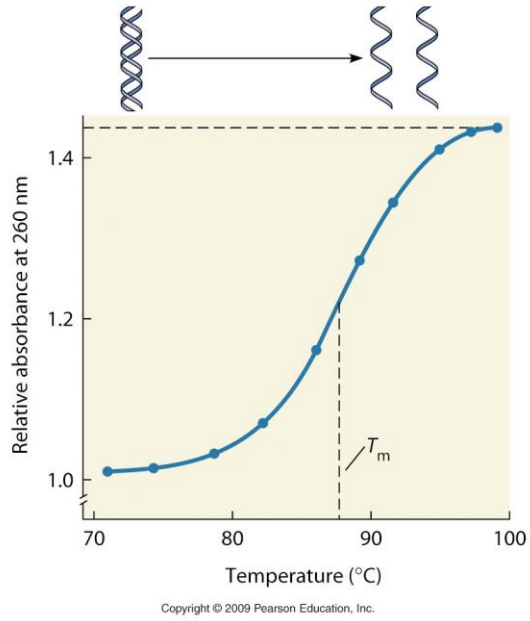

Image adapted from: National Human Genome Research Institute. Talking Glossary of Genetic Terms. Available at: www.genome.gov/Pages/Hyperion//DIR/VIP/Glossary/Illustration/antisense.shtml.

# Base pairing during RNA synthesis

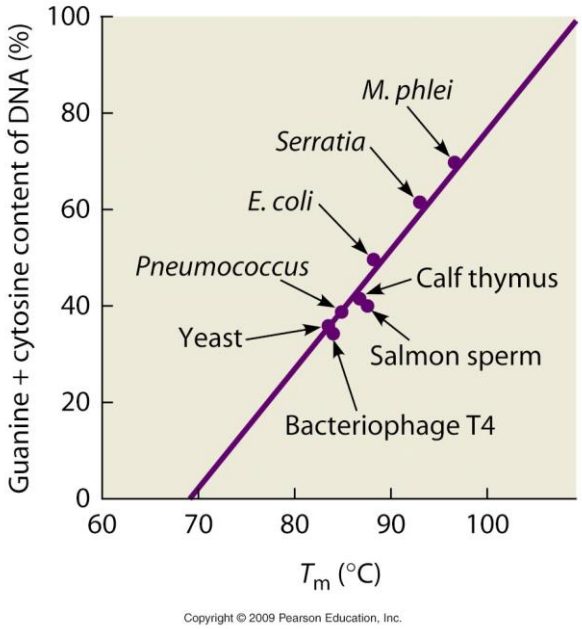
## Denaturation of DNA

- Extremes in pH or high temperatures cause the DNA to denature.
- A-T rich regions denature first.
- Cooperative unwinding of the DNA strand
- Can determine degree of denaturation by measuring absorbance at 260 nm.
- Increased single strandedness causes increase in absorbance
- Base stacking causes less absorbance

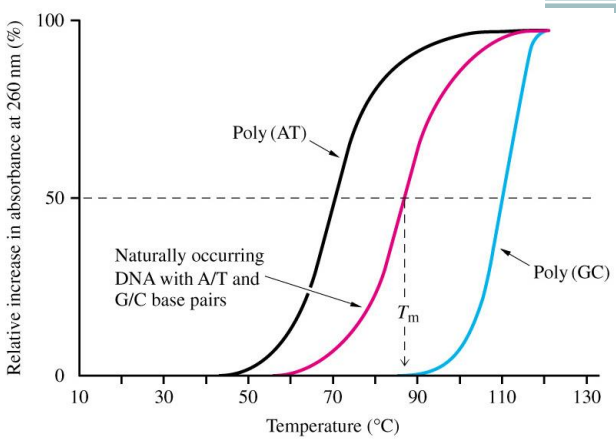




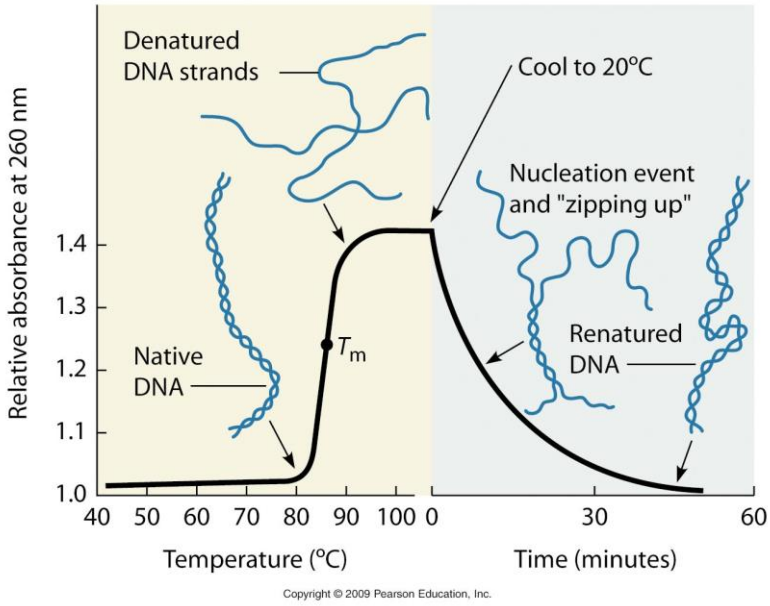

Melting temperature and UV absorbance




Melting temperature goes up as the G-C content goes up




-Melting temperature related to G:C and A:T content.
-3 H-bonds of G:C pair require higher temperatures to denture than 2 H-bonds of A:T pair.




Heat DNA and allow to cool down

# Some nomenclature

- **dsDNA:** double stranded DNA
- **ssDNA:** single stranded DNA
- A standard unit of size in DNA is **kilobase (kb)**
  - **1 kb = 1000 bp**
  - **1 Mb = 1,000,000 bp**
- One thousand kilobases is a **megabase (Mb)**
- **Genotype**: An organism's genetic constitution.
- **Phenotype:** The observed characteristics of an organism, as determined by the genetic makeup (and the environment
- **n** = number of chromosomes in a haploid genome
- **2n** = number of chromosomes in a diploid genome

**Table 18-2** Examples of Sequenced Genomes*

| Organism | Genome Size | Estimated Gene Number |
|---|---|---|
| Bacteria | | |
| *Mycoplasma genitalium* | 0.6 Mb | 470 |
| *Haemophilus influenza* | 1.8 Mb | 1,740 |
| *Streptococcus pneumoniae* | 2.2 Mb | 2,240 |
| *Escherichia coli* | 4.6 Mb | 4,400 |
| Yeast (*S. cerevisiae*) | 12.1 Mb | 6,200 |
| Roundworm (*C. elegans*) | 97 Mb | 19,700 |
| Mustard plant (*A. thaliana*) | 125 Mb | 25,500 |
| Fruit fly (*D. melanogaster*) | 180 Mb | 13,600 |
| Rice (*O. sativa*) | 389 Mb | 37,500 |
| Mouse (*Mus musculus*) | 2500 Mb | 25,000 |
| Human (*H. sapiens*) | 3200 Mb | 25,000 |

*As of October 2007, complete genome sequences had been published for 657 organisms (535 bacteria, 44 archaea, and 78 eukaryotes).

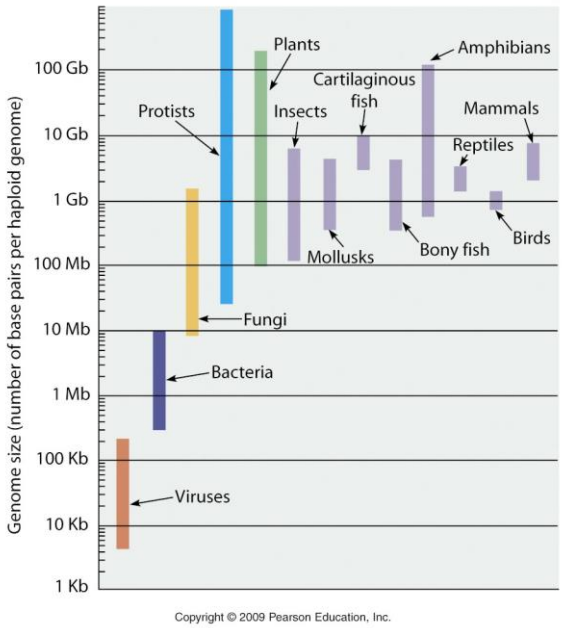




**Table 4–1 Some Vital Statistics for the Human Genome**

| | HUMAN GENOME |
|---|---|
| DNA length | $3.2 \times 10^9$ nucleotide pairs* |
| Number of genes | approximately 25,000 |
| Largest gene | $2.4 \times 10^6$ nucleotide pairs |
| Mean gene size | 27,000 nucleotide pairs |
| Smallest number of exons per gene | 1 |
| Largest number of exons per gene | 178 |
| Mean number of exons per gene | 10.4 |
| Largest exon size | 17,106 nucleotide pairs |
| Mean exon size | 145 nucleotide pairs |
| Number of pseudogenes** | more than 20,000 |
| Percentage of DNA sequence in exons (protein coding sequences) | 1.5% |
| Percentage of DNA in other highly conserved sequences*** | 3.5% |
| Percentage of DNA in high-copy repetitive elements | approximately 50% |

* The sequence of 2.85 billion nucleotides is known precisely (error rate of only about one in 100,000 nucleotides). The remaining DNA primarily consists of short highly repeated sequences that are tandemly repeated, with repeat numbers differing from one individual to the next.

** A pseudogene is a nucleotide sequence of DNA closely resembling that of a functional gene, but containing numerous mutations that prevent its proper expression. Most pseudogenes arise from the duplication of a functional gene followed by the accumulation of damaging mutations in one copy.

*** Preserved functional regions; these include DNA encoding 5′ and 3′ UTRs (untranslated regions), structural and functional RNAs, and conserved protein-binding sites on the DNA.