DNA is the Chemical Basis of Heredity.
[Avery, MacLeod, and McCarthy, 1944]

Wild-type *Pneumonococcus* (Smooth phenotype –S), Virulent

Heat-killed *Pneumonococcus* (Smooth phenotype-S), Avirulent

Mutant strain of *Pneumonococcus* (Rough phenotype –R), Avirulent

Infect mouse with mixture of heat-killed S-strain and live R-strain of *Pneumonococcus*. None of them can kill mice on their own.

Mouse dies

Dead mouse has live, smooth bacteria

Extract DNA and add to R-type bacteria

DNA transforms R-bacteria to S-bacteria
Nucleic acids are required for the storage and expression of genetic information.

- All visible properties and functions of a cell (phenotype) are determined by proteins.

- The structure (and therefore the function) of proteins is determined by their primary amino acid sequence.

- The amino acid sequence of the proteins is determined by the nucleotide sequence of DNA.
Chemical structure of the DNA molecule: 
the building blocks of DNA

- DNA is a polymer of deoxyribonucleoside monophosphates.
- Each deoxyribonucleoside monophosphate contains:
  - A pentose sugar (2-deoxyribose)
  - Phosphate
  - One heterocyclic base, which can be either a purine (adenine or guanine), or a pyrimidine (thymine or cytosine).
- The C atoms of the 2-deoxyribose are numbered with a prime (‘) in order to be distinguished from the C or N atoms of the bases.
Chemical structure of the DNA molecule: the primary structure

The backbone of the DNA molecule is formed by alternating phosphate and 2-deoxyribose residues held by phosphodiester bonds.

The DNA molecule has polarity. By convention, the 5’ -terminus is written at the left end of the sequence and the 3’ -terminus at the right end.

Variability of DNA structure: $4^n$ sequence possibilities exist for a sequence of $n$ nucleotides.

Deoxyribonucleases (endonucleases and exonucleases hydrolytically cleave the phosphodiester bonds)
Secondary structure of the DNA molecule: The Watson-Crick double helix.

- The two strands of the double helix have opposite polarity.
- The hydrophilic deoxyribose-phosphate backbones of the two DNA strands form two ridges on the surface of the molecule. Phosphate groups have a negative charge at physiologic pH.
- Bases in opposite strands interact by hydrogen bonds.
- The double-helix structure is stabilized by the hydrogen bonds plus the hydrophobic interactions between the stacked bases.
- The hydrophobic bases are stacked flat on top of each other inside the molecule, creating a major and a minor groove.
- These grooves provide binding to regulatory proteins.
- Anticancer drugs (dactinomycin) intercalate into the minor groove interfering with DNA or RNA synthesis.
The interactions between the bases in opposite DNA strands are highly specific:

- Adenine always pairs with thymine; guanine always pairs with cytosine.
- Each A-T base pair is held together by two hydrogen bonds; each C-G pair by three.
- Double-stranded DNA molecules always contain equal molar amounts of purines (A and T), and equal molar amounts of pyrimidines (G and C) ("Chargaff Rule").
- Each polynucleotide chain of the DNA double–helix is always the complement of the other.
Denaturation of the double-helix DNA molecule

- The two DNA strands can be separated when the hydrogen bonds between the paired bases are disrupted.
- Denaturation of DNA can occur under the following conditions:
  - Increased temperature
  - Changes in the pH
  - Decreased salt concentration
  - Presence of denaturing agents, such as formamide
- Denaturation begins in areas with high A-T content. G-C pairs (three hydrogen bonds) are more difficult to separate than A-T pairs (two hydrogen bonds).
- Single-stranded DNA has higher (~37%) absorbance at OD 260nm than double-stranded DNA, due to loss of the base stacking interactions.
- Renaturation (annealing) of the DNA molecule occurs spontaneously when cooled down slowly (5°C to 20° C below the Tm).
### Structural forms of the double-helix

<table>
<thead>
<tr>
<th>Property</th>
<th>A-DNA</th>
<th>B-DNA</th>
<th>Z-DNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Helix sense</td>
<td>Right</td>
<td>Right</td>
<td>Left</td>
</tr>
<tr>
<td>Base pairs/turn</td>
<td>11</td>
<td>10</td>
<td>12</td>
</tr>
<tr>
<td>Rise/base pair*</td>
<td>2.6 Å</td>
<td>3.4 Å</td>
<td>3.7 Å</td>
</tr>
<tr>
<td>Helix pitch*</td>
<td>28.6 Å</td>
<td>34 Å</td>
<td>44.4 Å</td>
</tr>
<tr>
<td>Widest diameter</td>
<td>25.5 Å</td>
<td>23.7 Å</td>
<td>18.4 Å</td>
</tr>
<tr>
<td>Major groove</td>
<td>Narrow, very deep</td>
<td>Wide and deep</td>
<td>Wide and flat</td>
</tr>
<tr>
<td>Minor groove</td>
<td>Wide and shallow</td>
<td>Narrow and deep</td>
<td>Very narrow</td>
</tr>
</tbody>
</table>

*One angstrom (Å) is 0.1 nm or $10^{-10}$ m*
The chromosome of a typical bacterium consists of a single, double-stranded DNA molecule (*E. coli*: $4.6 \times 10^6$ bp), usually associated with proteins.

The bacterial chromosome contains only one copy of each gene (*haploid*) (*E. coli*: 4,300 genes encoding proteins and 115 genes for RNA).

Every gene is precisely colinear with the amino acid sequence (or RNA) it encodes.

Many bacteria additionally contain small, circular extrachromosomal DNA molecules, called *plasmids*. 
Plasmids are small, circular, extrachromosomal DNA molecules, present in bacteria and some lower eukaryotes.

- Plasmid size can vary from 2,000 to 200,000 bp.
- Copy number may vary.
- Genes carried on plasmids are usually not essential for the survival of the cells.
- Plasmids are self-replicating entities. Their replication may or may not be synchronized to chromosomal division.
Biological importance of plasmids

- Gene products encoded by plasmids are usually **antibiotic resistance** determinants (R factors), toxin production, ability to degrade unusual substrates.
- Plasmids can be transferred from one cell to another, some of them independently via conjugation (e.g., the F “fertility” factor of *E. coli*).
- For these reasons, plasmids are involved in the development of bacterial strains that are resistant to particular classes of antibiotics.
- Frequent prescription of low doses of antibiotics, if this is not strictly indicated, provides the selective pressure for the emergence of resistant strains and should be avoided.
- Plasmids are also important tools in molecular biology and the recombinant technology, because they can serve as **vectors** for gene cloning.
Organization of the Human Genome

- The haploid human genome:
  - contains ~3.9 billion base pairs
  - Has a stretched-out length of 130 cm
  - Is packaged in 23 chromosomes
- Each chromosome contains one, linear, double stranded DNA molecule, which is several cm long, and contains 50,000-100,000 genes, plus non-coding DNA.
In a nondividing eukaryotic cell the chromosomal material (chromatine) appears to be amorphous, and dispersed throughout the nucleus.

Chromatin is a nucleoprotein complex consisting of:
- DNA (~50%)
- Histones
- Non-histone proteins (proteins involved in DNA replication and in transcription)
- Small amounts of RNA

There are two types of chromatin:
- Euchromatin, less densely packed, genetically expressed
- Heterochromatin, more condensed, deeper stained, not genetically expressed
Histones and the formation of nucleosomes

- Histones are small basic proteins which associate with the negatively charged DNA via electrostatic interactions. Their function is to condense the DNA.

- There are five types of histones in eukaryotic cells:
  - H3 and H4: highly conserved. Arginine-rich. They form tetramers containing two molecules of each: (H3-H4)₂.
  - H1: not very conserved among species (tissue-specific and species-specific. Do not bind to chromatin tightly. Help packing the nucleosomes into tighter structures.

- Nucleosome formation:
  - Histones form octamers consisting of one (H3-H4)₂ tetramer plus two (H2A-H2B) dimers.
  - DNA (about 146 bp) wraps around the histone octamer, making 1.75 superhelical turns.

Nucleosomes are connected by a variable length of DNA (50-60 bp) which is associated with a single H1.
Higher order structures contribute to the further compaction of chromatin

- **10-nm fibril**: nucleosomes arranged with their flat surfaces parallel to the fibril axis.

- **30-nm chromatin fiber**: 6 to 7 nucleosomes per turn, with their flat surfaces parallel to each other. Stabilized by H1.

- Chromatin fibers form **loops (domains)** (30,000 to 100,000 pb) anchored in a supporting matrix (scaffold). Each loop may correspond to a separate genetic function.
DNA Supercoiling

- Supercoiling is an important intrinsic property of the DNA structure which allows the high compaction of the DNA. It is a highly regulated process.
- DNA is coiled in the form of a double helix, in which both strands of the DNA coil around an axis. The further coiling of that axis upon itself produces DNA supercoiling.
- When there is no bending of the DNA axis upon itself, the DNA is said to be in a relaxed state.
Effects of negative supercoiling (underwining)
Topoisomerases increase or decrease the extend of DNA underwinding.

**Type I topoisomerases:**
- cut a single strand of the double helix
- have both nuclease and ligase activities
- do not require ATP
- relaxes negative supercoils in bacteria, negative and positive supercoils in eukaryotic cells.

**Type II topoisomerases:**
- bind to both strands of the double helix
- relaxes both negative and positive supercoils in prokaryotic and eukaryotic cells
- required for the separation of the interlocked molecules of DNA following chromosomal replication
- do not require ATP

**DNA gyrase:**
- Unusual type II topoisomerase found in *E. coli.*
- introduces negative supercoils in to resting circular DNA, facilitating future replication
- requires ATP
- inactivated by quinolones
Many eukaryotic genes contain intervening non-transcribed sequences

- The coding regions of the eukaryotic genom (exons) are interrupted by large intervening sequences of non-coding DNA (introns).
- Introns may be important in the evolution of biological function, because they separate functional domains of coding information allowing genetic rearrangements to occur more rapidly through recombination.
- Only ~ 100,000 proteins are expressed in human cells...that implies that much of the human DNA is not expressed.
Significant portion of eukaryotic DNA consists of Repetitive Sequences

- Highly repetitive sequences (simple-sequence DNA or “satellite” DNA
  - 5-500 bp long, repeated millions of times per cell
  - Usually clustered in the centromeres and telomeres
  - Transcriptionally inactive (may play structural role)
- Moderately repetitive sequences
  - Repeated less than $10^6$ times
  - Not clustered, usually transcribed
  - Most arise from transposition (retroposons)
    - **Long interspersed repeat sequences (LINEs)**
      - Species specific, 6-7 kbp long, 20-50 thousand copies in mammalian genomes
    - **Short interspersed repeat sequences (SINEs)**
      - 70-300 bp long, may be repeated 100,000 times
      - Best characterized moderately repetitive sequence in humans is the “**Alu family**”. Comprises 1-3% of the total DNA. Most of them are “mobile” elements.
Single Nucleotide Polymorphisms (SNPs, “snips”) 

- DNA sequence variations that occur when a single nucleotide in the genome sequence is altered (e.g., AGGCTAA to ATGGCTAA)
- Must occur in at least 1% of the population
- Occur every 100 to 300 bases in the human genome.
- May occur in both coding and noncoding regions of the genome.
- Each person’s genetic material contains a unique SNP pattern.
- Most of SNPs have no effect on cell function, but others could predispose to disease or influence response to a drug.
Mitochondrial DNA

- In typical somatic cells, mitochondria contain less than 0.1% of the cell’s DNA.
- It encodes for mitochondrial tRNAs, rRNAs, and few mitochondrial proteins (less than 5% of them).
- Mitochondrial DNA gets replicated before each mitochondrial division, which occurs before cell division.
- Mitochondria may have evolved from ancient bacteria, which gained access to the cytoplasm of host cells.
- A number of diseases, including some myopathies, neurologic disorders and some cases of diabetes melitus are due to mutations in mt DNA.
Viral genome

- Viruses contain generally much less DNA than cells, because they use the resources of the host cell for their propagation.
- Many viruses, including almost all the ones that infect plants, have single-stranded, RNA genomes.
- The shape of the viral genome changes during replication within the host: many linear DNAs become circular and all single-stranded DNAs become double-stranded.