

Chapter 5 - The Structure and Function of Macromolecules

Overview: The Molecules of Life

- Within all cells, small organic molecules are joined together to form larger molecules.
- All living things are made up of four main classes of **macromolecules**: carbohydrates, lipids, proteins, and nucleic acids.
- These large macromolecules may consist of thousands of covalently bonded atoms, some with mass greater than 100,000 daltons.
- Biochemists have determined the detailed structures of many macromolecules, which exhibit unique emergent properties arising from the orderly arrangement of their atoms.

Concept 5.1 Most macromolecules are polymers, built from monomers.

- Three of the four classes of macromolecules—carbohydrates, proteins, and nucleic acids—form chain-like molecules called polymers.

A **polymer** is a long molecule consisting of many similar or identical building blocks linked by covalent bonds.

The repeated units are small molecules called **monomers**.

Some of the molecules that serve as monomers have other functions of their own.

- The chemical mechanisms that cells use to make and break polymers are similar for all classes of macromolecules.
- Monomers are connected by covalent bonds that form through the loss of a water molecule.

This reaction is called a **condensation reaction** or **dehydration reaction**.

When a bond forms between two monomers, each monomer contributes part of the water molecule that is lost. One monomer provides a hydroxyl group (—OH), while the other provides a hydrogen atom (—H).

Cells invest energy to carry out dehydration reactions.

Dehydration is facilitated by enzymes, specialized macromolecules that speed up chemical reactions in cells.

- The covalent bonds that connect monomers in a polymer are disassembled by **hydrolysis**, a reaction that is effectively the reverse of dehydration.

In hydrolysis, bonds are broken by the addition of water molecules. A hydrogen atom attaches to one monomer, and a hydroxyl group attaches to the adjacent monomer.

We take in food as organic polymers that are too large for our cells to absorb. In the digestive tract, enzymes direct the hydrolysis of specific polymers. The resulting monomers are absorbed by the cells lining the gut and transported to the bloodstream for distribution to body cells.

The cells of our body then use dehydration reactions to assemble the monomers into new and different polymers that carry out functions specific to the particular cell type.

- *An immense variety of polymers can be built from a small number of monomers.*
- Each cell has thousands of different kinds of macromolecules.
- Macromolecules vary among cells of the same individual. They vary more among unrelated individuals of a species, and even more between species.
- This diversity comes from various combinations of the 40–50 common monomers and some others that occur rarely.

These monomers can be connected in a great many combinations, just as the 26 letters in the alphabet are used to create a whole dictionary of words.

- Despite the great diversity in organic macromolecules, members of each of the four major classes of macromolecules are similar in structure and function.

Concept 5.2 Carbohydrates serve as fuel and building material.

- **Carbohydrates** include sugars and their polymers.

The simplest carbohydrates are monosaccharides, or simple sugars.

Disaccharides, or double sugars, consist of two monosaccharides joined by a condensation reaction.

Polysaccharides are polymers of many monosaccharides.

- **Sugars, the smallest carbohydrates, serve as fuel and a source of carbon.**

- **Monosaccharides** generally have molecular formulas that are some multiple of the unit CH_2O .

For example, glucose has the formula $\text{C}_6\text{H}_{12}\text{O}_6$.

- Monosaccharides have a carbonyl group ($>\text{C}=\text{O}$) and multiple hydroxyl groups ($-\text{OH}$).

Depending on the location of the carbonyl group, the sugar is an aldose (aldehyde sugar) or a ketose (ketone sugar).

Most names for sugars end in *-ose*.

Glucose, an aldose, and fructose, a ketose, are structural isomers.

- Monosaccharides are also classified by the number of carbon atoms in the carbon skeleton.

The carbon skeleton of a sugar ranges from three to seven carbons long.

Glucose and other six-carbon sugars are hexoses.

Five-carbon sugars are pentoses; three-carbon sugars are trioses.

- Another source of diversity for simple sugars is the spatial arrangement of their parts around asymmetric carbon atoms.

For example, glucose and galactose, both six-carbon aldoses, differ only in the spatial arrangement of their parts around asymmetric carbons.

- Although glucose is often drawn with a linear carbon skeleton, most sugars (including glucose) form rings in aqueous solution.

- Monosaccharides, particularly glucose, are major nutrients for cellular work.

Cells extract energy from glucose molecules in the process of cellular respiration.

- Simple sugars also function as the raw material for the synthesis of other monomers, such as amino acids and fatty acids.

- Two monosaccharides can join with a **glycosidic linkage** to form a **disaccharide** via dehydration.

Maltose, malt sugar, is formed by joining two glucose molecules.

Sucrose, table sugar, is formed by joining glucose and fructose. Sucrose is the major transport form of sugars in plants.

Lactose, milk sugar, is formed by joining glucose and galactose.

- **Polysaccharides, the polymers of sugars, have storage and structural roles.**

- **Polysaccharides** are polymers of hundreds to thousands of monosaccharides joined by glycosidic linkages.

Some polysaccharides serve for storage and are hydrolyzed as sugars are needed.

Other polysaccharides serve as building materials for the cell or the whole organism.

- **Starch** is a storage polysaccharide composed entirely of glucose monomers.

- Plants store surplus glucose as starch granules within plastids, including chloroplasts, and

withdraw it as needed for energy or carbon.

Animals that feed on plants, especially parts rich in starch, have digestive enzymes that can hydrolyze starch to glucose, making the glucose available as a nutrient for cells. Most of the glucose monomers in starch are joined by 1–4 linkages (number 1 carbon to number 4 carbon).

The simplest form of starch, amylose, is unbranched.

Branched forms such as amylopectin are more complex.

- Animals store glucose in a polysaccharide called **glycogen**.

Glycogen is highly branched, like amylopectin.

Humans and other vertebrates store a day's supply of glycogen in the liver and muscles.

- **Cellulose** is a major component of the tough walls of plant cells.

Plants produce almost 100 billion tons of cellulose per year. It is the most abundant organic compound on Earth.

- Like starch, cellulose is a polymer of glucose. However, the glycosidic linkages in these two polymers differ.

The linkages are different because glucose has two slightly different ring structures.

These two ring forms differ in whether the hydroxyl group attached to the number 1 carbon is fixed above (α glucose) or below (β glucose) the plane of the ring.

- Starch is a polysaccharide of α () glucose monomers.
- Cellulose is a polysaccharide of β () glucose monomers, making every other glucose monomer upside down with respect to its neighbors.
- The differing glycosidic linkages in starch and cellulose give the two molecules distinct three-dimensional shapes.

While polymers built with α glucose form helical structures, polymers built with β glucose form straight structures.

The straight structures built with β glucose allow H atoms on one strand to form hydrogen bonds with OH groups on other strands.

In plant cell walls, parallel cellulose molecules held together in this way are grouped into units called microfibrils, which form strong building materials for plants.

Cellulose microfibrils are important constituents of wood, paper, and cotton.

- The enzymes that digest starch by hydrolyzing its linkages cannot hydrolyze the linkages in cellulose.

Cellulose in human food passes through the digestive tract and is eliminated in feces as “insoluble fiber.”

As it travels through the digestive tract, cellulose abrades the intestinal walls and stimulates the secretion of mucus, which aids in the passage of food.

- Some microbes can digest cellulose to its glucose monomers through the use of cellulase enzymes.
- Many eukaryotic herbivores, from cows to termites, have symbiotic relationships with cellulose-digesting prokaryotes, providing the prokaryote and the host animal access to a rich source of energy.

Some fungi can also digest cellulose.

- Another important structural polysaccharide is **chitin**, found in the exoskeletons of arthropods (including insects, spiders, and crustaceans).

Chitin is similar to cellulose, except that it has a nitrogen-containing appendage on each glucose monomer.

Pure chitin is leathery but can be hardened by the addition of calcium carbonate.

Chitin also provides structural support for the cell walls of many fungi.

Concept 5.3 Lipids are a diverse group of hydrophobic molecules.

- Unlike other macromolecules, lipids do not form polymers.
- The unifying feature of **lipids** is that they have little or no affinity for water because they consist of mostly hydrocarbons, which form nonpolar covalent bonds.
- Lipids are highly diverse in form and function.
- **Fats store large amounts of energy.**
- Although fats are not strictly polymers, they are large molecules assembled from smaller molecules via dehydration reactions.
- A **fat** is constructed from two kinds of smaller molecules: glycerol and fatty acids.

Glycerol is a three-carbon alcohol with a hydroxyl group attached to each carbon.

A **fatty acid** consists of a carboxyl group attached to a long carbon skeleton, often 16 to 18 carbons long.

The many nonpolar C—H bonds in the long hydrocarbon skeleton make fats hydrophobic.

Fats separate from water because the water molecules hydrogen-bond to one another and exclude the fats.

- In a fat, three fatty acids are joined to glycerol by an ester linkage, creating a **triacylglycerol**, or *triglyceride*.
 - The three fatty acids in a fat can be the same or different.
- Fatty acids vary in length (number of carbons) and in the number and locations of double bonds.

If the fatty acid has no carbon-carbon double bonds, then the molecule is a **saturated fatty acid**, saturated with hydrogens at every possible position.

If the fatty acid has one or more carbon-carbon double bonds formed by the removal of hydrogen atoms from the carbon skeleton, then the molecule is an **unsaturated fatty acid**.

- A saturated fatty acid is a straight chain, but an unsaturated fatty acid has a kink wherever there is a *cis* double bond.
- Fats made from saturated fatty acids are saturated fats. Fats made from unsaturated fatty acids are unsaturated fats.

Most animal fats are saturated.

Saturated fats are solid at room temperature.

Plant and fish fats are liquid at room temperature and are known as oils. The kinks caused by the *cis* double bonds prevent the molecules from packing tightly enough to solidify at room temperature.

The phrase “hydrogenated vegetable oils” on food labels means that unsaturated fats have been synthetically converted to saturated fats by the addition of hydrogen.

Peanut butter and margarine are hydrogenated to prevent lipids from separating out as oil.

A diet rich in saturated fats may contribute to cardiovascular disease (atherosclerosis) through plaque deposits.

The process of hydrogenating vegetable oils produces saturated fats and also unsaturated fats with *trans* double bonds. These *trans* fat molecules contribute more than saturated fats to atherosclerosis.

- The major function of fats is energy storage.

A gram of fat stores more than twice as much energy as a gram of a polysaccharide such as starch.

Because plants are immobile, they can function with bulky energy storage in the form of starch. Plants use oils when dispersal and compact storage are important, as in seeds.

Animals must carry their energy stores with them, so they benefit from having a more compact fuel reservoir of fat.

Humans and other mammals store fats as long-term energy reserves in adipose cells that swell and shrink as fat is deposited and withdrawn from storage.

- Adipose tissue also functions to cushion vital organs, such as the kidneys.
- A layer of fat can function as insulation.

This subcutaneous layer is especially thick in whales, seals, and most other marine mammals.

- ***Phospholipids are major components of cell membranes.***

- **Phospholipids** have two fatty acids attached to glycerol and a phosphate group at the third position.

The phosphate group carries a negative charge.

Additional smaller groups (usually charged or polar) may be attached to the phosphate group to form a variety of phospholipids.

- The interaction of phospholipids with water is complex.

The fatty acid tails are hydrophobic, but the phosphate group and its attachments form a hydrophilic head.

- When phospholipids are added to water, they form assemblages with the hydrophobic tails pointing toward the interior.

- Phospholipids are arranged as a bilayer at the surface of a cell.

The hydrophilic heads are on the outside of the bilayer, in contact with the aqueous solution, and the hydrophobic tails point toward the interior of the bilayer.

The phospholipid bilayer forms a barrier between the cell and the external environment.

Phospholipids are the major component of all cell membranes.

- ***Steroids include cholesterol and certain hormones.***

- **Steroids** are lipids with a carbon skeleton consisting of four fused rings.
- Different steroids are created by varying the functional groups attached to the rings.
- **Cholesterol**, an important steroid, is a component in animal cell membranes.
- Cholesterol is the precursor from which all other steroids are synthesized.

Many of these other steroids are hormones, including the vertebrate sex hormones.

- Although cholesterol is an essential molecule in animals, high levels of cholesterol in the blood may contribute to cardiovascular disease.
 - Both saturated fats and *trans* fats exert their negative impact on health by affecting cholesterol levels.

Concept 5.4 Proteins have many structures, resulting in a wide range of functions.

- Proteins account for more than 50% of the dry mass of most cells. They are instrumental in almost everything an organism does.

Protein functions include structural support, storage, transport, cellular signaling, movement, and defense against foreign substances.

Most important, protein enzymes function as **catalysts** in cells, regulating metabolism by selectively accelerating certain chemical reactions without being consumed.

- Humans have tens of thousands of different proteins, each with a specific structure and

function.

- Proteins are the most structurally complex molecules known.
 - Each type of protein has a complex three-dimensional shape.
- All protein polymers are constructed from the same 20 amino acid monomers.
- Polymers of proteins are called **polypeptides**.
- A **protein** consists of one or more polypeptides folded and coiled into a specific conformation.
- ***Amino acids are the monomers from which proteins are constructed.***
- **Amino acids** are organic molecules with both carboxyl and amino groups.
- At the center of an amino acid is an asymmetric carbon atom called the *alpha* (α) carbon.
- Four components are attached to the α carbon: a hydrogen atom, a carboxyl group, an amino group, and a variable R group (or side chain).
 - Different R groups characterize the 20 different amino acids.
 - An R group may be as simple as a hydrogen atom (as in the amino acid glycine), or it may be a carbon skeleton with various functional groups attached (as in glutamine).
- The physical and chemical properties of the R group determine the unique characteristics of a particular amino acid.
 - One group of amino acids has nonpolar R groups, which are hydrophobic.
 - Another group of amino acids has polar R groups, which are hydrophilic.
 - A third group of amino acids has functional groups that are charged (ionized) at cellular pH.
 - Some acidic R groups have negative charge due to the presence of a carboxyl group.
 - Basic R groups have amino groups with positive charge.
 - *All* amino acids have carboxyl and amino groups. The terms *acidic* and *basic* in this context refer only to these groups in the R groups.
- Amino acids are joined together when a dehydration reaction removes a hydroxyl group from the carboxyl end of one amino acid and a hydrogen atom from the amino group of another.
 - The resulting covalent bond is called a **peptide bond**.
- Repeating the process over and over creates a polypeptide chain.
 - At one end is an amino acid with a free amino group (the N-terminus), and at the other end is an amino acid with a free carboxyl group (the C-terminus).
- Polypeptides range in size from a few monomers to thousands.
- Each polypeptide has a unique linear sequence of amino acids.
- ***Scientists have determined the amino acid sequences of polypeptides.***
- Frederick Sanger and his colleagues at Cambridge University determined the amino acid sequence of insulin in the early 1950s.
 - Sanger used various protein-digesting enzymes and other catalysts to hydrolyze the insulin at specific places, yielding groups of small fragments.
 - Sanger used chemical methods to determine the sequences of amino acids in the small fragments.
 - After years of effort, Sanger was able to reconstruct the complete primary structure of insulin.
- Most of the steps in sequencing a polypeptide have since been automated.
- ***Protein conformation determines protein function.***
- A functional protein consists of one or more polypeptides that have been twisted, folded,

and coiled into a unique shape.

- It is the order of amino acids that determines the three-dimensional structure of the protein.
- A protein's specific structure determines its function.
- When a cell synthesizes a polypeptide, the chain generally folds spontaneously to assume the functional structure for that protein.
- The folding is reinforced by a variety of bonds between parts of the chain, which in turn depend on the sequence of amino acids.
- Many proteins are *globular*, while others are *fibrous* in shape.
- In almost every case, the function of a protein depends on its ability to recognize and bind to some other molecule.

For example, an antibody binds to a particular foreign substance.

An enzyme recognizes and binds to a specific substrate, facilitating a chemical reaction.

Natural signal molecules called endorphins bind to specific receptor proteins on the surface of brain cells in humans, producing euphoria and relieving pain.

Morphine, heroin, and other opiate drugs mimic endorphins because they are similar in shape and can bind to the brain's endorphin receptors.

- The function of a protein is an emergent property resulting from its specific molecular order.
- Three levels of structure—primary, secondary, and tertiary structures—organize the folding within a single polypeptide.

Quaternary structure arises when two or more polypeptides join to form a protein.

- The **primary structure** of a protein is its unique sequence of amino acids.

Transthyretin is a globular protein in the blood that transports vitamin A and a particular thyroid hormone throughout the body.

Each of the four identical polypeptide chains that, together, make up transthyretin is composed of 127 amino acids.

The precise primary structure of each of the polypeptide chains is determined by inherited genetic information, not by the random linking of amino acids.

- Even a slight change in the primary structure can affect a protein's conformation and ability to function.

The substitution of one amino acid (valine) for the normal one (glutamic acid) at a particular position in the primary structure of hemoglobin, the protein that carries oxygen in red blood cells, can cause *sickle-cell disease*, an inherited blood disorder.

The abnormal hemoglobins crystallize, deforming some of the red blood cells into a sickle shape and clogging capillaries.

- Most proteins have segments of their polypeptide chains repeatedly coiled or folded.
- These coils and folds are referred to as **secondary structure** and result from hydrogen bonds between the repeating constituents of the polypeptide backbone.

The weakly positive hydrogen atom attached to the nitrogen atom has an affinity for the oxygen atom of a nearby peptide bond.

Each hydrogen bond is weak, but the sum of many hydrogen bonds stabilizes the structure of part of the protein.

- One secondary structure is the **helix**, a delicate coil held together by hydrogen bonding between every fourth amino acid, as found in transthyretin.

Some fibrous proteins, such as -keratin, the structural protein of hair, have the helix formation over most of their length.

- The other main type of secondary structure is the **pleated sheet**.

In this structure, two or more regions of the polypeptide chain lying side by side are connected by hydrogen bonds between parts of the two parallel polypeptide backbones.

Pleated sheets are found in many globular proteins, such as transthyretin, and they dominate some fibrous proteins, including the silk protein of a spider's web.

The presence of so many hydrogen bonds makes each silk fiber stronger than a steel strand of the same weight.

- **Tertiary structure** is determined by interactions among various R groups.

These interactions include hydrogen bonds between polar and/or charged areas, ionic bonds between charged R groups, and **hydrophobic interactions** and van der Waals interactions among hydrophobic R groups.

Although these three interactions are relatively weak, their cumulative effect helps give the protein a unique shape.

Strong covalent bonds called **disulfide bridges** that form between the sulfhydryl groups (SH) of two cysteine monomers act to rivet parts of the protein together.

- **Quaternary structure** results from the aggregation of two or more polypeptide subunits.

For example, globular transthyretin protein is made up of four polypeptides.

Collagen is a fibrous protein made up of three polypeptides that are supercoiled into a larger triple helix.

The helix provides structural strength for collagen's role in connective tissue.

Hemoglobin is a globular protein with quaternary structure.

Hemoglobin consists of four polypeptide subunits: two α and two β chains.

Both types of subunits consist of primarily α -helical secondary structure.

Each subunit has a nonpeptide heme component with an iron atom that binds oxygen.

- A polypeptide chain with a given amino acid sequence can spontaneously arrange itself into a three-dimensional shape determined and maintained by the interactions responsible for secondary and tertiary structure.

The folding occurs as the protein is synthesized within the cell.

- Protein structure also depends on the physical and chemical conditions of the protein's environment.

Alterations in pH, salt concentration, temperature, or other factors can unravel or **denature** a protein.

These forces disrupt the hydrogen bonds, ionic bonds, and disulfide bridges that maintain the protein's shape.

Because it is misshapen, a denatured protein is biologically inactive.

- Most proteins become denatured if they are transferred from an aqueous environment to an organic solvent.

The polypeptide chain refolds so that its hydrophobic regions face outward, toward the solvent.

- Other denaturation agents include chemicals that disrupt the hydrogen bonds, ionic bonds, and disulfide bridges that maintain a protein's shape.

- Denaturation can also be caused by heat, which disrupts the weak interactions that stabilize conformation.

This explains why extremely high fevers can be fatal. Proteins in the blood become denatured by the high body temperatures.

- Some, but not all, proteins can return to their functional shape after denaturation.

This suggests that the information for building a specific shape is intrinsic to the protein's primary structure.

In the crowded environment inside a cell, specific proteins may assist with the folding of other proteins.

- Biochemists now know the amino acid sequences of more than 1.2 million proteins and the three-dimensional shapes of about 8,500 proteins.
 - Nevertheless, it is still difficult to predict the shape of a protein from its primary structure alone.
- Most proteins appear to undergo several intermediate stages before reaching their “mature” structure.
- The folding of many proteins is assisted by **chaperonins**, or chaperone proteins.
 - Chaperonins do not specify the final structure of a polypeptide but rather work to segregate and protect the polypeptide while it folds spontaneously.
 - Molecular systems in the cell interact with chaperonins, marking incorrectly folded proteins for refolding or for destruction.
- Accumulation of incorrectly folded polypeptides is associated with many diseases, including Alzheimer's disease and Parkinson's disease.
- In 2006, Roger Kornberg received the Nobel Prize in Chemistry for elucidating the structure of RNA polymerase, which plays a crucial role in the expression of genes.
 - Kornberg and his colleagues used **X-ray crystallography** to determine the three-dimensional structure of this protein.
- Nuclear magnetic resonance (NMR) spectroscopy, which does not require protein crystallization, can also be used to determine protein structure.
- Bioinformatics uses computer programs to predict the three-dimensional structures of amino acid sequences.
 - In 2005, Austrian scientists Heimo Breiteneder, Clare Mills, and their colleagues used bioinformatics to analyze the amino acid sequences and structures of 129 common plant protein allergens.
 - They classified 65% of these allergens into just 4 of 3,849 possible protein families.
 - This high degree of similarity suggests that shared structures in these protein families may play a role in generating allergic reactions.
 - This information may be useful in the development of medications to relieve allergy symptoms.
- X-ray crystallography, NMR spectroscopy, and bioinformatics are complementary approaches, contributing a great deal to our understanding of protein structure.

Concept 5.5 Nucleic acids store and transmit hereditary information.

- The amino acid sequence of a polypeptide is programmed by a unit of inheritance known as a **gene**.
- A gene consists of DNA, a polymer known as a **nucleic acid**.
- ***There are two types of nucleic acids: RNA and DNA.***
- The two types of nucleic acids are **ribonucleic acid (RNA)** and **deoxyribonucleic acid (DNA)**.
- RNA and DNA are the molecules that enable living organisms to reproduce their complex components from generation to generation.
 - DNA provides directions for its own replication.
 - DNA also directs RNA synthesis and, through RNA, controls protein synthesis.
- Organisms inherit DNA from their parents.

Each DNA molecule is very long, consisting of hundreds to more than a thousand genes.

Before a cell reproduces itself by dividing, its DNA is copied. The copies are then passed to the next generation of cells.

- Although DNA encodes the information that programs all the cell's activities, it is not directly involved in the day-to-day operations of the cell.
- Proteins are responsible for implementing the instructions contained in DNA.
- Each gene along a DNA molecule directs the synthesis of a specific type of RNA called *messenger RNA* (mRNA).
- The mRNA molecule interacts with the cell's protein-synthesizing machinery to direct the ordering of amino acids in a polypeptide.
- The flow of genetic information is DNA → RNA → protein.
- Protein synthesis occurs on cellular structures called ribosomes.
- In eukaryotes, DNA is located in the nucleus, but most ribosomes are in the cytoplasm. mRNA functions as an intermediary, moving information and directions from the nucleus to the cytoplasm.
- Prokaryotes lack nuclei but still use RNA as an intermediary to carry a message from DNA to the ribosomes.
- ***A nucleic acid strand is a polymer of nucleotides.***
- Nucleic acids are polymers made of **nucleotide** monomers organized as **polynucleotides**.
- Each nucleotide consists of three parts: a nitrogenous base, a pentose sugar, and a phosphate group.
- The nitrogenous bases are rings of carbon and nitrogen that come in two types: purines and pyrimidines.

Pyrimidines have a single six-membered ring of carbon and nitrogen atoms.

There are three different pyrimidines: cytosine (C), thymine (T), and uracil (U).

Thymine is found only in DNA and uracil is found only in RNA.

Purines have a six-membered ring joined to a five-membered ring.

The two purines are adenine (A) and guanine (G).

- The pentose joined to the nitrogenous base is **ribose** in nucleotides of RNA and **deoxyribose** in DNA.

The only difference between the sugars is the lack of an oxygen atom on carbon 2 in *deoxyribose*.

Because the atoms in both the nitrogenous base and the sugar are numbered, the sugar atoms are distinguished by a prime (') after the number.

Thus, the second carbon in the sugar ring is the 2' (2 prime) carbon, and the carbon that sticks up from the ring is the 5' carbon.

The combination of a pentose and a nitrogenous base is a *nucleoside*.

- The addition of a phosphate group creates a nucleoside monophosphate or *nucleotide*.
- **Polynucleotides** are synthesized when adjacent nucleotides are joined by covalent bonds called phosphodiester linkages that form between the —OH group on the 3' of one nucleotide and the phosphate on the 5' carbon of the next.

This process creates a repeating backbone of sugar-phosphate units, with appendages consisting of the nitrogenous bases.

- The two free ends of the polymer are distinct.

One end has a phosphate attached to a 5' carbon; this is the 5' end.

The other end has a hydroxyl group on a 3' carbon; this is the 3' end.

- The sequence of bases along a DNA or mRNA polymer is unique for each gene.
 - Because genes are normally hundreds to thousands of nucleotides long, the number of possible base combinations is virtually limitless.
- The linear order of bases in a gene specifies the order of amino acids—the primary structure—of a protein, which in turn determines three-dimensional structure and function.
- ***Inheritance is based on replication of the DNA double helix.***
- An RNA molecule is a single polynucleotide chain.
- DNA molecules have two polynucleotide strands that spiral around an imaginary axis to form a **double helix**.
 - The double helix was first proposed as the structure of DNA in 1953 by James Watson and Francis Crick.
- The sugar-phosphate backbones of the two polynucleotides are on the outside of the helix.
 - The two backbones run in opposite 5' 3' directions from each other, an arrangement referred to as **antiparallel**.
- The two polynucleotides or strands are held together by hydrogen bonds between the paired bases and by van der Waals interactions between the stacked bases.
- Most DNA molecules have thousands to millions of base pairs.
- Because of their shapes, only some bases are compatible with each other.
 - Adenine (A) always pairs with thymine (T) and guanine (G) with cytosine (C).
 - With these base-pairing rules, if we know the sequence of bases on one strand, we know the sequence on the opposite strand.
 - The two strands are *complementary*.
- Prior to cell division, each of the strands serves as a template to order nucleotides in a new complementary strand.
 - This results in two identical copies of the original double-stranded DNA molecule, which are then distributed to the daughter cells.
 - This mechanism ensures that a full set of genetic information is transmitted whenever a cell reproduces.
- ***We can use DNA and proteins as tape measures of evolution.***
- Genes (DNA) and their products (proteins) document the hereditary background of an organism.
- Because DNA molecules are passed from parents to offspring, siblings have greater similarity in their DNA and protein than do unrelated individuals of the same species.
- This argument can be extended to develop a “molecular genealogy” to relationships *between* species.
- Two species that appear to be closely related based on fossil and molecular evidence should also be more similar in DNA and protein sequences than are more distantly related species.
 - Scientists can compare the sequence of 146 amino acids in the polypeptide chain of human hemoglobin to the sequences in five other vertebrates.
 - Humans and gorillas differ in just 1 amino acid, while humans and frogs differ in 67 amino acids.
 - Despite these differences, all the species have functional hemoglobin.