INTRODUCTION TO ORGANIC AND BIOCHEMISTRY

Muhammad Arif Malik
Hampton University

LibreTexts™

Introduction to Organic and Biochemistry

Muhammad Arif Malik

Hampton University, Hampton, VA

The LibreTexts libraries are Powered by NICE CXOne and are supported by the Department of Education Open Textbook Pilot Project, the UC Davis Office of the Provost, the UC Davis Library, the California State University Affordable Learning Solutions Program, and Merlot. This material is based upon work supported by the National Science Foundation under Grant No. 1246120, 1525057, and 1413739.

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation nor the US Department of Education.

TABLE OF CONTENTS

6: Lipids

7: Proteins

8: Nucleic acids

9: Food to energy metabolic pathways

Licensing

*A detailed breakdown of this resource's licensing can be found in **Back Matter/Detailed Licensing**.*

Dedication

This book is dedicated to Karl H. Schoenbach - Post Doctoral Supervisor of the author

Karl H. Schoenbach, Eminent Scholar Emeritus, Professor Emeritus of Electrical and Computer, and Founding Director of Frank Reidy Research Center for Bioelectrics at Old Dominion University, Norfolk, Virginia, helped bring the author to the US. The most prominent of his achievements is the development of nanosecond pulsed high voltage techniques for medical treatments proven to cure some forms of cancer.

CHAPTER OVERVIEW

1: Bonding in organic compounds

1.1: What is organic chemistry?

Learning Objectives

- Recognize organic compounds from their formula.
- Understand some logical reason for why nature has selected C and H as the main constituent of organic compounds.

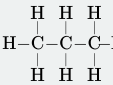
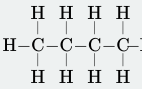
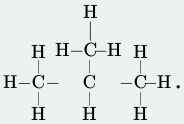
What are organic compounds?

The compounds usually synthesized in living things are called organic compounds. The organic compounds are primarily composed of carbon (C) and hydrogen (H). For example, methane (CH_4) produced by decaying plant materials is composed of one C and four H′s. Often one or more atoms of elements other than C and H are also present in organic compounds, like oxygen (O), nitrogen (N), phosphorous (P), sulfur (S), etc. For example, O atoms are present in glucose ($C_6H_{12}O_6$). The atoms other than C and H, e.g., O in $C_6H_{12}O_6$, are called **heteroatoms**. Figure 1.1.1 illustrates fruits and vegetables composed of organic compounds.

Figure 1.1.1: Fruits and vegetables comprise organic compounds. (Copyright; oy, CC BY 2.0, via Wikimedia Commons)

Why has nature chosen C and H as primary constituents of organic compounds?

Nature has chosen C and H as primary constituents composing the organic compounds because of several reasons, some of which are the following.

- C is a member of second-row elements in the periodic table that usually make stronger and more stable bonds than the elements in the higher rows.
- C makes four bonds in neutral molecules, which is higher than any other element of the second row can make. For example, C has for bonds in a methane molecule represent as:  , where each line represents a bond.
- C can make chains and rings, e.g., ethane:  , and a propane:  are chains of two and three C′s. The ability of C to make a chain of atoms is called **catenation**.
- A C atom can make bonds with more than two C′s resulting in branched compounds that increases the number of compounds possible, e.g. four C′s molecule can be in a straight chain as in n-butane:  or in a branched chain as in isobutane:  .
- Two C′s can make single, double, or triple bonds with each other allowing more variations. For example, ethane () has all single bonds; ethene () as double bond; and acetylene ( ) has a triple bond between C′s.
- H is the lightest monovalent atom that can occupy the valencies of C′ not used in C to C bonds.

- H bonded with a strongly electronegative atom like O or N can interact with a O or N atom of a neighboring molecule through hydrogen bonding that plays a vital role in the functioning of organic molecules in living things.
- C or H in an organic compound can be replaced with a heteroatom that tremendously increases the variety of organic compounds available to living things.

Some other elements have the desired characteristics of C and H but are associated with significant disadvantages. For example, silicon (Si), like C, makes i) four bonds, ii) straight and branched chains, and iii) a single bond with hydrogen. The drawbacks of Si are i) it is two times heavier than C, ii) it makes weaker unstable Si−Si, and Si−H bonds compared to C−C and C−H bonds, and iii) its oxidation product is solid silicon dioxide (SiO_2) which is insoluble in water and would have been difficult to excrete than the gaseous CO_2 from the C compounds which are easier to exhale. Similarly, halogens are monovalent like hydrogen, but halogens are significantly heavier than H and make weaker bonds with C than C−H bonds.

What is organic chemistry?

Organic chemistry is the study of the properties and reactions of organic compounds.

The following section describes chemical bonding that determines the compound's physical properties and chemical reactivities.

1.2: What is a chemical bond

Learning Objectives

- Describe the nature of ionic, covalent, and coordinate covalent bonds.
- Understand the nature of a bond's polarity and molecules' polarity.
- Describe the nature of intermolecular forces and their effect on some properties of the compounds.
- Compare the properties of organic and inorganic compounds.

Chemical bonds were introduced in general chemistry. This is a review of chemical bonds for understanding the differences between inorganic and organic compounds. The chemical bonds connect the atoms in the compounds. Transferring or sharing some valance electrons from one atom to another makes the bonds. There are three major types of chemical bonds, ionic, covalent, and coordinate covalent bonds, as described below.

Ionic bond

An ionic bond is formed by transferring some valence electrons from one atom (usually from a metal atom) to another (usually to a nonmetal atom). The atom that loses an electron becomes a positively charged specie called a cation. For example, sodium atom Na loses one electron and become Na^{+}, or calcium Ca loses two electrons and becomes Ca^{2+}. The atom that gains an electron becomes a negatively charged specie called an anion. For example, chlorine atom Cl gains one electron and become Cl^{-}, or oxygen O gains two electrons and becomes O^{2-}.

The same charges repel each other, and opposite charges attract each other. The cations and the anions come together in the structure of the compound in such a way that the distances between the centers of opposite charges are minimized to increase attraction. The spaces between the centers of the same charges are maximized to decrease repulsion, as illustrated in Figure $1.2.1$ for the case of NaCl. An ionic bond is the net attractive force holding the ions together in an ionic compound. ionic compounds are hard and have high melting and boiling points due to the strong electrostatic attractive forces. Ionic compounds are brittle because a slight displacement of one layer of atoms relative to the other layer can place similar charges next to each other and split due to repulsive forces between similar charges.

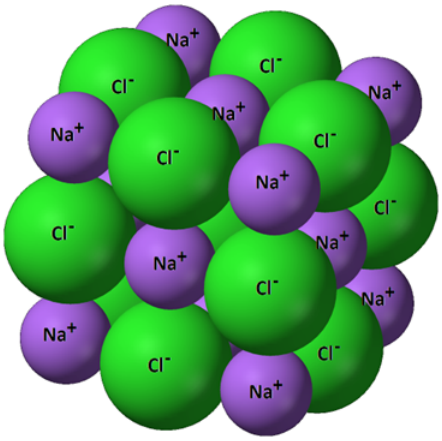


Figure 1.2.1: Crystal lattice of an ionic compound NaCl illustrated. (Copyright; Public domain)

Covalent bond

What is a covalent bond?

A bond formed by sharing valence electrons is a covalent bond. Each bonded atom contributes one electron in a shared pair of electrons called a covalent bond, e.g., $\mathrm{H-Cl}$ bond illustrated in Figure $1.2.2$. The shared pair of electrons is called a **bonding pair,** and it is usually represented as a single line between the bonded atoms as in $\mathrm{H-Cl}$. Note that there are three unshared pairs of electrons on Cl atom in $\mathrm{H-Cl}$ molecule as illustrated in Figure $1.2.2$. The unshared electron pairs are called nonbonding or **lone pairs** of electrons. Atoms may share two electrons each to make a double bond, e.g., in $\mathrm{O=O}$, or three electrons each to create a triple bond, e.g., in $\mathrm{N \equiv N}$, as illustrated in Figure $1.2.2$.

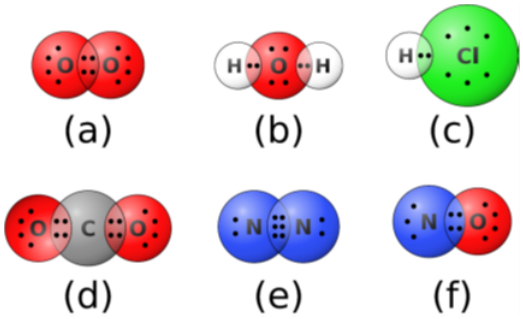


Figure 1.2.2: Covalent bonds formed by sharing some of the valence electrons in the overlapped portions of the electron clouds of the bonded atoms are illustrated for the cases of: a) O=O, b) H−O−H, c) H−Cl, d) O=C=O e) N ≡ N, and f) N=O. Each line represents a bonding electron pair. (Copyright; MikeRun, CC BY-SA 4.0, via Wikimedia Commons)

How is a covalent bond formed?

Isolated atoms have **potential energy** due to several factors, like movements of atoms, movements of and electrostatic interactions among subatomic particles, etc. When two atoms approach each other, attractive and repulsive forces develop. The attractive forces are the attraction between opposite charges, i.e., between the nucleus of one atom and the electrons of the other. The repulsive forces are the repulsion between the same charges. When the two atoms can place more electrons between the nuclei by overlapping their atomic orbitals, the attractive forces between nuclei and electrons become stronger than the repulsive forces. Figure 1.2.3 illustrates this situation for the case of H−H covalent bond formation. The two atoms move towards each other due to the attractive force resulting in a decrease in their potential energy until the minimum potential energy is reached. Decreasing the internuclear distance to less than the bond distance results in repulsive forces increasing more than attractive forces, and the overall potential energy increases.

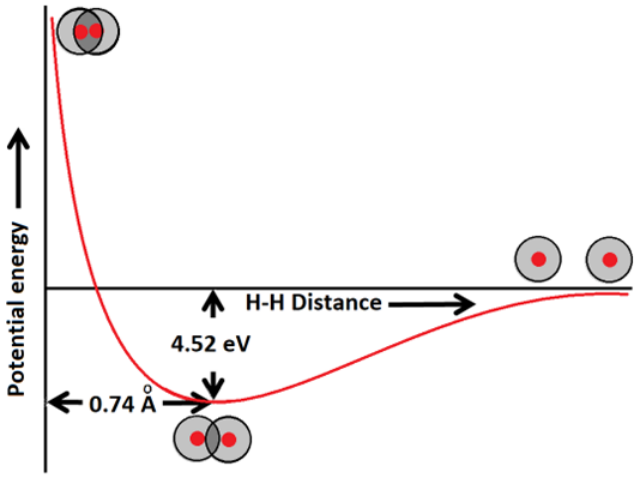


Figure 1.2.3: Illustration of potential energy vs. internuclear distance for the case of covalent bond formation between two H atoms. (Copyright; Public domain)

Bond length, bond energy, and the bond strength

- **Bond lenght** is the distance between the centers of the two atoms at the potential energy minimum. For example, the bond distance H−H is 0.74 Å.
- The difference in the energy of isolated atoms and the energy state at the bonding distance is the **bond energy**. For example, the bond energy of H−H is 4.52 eV as shown in Figure 1.2.3
- Bond energy is always released when the bond is formed, and the same amount of energy is absorbed when the same bond is broken. **Stronger bonds** have higher bond energy, and **weaker bonds** have lower bond energy.

The polarity of a covalent bond

The polarity was introduced in general chemistry. Here is a quick review. When the bonded atoms are the same, the bonding electron pair equally shared between them and the bond is **nonpolar**. For example, H−H, F−F, and C−C bonds are nonpolar. When the bonded atoms differ, one atom attracts the bonding election more than the other. The ability of an atom to attract the

bonding electrons to itself is called **electronegativity**. If the difference in the electronegativity values of the bonded atoms is less than 0.5, the bond is still considered **nonpolar**. However, if the difference in the electronegativity values of the bonded atoms is between 0.5 to 1.9, the bond is a **polar covalent**. When the electronegativity difference is more than 1.9, it is usually an **ionic bond**.

In a polar covalent bond, the more electronegative atom has a partial negative (δ-) charge because the electronegative atom has more share of negatively charged electrons, and the other atom has partial positive (δ+) charge, e.g., $\overset{\delta+}{\mathrm{H}}-\overset{\delta-}{\mathrm{Cl}}$, $\overset{\delta+}{\mathrm{C}}-\overset{\delta-}{\mathrm{O}}$, and $\overset{\delta-}{\mathrm{O}}-\overset{\delta+}{\mathrm{H}}$.

The polarity of a molecule

The polarity of molecules is described in general chemistry. Recall that polarity is a vector quantity. It is also shown as an arrow over the bond with the arrowhead pointing to δ- end and the tail with a plus sign starting from δ+ end of the polar bond, e.g., $\overset{\delta+}{\mathrm{H}}-\overset{\delta-}{\mathrm{Cl}}$ can be represented in vector form as: $\overset{+\longrightarrow}{\mathrm{H}-\mathrm{Cl}}$. Suppose there is more than one polar bond in a molecule. In that case, the polarity vectors may cancel out each other, e.g., two equal and opposite polarity vectors in a carbon dioxide molecule ($\overset{\longleftarrow+\longrightarrow}{\mathrm{O}=\mathrm{C}=\mathrm{O}}$) cancel each other resulting in a nonpolar molecule. Therefore, the molecule with more than one polar bond may or may not be polar, depending on whether the polarities cancel each other. The following rules determine the polarity of a molecule:

1. A molecule is polar if there is only one polar bond in it.
2. A molecule is polar if there are polar bonds, but the molecule is not symmetric. In this case, the polarities do not entirely cancel out.
3. A molecule is nonpolar if there are polar bonds, but the molecule is symmetric, e.g., i) linear with two equal bonds, ii) trigonal planar with three equal bonds, iii) tetrahedral with four equal bonds. In these cases, the polarities cancel out.
4. A molecule is nonpolar if there is no polar bond in it.

Coordinate covalent bond

The coordinate covalent bond is the same as the covalent bond, except that both bonding electrons are donated by one of the two bonded atoms. The coordinate covalent bond is also called a dative bond. For example, the bond formed between the boron atom of boron trifluoride ($\mathrm{BF_3}$) and the nitrogen atom of ammonia ($\mathrm{:NH_3}$) in the following reaction: $\mathrm{F_3B + :NH_3 \longrightarrow F_3\overset{-}{B}-\overset{+}{N}H_3}$, where N donates its lone pair to make the bond. The charges developed on the N and B after bonding is explained later in the formal charge section.

Intermolecular interactions

Although a covalent bond is almost as strong as an ionic bond, the covalent bond forces operate within a molecule. The question is, what holds the molecules together in solid molecular compounds? These are electrostatic interactions called intermolecular forces that were introduced in general chemistry. Intermolecular forces are much weaker than ionic, covalent, or coordinate covalent bonds. A line shows a covalent bond, while dotted lines usually show intermolecular forces. Three major intermolecular forces exist: i) London dispersion forces, ii) dipole-dipole interaction, and iii) hydrogen bonding.

London dispersion

Atoms have +ve protons in the nucleus and equal -ve electrons outside the nucleus. The atom is nonpolar because the electrons and protons are equal, and the -ve electrons are symmetrically distributed around the +ve nucleus. Electrons can be considered like clouds that can temporarily sway to one side or the other. If the electron cloud sways to one side, the atom is no more nonpolar, the side of the nucleus becomes a partial positive (δ+) pole, and the side where electrons move to becomes a partial negative (δ-) pole. This transient dipole in one atom induces a temporary dipole in the neighboring atom due to repulsion between the same charges or attraction between the opposite charges. The dipoles orient themselves to maximize attractive force between opposite charges and minimize repulsion between the same changes resulting in a net attractive force. This attraction between transient dipole-induced dipole is called the **London dispersion** force. The same phenomenon happens in the molecules.

The transient dipole appears and disappears randomly. **London dispersion force is proportional to molecular mass** because more mass means more electrons and a higher probability of temporary dipole. That is why smaller nonpolar organic compounds, like methane ($\mathrm{CH_4}$, boiling point ~-160 ^{o}C) are gases, large like decane ($\mathrm{C_{10}H_{22}}$, boiling point ~+174 ^{o}C) are liquid, and even larger like eicosane ($\mathrm{C_{20}H_{42}}$, melting point ~37 ^{o}C, boiling point ~+343 ^{o}C) are solid.

Dipole-dipole interactions

Polar molecules have a permanent dipole in addition to the London dispersions force (transient dipoles). Therefore, polar molecules have higher intermolecular interaction and, consequently, higher melting and boiling points than nonpolar molecules of comparable molecular masses. For example, acetone (C_2H_6O), a polar molecule, is liquid (melting point ~-95 ^oC and boiling point ~+56 ^oC), and butane (C_4H_{10}), a nonpolar molecule of the same molecular mass (58 g/mole), is a gas (melting point ~-137 ^oC and boiling point ~0 ^oC).

Hydrogen bonding

Hydrogen bonding is a particular class of dipole-dipole interactions that involve $\overset{\delta-}{O}-\overset{\delta+}{H}$, $\overset{\delta-}{N}-\overset{\delta+}{H}$, or $\overset{\delta-}{F}-\overset{\delta+}{H}$ dipole. Hydrogen bonding is usually stronger than dipole-dipole interactions because i) the dipoles in hydrogen bonding are usually stronger than the other dipoles, ii) $\delta+$ charge is denser on a small H than the same charge on larger atoms, and iii) being small, H can penetrate even in tight spaces to establish hydrogen bonding. For example, methanol (CH_3OH), capable of hydrogen bonding, is a liquid (boiling point ~+65 ^oC), formaldehyde ($H_2C=O$), a polar molecule without hydrogen bonding capability, is a gas (boiling point ~-19 ^oC), and ethane (C_2H_2), a nonpolar molecule, is a gas with much lower boiling point (~-89 ^oC), all three having comparable molar masses 32 g/mole, 30 g/mole, and 30 g/mole, respectively. Further, hydrogen bonding is more important in living things because H is one of the primary constituents of organic compounds.

The intermolecular forces play an essential role in the chemistry of living things, e.g., as illustrated in Figure $1.2.4$ for defining the shape of a protein molecule.

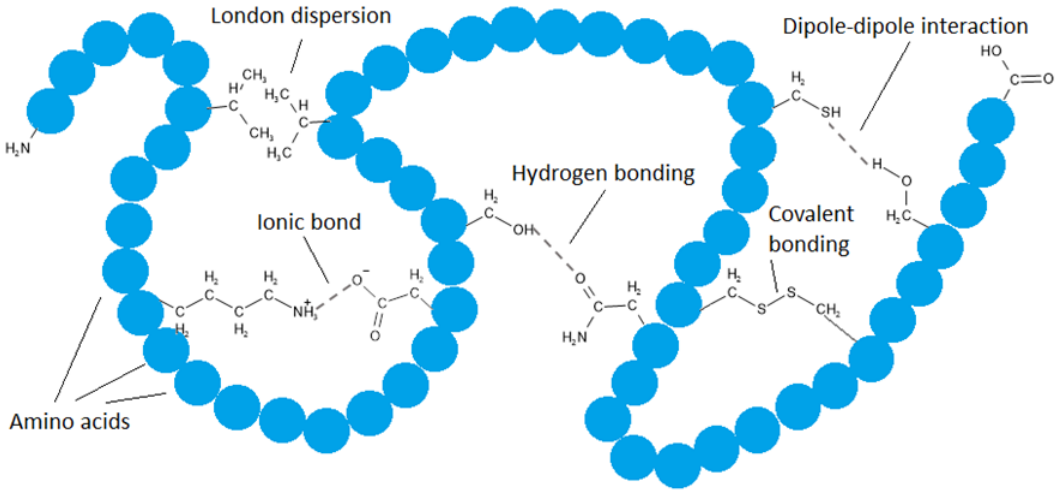


Figure 1.2.4: Illustration of bonding (ionic and covalent) and intermolecular interactions (London dispersion forces, dipole-dipole interactions, and hydrogen bonding) in a protein molecule. (Copyright; Public domain)

Comparison of bonding and properties of organic and inorganic compounds

Organic compounds

1. Organic compounds are usually covalently bonded molecules, e.g., methane (CH_4), glucose ($C_6H_{12}O_6$), etc.
2. Organic compound are primarily composed of C and H. Other elements, like O, N, P, S, etc., may also be present in small amounts.
3. The polarity of organic compounds varies from nonpolar to highly polar, e.g., decane ($C_{10}H_{22}$) is nonpolar, and glucose ($C_6H_{12}O_6$) is polar.
4. Organic compounds are usually insoluble in water, e.g., decane ($C_{10}H_{22}$) is insoluble in water, except a few highly-polar compounds that are soluble, e.g., glucose ($C_6H_{12}O_6$) is soluble in water.
5. Organic compounds usually have lower melting and boiling points, e.g., decane ($C_{10}H_{22}$) melts at ~-30 ^oC and boils at ~174 ^oC.
6. Organic compounds are usually soft solids, liquids, or sometimes gases, e.g., glucose ($C_6H_{12}O_6$) is a soft solid, decane ($C_{10}H_{22}$) is liquid, and methane (CH_4) is a gas.

Inorganic compounds

1. Inorganic compounds are usually ionic or highly polar covalent, e.g., table salt (NaCl) is ionic and ammonium phosphate ($(\mathrm{NH_4})_3\mathrm{PO_4}$ -a fertilizer, is a combination of covalent and ionic.
2. Inorganic compounds are usually composed of a combination of metals and nonmetals. e.g., calcium carbonate ($\mathrm{CaCO_3}$ is composed of metal (Ca) and nonmetals (C) and (O).
3. Inorganic compounds are usually soluble in water, e.g., NaCl and $(\mathrm{NH_4})_3\mathrm{PO_4}$ are soluble in water. Some exceptions exist, e.g., $\mathrm{CaCO_3}$ is not soluble in water.
4. Inorganic compounds usually have high melting and boiling points, e.g., NaCl melts at ~-801 ^{o}C and boils at ~1465 ^{o}C.
5. Inorganic compounds are usually hard and brittle solids, with few liquids and gases, e.g., NaCl and $\mathrm{CaCO_3}$ are hard and brittle solids.

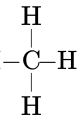
1.3: Hybridization of orbitals and 3D structures of simple organic compounds

Learning Objectives

- Understand sp^3, sp^2, and sp hybridization of C, N, O, and halogens in organic compounds.
- Understand bond angels and geometries in simple molecules having sp^3, sp^2, and sp hybridization of C, N, O.
- Read an electrostatic potential map of a simple organic molecule and recognize the partially charged regions and the lone pairs.
- Understand free rotation around single and triple bonds but not around double bonds.

Prelude to the hybridization of atomic orbitals

Electron configuration of C atom is: $1s^2$, $2s^2$, $2p_x^1$, $2p_y^1$, $2p_z^0$, where 1s is the core-shell and 2s and 2p are in the valence shell. Since two unpaired electrons are in the valence shell in $2p_x^1$, $2p_y^1$ orbitals, it implies that C should have two bonds in its compound. In reality, C has four bonds, e.g., there are four C−H covalent bonds in methane (CH_4). It can be explained by assuming that one electron is promoted from 2s to 2p orbital of C, resulting in the electron configuration of $1s^2$, $2s^1$, $2p_x^1$, $2p_y^1$, $2p_z^0$, which allows four bonds. The rationale is that making tow addition bonds releases more energy that more than compensates the energy consumed in promoting an electron from 2s to 2p orbital.

The molecular formula of methane is (CH_4), which shows the elemental composition but does not show bonds. The Lewis structure of methane is  which shows covalent bonds as lines but does not show the actual geometry of the molecule. The geometry of the methane molecule is tetrahedral. C is at the center of the tetrahedron, and four H atoms at the corners of the tetrahedron as illustrated in Figure 1.3.1 a and b. The four bonds are equal and at 109.5° from each other.

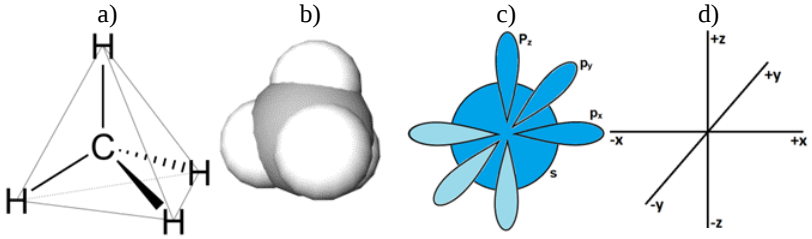


Figure 1.3.1: a) Illustration of tetrahedral geometry of CH_4, b) model of methane molecule where white spheres on the periphery are H and gray sphere in the center is C, c) illustration of valence orbitals of C atom, i.e., (2s, $2p_x$, $2p_y$, and $2p_z$), and d) 3D-coordinate axis for reference (Copyright; a) Hbf878, CC0, via Wikimedia Commons, b), c, and d) Public domain)

The geometry of the methane molecule, shown in Figure 1.3.1 a and b, disagrees with the geometry of valence orbitals of C illustrated in Figure 1.3.1 c. The s orbital is a spherical shape with a single phase shown by a single shade, while p orbitals have two lobes of opposite phases along the axis indicated by two different shades. The three p orbitals are perpendicular to each other, each lying along one of the three axes: p_x along the x-axis, p_y along the y-axis, and p_z along the z-axis. Suppose the valence s and the three p orbitals of C were involved in the bonding. In that case, there should have been three bonds from the p orbitals perpendicular to each other and one bond from the s orbital different from the other three at any angle around the C nucleus. This difference between the actual and the expected geometry of CH_4 molecule is explained by the orbital hybridization concept.

According to the **orbital hybridization** concept, atomic orbitals can combine and produce an equal number of hybrid orbitals of energy and orientation different from the constituent orbitals. The s and all three, any two, or any one of the p valence orbitals of C can mix and produce four sp^3, three sp^2, or two sp hybrid orbitals.

The geometry around C atoms in organic molecules agrees with the geometry of the hybrid orbitals, as described in the following sections.

sp^3 Hybridization of C

sp^3 Hybridization process

The s-orbital has a spherical shape and has a single phase as illustrated in 1.3.2 a. Each p orbital has two lobes lying on a straight line with the nucleus at the center. The two lobes have opposite phases, represented by shades as illustrated in 1.3.2 a. The term phase is about the wave nature of electrons. Waves have crests and troughs that are opposite phases.

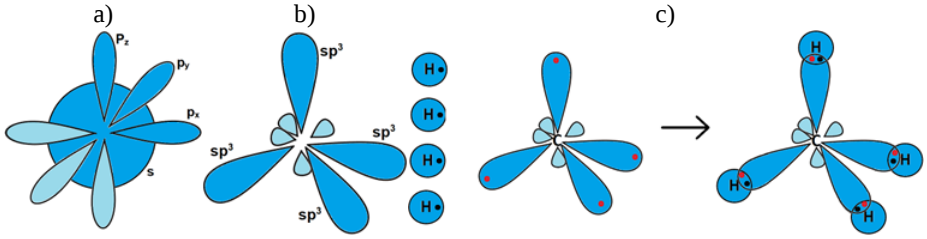


Figure 1.3.2: a) An s and three p valence orbitals before hybridization, b) a set of four equal sp^3 hybrid orbitals formed from the s and the three p orbitals, c) illustration of four covalent bonds formed by the overlap of one sp^3 orbital of C and one s orbital of H each, called $\sigma_{sp^3\text{-}s}$ bond, resulting in CH_4 molecule. Each orbital contributes one electron to the bond, represented by a red dot in the case of sp^3 and a black dot in the case of s orbital for illustration (Copyright; Public domain).

When one s and three p orbitals mix, a set of four equal hybrid orbitals, called sp^3 orbitals, is formed as illustrated in 1.3.2 b. An sp^3 orbital has two lobes of opposite phases, a more prominent lobe and a smaller lobe, lying along a straight line with the nucleus between the two lobes. Each sp^3 orbital has one-fourth of s and three-fourths of p orbital character. The sp^3 orbitals are arranged in a tetrahedral geometry around the nucleus, as illustrated in Figure 1.3.2 b. In the case of C, the four valence electrons distribute one each in the four sp^3 hybrid orbitals.

Formation of methane (CH_4) molecule

Each of the four sp^3 orbitals of C overlaps with one 1s orbital of H, making a set of four covalent bonds as illustrated in Figure 1.3.2 c. The bond formed by a head-on overlap of atomic orbitals is called a sigma-bond (σ-bond), which in the present case is called $\sigma_{sp^3\text{-}s}$ bond. The result is a CH_4 molecule with all bonds equal in a tetrahedral geometry around C, which agrees with the observed geometry, as illustrated in Figure 1.3.1: a. All bonds are equal of 108.70 pm bond length and 109.5^o bond angles in a CH_4 molecule. Figure 1.3.3 presents an interactive model of CH_4 molecule for lea

Formation of ethane (CH_3CH_3) molecule

Carbon can form a chain of carbons by making covalent bonds between carbon atoms. For example, a C overlaps a bigger lobe of one of its sp^3 hybrid orbitals with the bigger lobe of the sp^3 orbital of the second carbon along the axis of the sp^3 orbitals to make a $\sigma_{sp^3\text{-}sp^3}$ bond as illustrated in Figure 1.3.4 a. The remaining sp^3 orbitals on each C make $\sigma_{sp^3\text{-}s}$ bond by overlapping with s orbital of H resulting in an ethane C_2H_6 molecule.

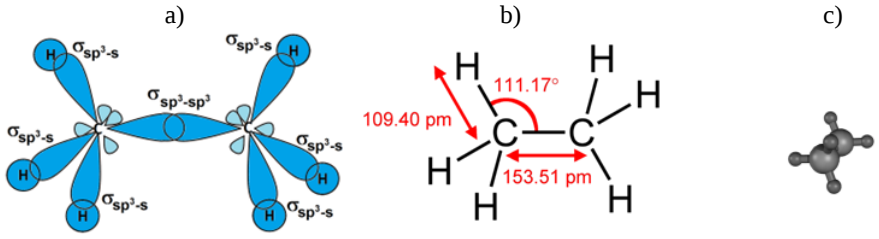


Figure 1.3.4: a) illustration of covalent bonds in a CH_3CH_3 molecule, b) structure showing bond lengths and angles in a CH_3CH_3 molecule, and c) Model of CH_3CH_3 molecule showing views from different angles and also showing rotation around $\sigma_{sp^3\text{-}sp^3}$ bond in a CH_3CH_3 molecule. Note that the bond angles are not precisely equal to but very close to the predicted 109.5^o, (Copyright; a) Public domain, b) Benjah-bmm27, Public domain, via Wikimedia Commons, and c) mailto:ralf@ark.in-berlin.de, CC BY 2.5 via Wikimedia Commons)

Rotation around a single bond is possible under ambient conditions. Rotation around a C−H bond does not change the shape or orientation of any atom in the molecule. However, rotation around C−C in CH_3CH_3 molecule results in different orientations of H atoms on one C relative to the H atoms on the other C, as illustrated in Figure 1.3.4 c. The structures with different orientations of atoms in the same molecule, as a result of rotation around a single bond, are called **conformers** of each other.

sp^2 Hybridization of C

sp^2 Hybridization process

One s and two p orbitals in a valence shell can mix and produce a set of three equal hybrid orbitals, called sp^2 orbitals, as illustrated in Figure 1.3.5. An sp^2 orbital has one-third s-orbital and two-third p-orbital character. An sp^2 orbital comprises two lobes, a more prominent lobe and a smaller lobe of opposite polarity along a straight line with the nucleus between the two lobes, as in the case of the sp^3 orbital. The three sp^2 are in a plane at 120° from each other, i.e., a trigonal planer geometry with the nucleus in the middle. One p orbital left out from the hybridization lies perpendicular to the plane of the three sp^2 orbitals.

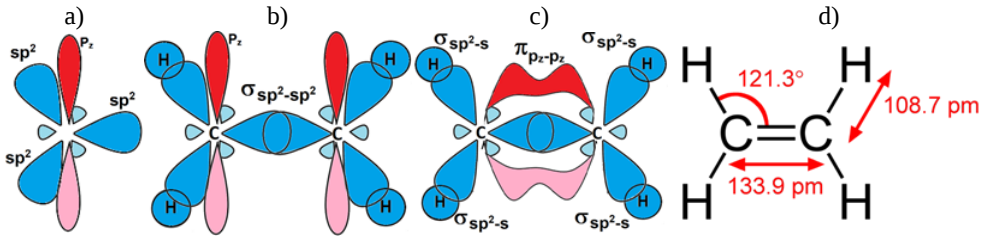


Figure 1.3.5: Cartoons of a) a set of three equal sp^2 hybrid orbitals in a trigonal planer geometry, formed from the mixing of one s and the two p orbitals along with the un-hybridized p perpendicular to the plane, b) a $\sigma_{sp^2\text{-}sp^2}$ bond formation between two sp^2 hybridized C atoms along with four $\sigma_{sp^2\text{-}s}$ bond formation between sp^2 hybridized C and H atoms, c) a π-bond formation between two parallel p orbitals on adjacent C atoms, and d) structure showing bond lengths and angles in a CH_2CH_2 molecule, (Copyright; a), b, and c) Public domain, d) Benjah-bmm27, Public domain, via Wikimedia Commons)

Formation of ethene ($H_2C{=}CH_2$) molecule

A C can form multiple bonds with a C or a heteroatom. For example, a C overlaps a bigger lobe of one of its sp^2 hybrid orbitals with the bigger lobe of the sp^2 orbital of the second carbon along the axis of the sp^2 orbital to make a $\sigma_{sp^2\text{-}sp^2}$ bond as illustrated in Figure 1.3.5b. The remaining sp^2 orbitals on each C make $\sigma_{sp^2\text{-}s}$ bond by overlapping with s orbital of H. Two p orbitals on adjacent carbons orient themselves parallel and overlap sideways, forming a pi-bond (π-bond). This results in the ethene molecule illustrated in Figure 1.3.5c.

The π-bond is weaker than a σ-bond because the sideways overall of two parallel p-orbitals making a π-bond is less than the head-one overlap of orbitals making a σ-bond. Therefore, a π-bond is formed only when a σ-bond can not be created. A single bond is always a σ-bond. A π-bond and a σ-bond together, i.e., C=C is called a double bond. A double bond is stronger than a single bond. The π-bond becomes more weak if the p-orbitals constituting it are not parallel because the overly is less between nonparllel p-orbitals than between parallel p-orbitals. Ther is no overall lap between p-orbitals perpendicular to each other, i.e., there is no π-bond when p-orbitals are at 90° (perpendicular). Therefore, rotation around a double does not happen until the π-bond is broken. There is no free rotation around C=C bond.

Geometry around each carbon in the ethene molecule is trigonal planar and bond angles close to the predicted value of 120°, as illustrated in Figure 1.3.5d. All six atoms in the ethene (CH_2CH_2) molecule are in the same plane and are locked in this geometry due to no free rotation around the C=C bond.

sp Hybridization of C

sp Hybridization process

One s, and one p orbitals in a valence shell can mix and produce a set of two equal hybrid orbitals, called sp orbitals, as illustrated in Figure 1.3.6a. An sp-orbital has 50% s-orbital and 50% p-orbital character. Like sp^3 and sp^2 orbitals, an sp orbital comprises two lobes, a more prominent lobe and a smaller lobe of opposite polarity along a straight line with the nucleus between the two lobes. The two sp orbitals are in a line at 180° from each other, i.e., a linear geometry with the nucleus in the middle. Two p orbitals left out from the hybridization lie perpendicular to the axis of the sp orbitals and perpendicular to each other, as illustrated in Figure 1.3.6a.

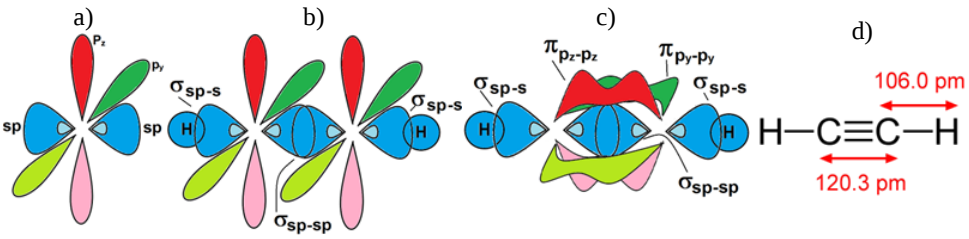


Figure 1.3.6: Cartoons of a) a set of two equal sp hybrid orbitals in a linear geometry, along with the two un-hybridized p-orbitals perpendicular to each other, b) a $\sigma_{sp\text{-}sp}$ bond formation between two sp orbitals on C atoms along with two $\sigma_{sp\text{-}s}$ bond formation between sp hybridized C orbital and s-orbital of H atoms, c) two π-bond formation between parallel p orbitals on adjacent C atoms, and d) structure showing bond lengths and angles in a CHCH molecule, (Copyright; a), b), and c) Public domain, d) Benjah-bmm27, Public domain, via Wikimedia Commons)

Formation of ethyne (HC≡CH) molecule

An sp-hybridized C overlaps a more prominent lobe of one of its sp orbitals with the bigger lobe of sp orbital of the second carbon along the axis of the sp orbitals to make a $\sigma_{sp\text{-}sp}$ bond as illustrated in Figure 1.3.6b. The remaining sp orbitals on each C make $\sigma_{sp\text{-}s}$ bond by overlapping with s orbital of H. One p orbitals on a C orient itself parallel to one p-orbital on the other and overlap sideways, forming a π-bond. The other two p orbitals are also parallel to each other and overlap sideways to form the second π-bond as illustrated in Figure 1.3.6c. This results in an ethyne molecule with a triple (C≡C) bond, i.e., a σ-bond and two π-bonds between the two C atoms, as illustrated in Figure 1.3.6d.

A triple-bond is stronger than a double-bond, and a double-bond is stronger than a single-bond. Unlike a double bond, rotation around a triple bond happens because as the π-bond starts breaking due to p-orbitals going away from parallel orientation, a new π-bond starts making with the alternate p-orbital on the neighboring C. This fact becomes apparent by comparing the π-bond electron clouds in the cases of a double bond of ethene and a triple bond of ethyne illustrated in Figure 1.3.7. However, the rotation around a triple-bond means less as its geometry is linear.

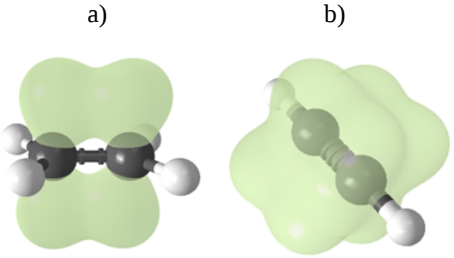


Figure 1.3.7: a) illustration of electron-cloud of a π-bond above and below the axis of the nuclei in ethene $H_2C{=}CH_2$ molecule, and b) the electron-cloud of two π-bonds around the axis of the nuclei in ethyne HC≡CH molecule, (Copyright; Jiří Janoušek, Public domain, via Wikimedia Commons)

Hybridization of N, O, and halogens

Heteroatoms like N, O, and halogen atoms (represented as: X) can also mix all or some of their valence shell s and p orbitals to have sp^3, sp^2, or sp hybridization like C. The differences are the following.

Number of covalent bonds and lone pairs in compounds

Table 1 shows the number of covalent bonds and the number of lone pairs on atoms of elements commonly found in organic compounds. C has four valence electrons that distribute one each in the hybrid orbitals and the remaining p orbitals. Lewis symbols of elements represent valence electrons around the symbol of the element. For example, Lewis symbol of C is $\cdot\dot{\underset{\cdot}{C}}\cdot$ that shows four unpaired electrons in the valence shell. Each unpaired valence electron can make one covalent bond. So, C makes four covalent bonds in organic compounds.

H has one valence electron, represented as $\dot{H}$ that shows one unpaired electron that makes one bond and no lone pair on H in their compounds.

N has five valence electrons represented in Lewis symbol as $\cdot\dot{\underset{\cdot\cdot}{N}}\cdot$ that shows three unpaired electrons and one electron pair. The unpaired electrons make one bond each while the paired electrons remain as a lone pair on N in its compounds.

O has six valence electrons represented in Lewis symbol as $\cdot\ddot{\underset{\cdot\cdot}{O}}\cdot$ that shows two unpaired electrons and to electron pairs. The unpaired electrons make one bond each while the paired electrons remain as two lone pairs on O in its compounds.

Halogen X (where X can be F, Cl, Br, or I) has seven valence electrons. For example, Cl represented as $:\ddot{\underset{\cdot\cdot}{Cl}}:$ that shows one unpaired electron that makes one bond while the paired electrons remain as three lone pairs on X in their compounds.

Table 1: Number of covalent bonds and lone pairs of electrons around elements commonly found in organic compounds.

Atom (Lewis symbol)	Number of Covalent bonds	Number of lone pairs
$\cdot\dot{\underset{\cdot}{C}}\cdot$	4	0
$\dot{H}$	1	0
$\cdot\ddot{N}\cdot$	3	1
$\cdot\ddot{\underset{\cdot\cdot}{O}}\cdot$	2	2
Halogen $:\ddot{\underset{\cdot\cdot}{X}}:$ (where X can be F, Cl, Br, or I)	1	3

Geometry around the central atom

sp^3 Hybridization

An sp^3 hybridized C makes four σ-bonds in a tetrahedral geometry around C where bonds are around 109° from eachother, as illustrated Figure 1.3.8a for the case of methane (CH_4). The electrostatic potential map of methane is also shown.

> An electrostatic potential map shows electron cloud in colors: green means neutral or nonpolar region, red means $\delta-$ region, blue means $\delta+$ region, yellow is between neutral and $\delta-$ and light blue is between neutral and $\delta+$.

Methane is nonpolar, shown by green in the electrostatic potential map.

The sp^3 orbitals of N have tetrahedral geometry where three orbitals make σ-bonds while a lone pair occupy the fourth. The three peripheral atoms are at the base forming a triangle while N is raised from the middle of the triangle resulting in a trigonal-pyramidal geometry, as illustrated in Figure 1.3.8b for NH_3 molecule. The trigonal-pyramidal geometry is a modified form of tetrahedral geometry where the fourth corner of the tetrahedron is missing due to being occupied by a lone pair. The bond angles are close to 109° as predicted by the tetrahedral geometry. The molecule is not symmetric with polar bonds, so, it is polar, which is evident from its electrostatic potential map. The red region is due to the polarity of N−H bonds and also due to the lone pair region showing red because it is in an sp^3-orbital which, unlike a p-orbital, is not symmetric with most of the electron located in the more prominent lobe on one side of the nucleus.

The sp^3 orbitals of O have tetrahedral geometry where two orbitals make σ-bonds while lone pairs occupy the other two. The two corners of a tetrahedron which are peripheral atoms and the O in the middle of the tetrahedron shape, results in a bent shape with the other two corners of the tetrahedron missing due to being occupied by lone pairs as illustrated in Figure 1.3.8c for H_2O molecule. The bent shape is a modified form of tetrahedral geometry where two corners of the tetrahedron are missing due to being occupied by lone pairs. The bond angle is close to 109° as predicted by the tetrahedral geometry. The electrostatic potential map reflects the polarity of two O−H bonds and the two lone pairs in sp^3-orbitals.

The sp^3 orbitals of halogen atoms have tetrahedral geometry where one orbital makes a σ-bond while lone pairs occupy the other three. One bond can make only a linear geometry as illustrated in Figure 1.3.8d for HCl molecule. The other three corners of the tetrahedron are missing due to being occupied by lone pairs. The electrostatic potential map reflects the polarity of the bond and the three lone pairs in sp^3-orbitals.

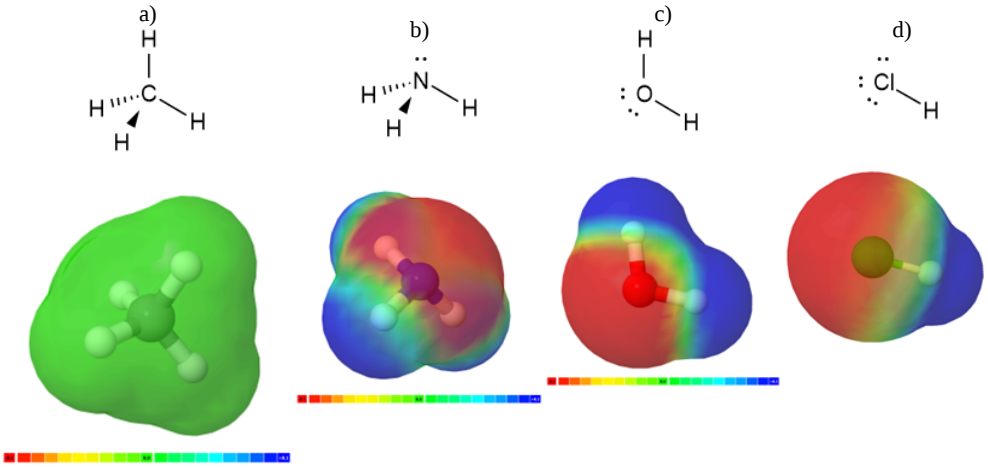


Figure 1.3.8: Structure of a) methane (CH_4) molecule, b) ammonia (NH_3) molecule, c,) water (H_2O) molecule, and d) hydrochloric acid (HCl) molecule, (Copyright; Public domain)

sp^2 Hybridization

An sp^2 hybridized C makes three σ-bonds using its sp^2 orbitals in a trigonal planer geometry around C and a π-bond using its p-orbital. The σ-bonds are around 120^o from each other, as illustrated in Figure 1.3.9a for ethene ($H_2C{=}CH_2$) molecule. The electrostatic potential map of $H_2C{=}CH_2$ in Figure 1.3.9a shows red region around the axis of the C=C bond which is due to the π-bondin electrons placed above and below the σ-bond. Further the $H's$ are bluish compared to those in CH_4 molecule indicating that the $C{-}H$ is slightly polar with $\delta+$ charge on the H.

The sp^2 orbitals of N have trigonal planer geometry where two of the sp^2 orbitals make σ-bonds, and the lone pair occupies the third sp^2 orbital. The p-orbital makes a \(\pi)-bond. The geometry around an sp^2 hybridized N is bent with a bond angle around 120^o as the trigonal planer geometry predicted. The bent geometry of sp^2 hybridized N is illustrated in Figure 1.3.9b for methanimine ($H_2C{=}NH$) molecule. The electrostatic potential map reflects the bond polarity and the lone pair of electrons in one sp^2-orbital as a red region.

The sp^2 orbitals of O have trigonal planer geometry where one sp^2 orbital makes a σ-bond, and a lone pair occupies the other two sp^2 orbitals. The p-orbital makes a \(\pi)-bond. An sp^2-hybridized O has a linear geometry around O atom, as illustrated in Figure 1.3.9c for formaldehyde ($H_2C{=}O$) molecule. The electrostatic potential map reflects the bond polarity and the lone pair of electrons in two sp^2-orbitals as a red region.

The sp^2 or hybridization of halogens is not observed except in a few cases, e.g., when one of the lone pairs is involved in resonance which will be described in a later section.

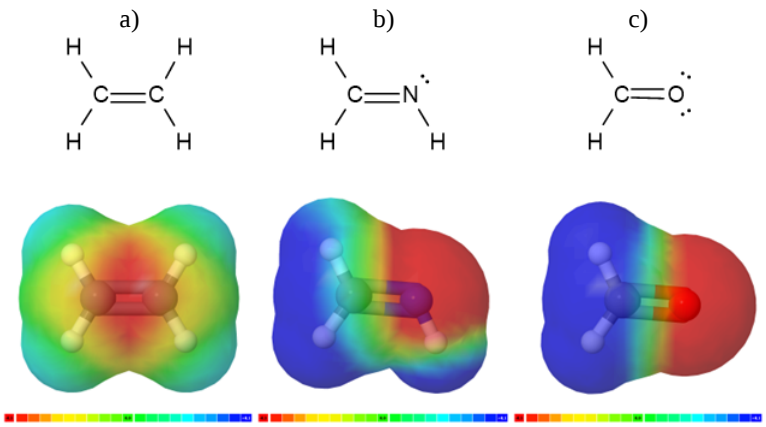


Figure 1.3.9: Structure of a) ethene (CH_2CH_2) molecule, b) methanimine (CH_2NH) molecule, and c,) formaldehyde (CH_2O) molecule, (Copyright; Public domain)

sp Hybridization

An sp hybridized C makes two σ-bonds using its sp-orbitals that are at 180^o from each other, i.e., in a linear geometry. The remaining two p-orbitals make two π-bonds perpendicular to each other and around the sigma bond resulting in a linear geometry around the sp hybridized carbon, as illustrated in Figure 1.3.10a for ethyne ($HC{\equiv}CH$) molecule. Electrostatic map shows red region around the $C{\equiv}C$ bond axis which is due to π-electrons and it also shows that the $C{-}H$ bond in ethyne is polar with $\delta+$ charge on H

The sp orbitals of N have linear geometry. One of the sp orbitals makes a σ-bond, and the other is occupied by a lone pair, and the two p-orbital make two π-bonds resulting in a triple bond. The geometry around an sp hybridized N is linear, as illustrated in Figure 1.3.10b for hydrogen cyanide (HCN) molecule. The electrostatic potential map reflects the bond polarity and shows the red region where the lone pair of electrons are located in an sp-orbital of nitrogen.

The sp hybridization of O is rare in organic compounds, but it does exist, e.g., in carbon monoxide molecule (CO) molecule. It involves formal charges on the atoms, which will be described in a later section.

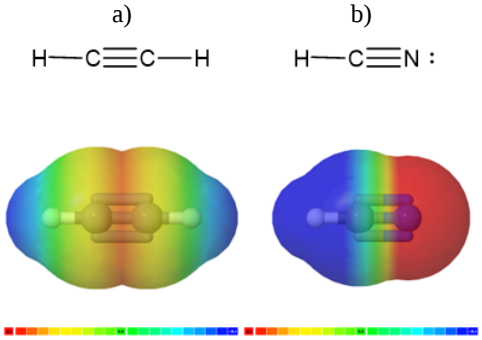


Figure 1.3.10: Structure of a) ethyne (CHCH) molecule, and b) hydrogen cyanide (HCN) molecule, (Copyright; Public domain)

1.4: Representing organic compounds

Learning Objectives

- Read and draw molecular, Lewis, condensed, and structural formulae of simple organic compounds.
- Understand the mixed versions of and slight variations in drawing the formulae of simple organic compounds.

The molecular formula, Lewis formula, condensed formula, skeletal formula, or a combination of these can represent an organic compound. These ways of representing organic compound and what they mean is described below.

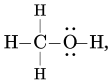
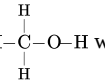
Molecular formula

The **molecular formula** tells the symbols of the elements that compose the compound, and the subscript to the element symbol denotes how many atoms of that element are in the molecule.

For example, (CH_4) is a molecule formula of methane which means there is one carbon and four hydrogen atoms in a methane molecule. C_2H_6 is a molecule formula of ethane, which means the ethane molecule has two carbon and six hydrogen atoms. Molecular formulas do not tell about the molecule's bonds and shapes.

Lewis formula

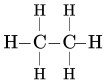
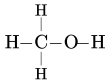
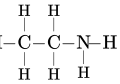
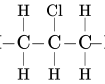
The **Lewis structure** or Lewis formula shows all the bonding electron pairs as lines (bonds) and lone pairs (non-bonding electron pairs) as pairs of dots around each atom in a molecule.

For example, Lewis formula of ethane (C_2H_6) is:  that shows each carbon is bonded with one C and three $H's$ by single bonds. Similarly, Lewis formula of formaldehyde (CH_2O) is: that shows C is bonded with two $H's$ by single bonds and with one O by a double bond and O has two lone pairs on it. The lone pairs are usually omitted from Lewis structures except when needed to emphasize their presence. For example, Lewis formula for methanol (CH_4O) with lone pairs is:  , but it can also be shown as  where the lone pairs are not shown on O but it is understood that two lone pairs are there.

Condensed formula

The Lewis formulas become complicated and time-consuming for larger organic compounds.

Condensed formula simplifies the Lewis formula by writing each C followed by H's attached with it. Subscripts are used to show more than one H's. If there is a heteroatom, i.e., any atom other than C or H in the chain, it is condensed like C, except for halogens which are condensed like H.

For example, the Lewis formula of ethane  is condensed as CH_3CH_3. Lewis formula for methanol  is condensed as CH_3OH. Lewis formula for ethanamine  is condensed as $CH_3CH_2NH_2$. Lewis formula for 2-chloropropane  , is condensed as $CH_3CHClCH_3$.

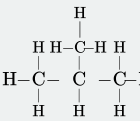
Skeletal formula

The Lewis and condensed formulas do not show the geometry of the organic compounds. Further, the Lewis and condensed formulas become complicated and time-consuming for large organic compounds.

Skeletal formulas or line-angle formulas overcome these drawbacks by simplifying the representation of organic compounds by omitting C's and H's in the formula and showing only the skeleton of the compound by C-to-C bonds as lines in a geometry that is closer to the actual geometry. Any heteroatom and H's attached to the heteroatom are shown in the skeletal formula.

Some terms related to the primary classification of organic compounds

- Before learning skeletal formulas, it is essential to understand the following terms related to the primary type of organic compounds.
- Organic compounds containing only C′s and H′s are called **hydrocarbons**. For example, ethane (CH_3CH_3), ethene (CH_2CH_2), and ethyne (CHCH) described in previous sections are hydrocarbons.
- The hydrocarbons contain only σ-bonds are called **alkanes**. For example, methane (CH_4) and ethane (CH_3CH_3) are alkanes.
  - Alkanes containing a chain of C′s where C′s are connected with either one or two C′s are called **straight chain alkanes** or normal-chain alkanes (**n-alkanes**). For example, n-alkane having four carbons is n-butane (($CH_3CH_2CH_2CH_3$), five carbons is n-pentane ($CH_3CH_2CH_2CH_2CH_3$), six carbons is n-hexane ($CH_3CH_2CH_2CH_2CH_2CH_3$), and so on.
- Alkanes in which at least on C connected with three or four other C′s are called **branched-chain alkanes**. For example, isopropane



is a branched-chain alkane
- The hydrocarbons containing at least one double bond, i.e., a σ-bond and a π-bond together, are called **alkenes**. For example, ethene

is an alkene.
- The hydrocarbons containing at least one triple-bond, i.e., a σ-bond and two π-bonds together, are called alkynes. For example, ethyne $H-C\equiv C-H$ is an alkyne.
- The hydrocarbons with a planer cyclic structure having alternating odd numbers of double bonds are a special class of hydrocarbons called **aromatic hydrocarbons** that will be described later.
- Organic compounds containing at least one heteroatom, i.e., O, N, S, P, etc. are not hydrocarbons. For example, methanol CH_3OH is not a hydrocarbon. Several classes of organic compounds are not hydrocarbons, which will be described later.

Skeletal formulas of n-alkanes

Methane (CH_4) is the simplest alkane that has a tetrahedral geometry around its C, as illustrated in Figure 1.4.1a. Gray lines in Figure 1.4.1a show the outline of the tetrahedron shape, and the black lines show C−H bonds. The plane defined by C, H on the top, and H on the right is in the plane of the paper in this perspective drawing (Figure 1.4.1a). The H on the hashed wedge is going below, and the H on the solid wedge is coming above the plane of the paper. The plane of the paper cuts through the middle of two H′s on the left of the drawing. The point of view is slightly above C towards the top-right corner.

Figure 1.4.1b shows the same structure without the tetrahedron layout drawing. All bonds are equal, and all bond angles are 109.5°. Figure 1.4.1c shows the model of CH_4 molecule from approximately the same view. Note that the geometry of CH_4 molecule is two V's of 109.5° internal angle, placed perpendicular to each other, and joined at the vertex. Figure 1.4.1d shows the same structure rotated such that the two H′s in the plane of the paper are on a straight line at the bottom of the drawing. Figure 1.4.1e shows the model rotated in the same orientation as the perspective drawing in Figure 1.4.1d.

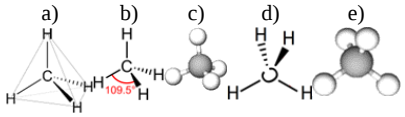


Figure 1.4.1: a) Perspective drawing of methane (CH_4) with the outlines of the tetrahedron drawing with gray lines. The plane of the paper (or the screen) is along the plane of C and two H′s to the right and cuts through the middle of the two H′s on the left (angel of view is slightly below the top-right corner) b) perspective drawing of CH_4 in the same view without drawing of the outlines of the tetrahedron, c) model of CH_4 from the same view, d) perspective drawing of CH_4 rotated to bring the two H′s in the plane of the page at the base making an inverted bigger V-shape, and e) model of CH_4 in the same view as the perspective drawing in d. (Copyright; Public domain)

Replacing any H with another sp^3 hybridized C results in ethane (CH_3CH_3) as shown in row two of Table 1. A line drawn to represent C−C bond in ethane is the skeletal formula of ethane, as shown in row two of Table 1. When 2nd H of methane is also replaced with another sp^3 hybridized C, it results in propane ($CH_3CH_2CH_3$), as shown in row three of Table 1. Two H′s of methane in the plane of the page have been replaced with C′s in this case resulting in C−C−C bonds in an inverted V-shape in the plane of the page. The skeletal formula representing a propane molecule is two lines connected in an inverted V-shape, as shown in row three of Table 1.

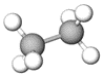
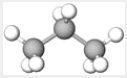
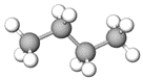
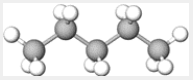
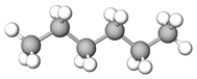
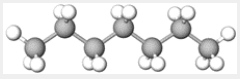
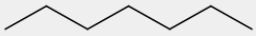
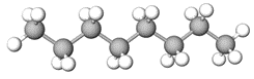
Replacing any H of a terminal C of a propane with another sp^3 hybridized C results in butane ($CH_3CH_2CH_2CH_3$) as shown in 4th row of Table 1. Remember rotation around C−C single bond happens. Therefore, the 4th C of butane can be placed at any angle relative to the plane defined by the other three C′s. However, placing all four C′s of butane in the same plane is the most stable arrangement because the terminal C′s, which are the bulkiest groups attached to the internal C′s, are farthest apart in this arrangement. Three lines connected by zigzag is the skeletal formula representing butane, as shown in row 4th of Table 1. The zigzag lines representing a chain of 4 C′s can be extended to represent a chain of 5 C′s, by adding one line to represent a chain of 6 C′s by adding two lines, and so on, as shown in Table 1 for the cases of n-alkanes having a chain of 2 to 12 C′s.

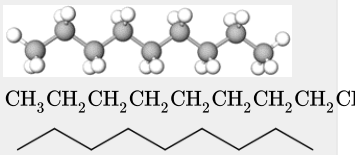
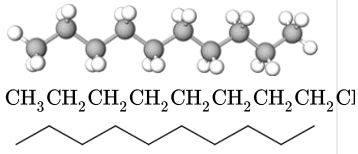
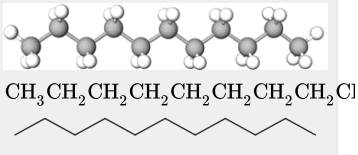
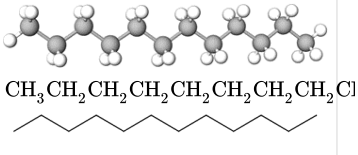
In summary, the **skeletal formula** of n-alkane is a line or lines connected zigzag representing C−C bonds. It is understood that:

- the terminals (end) and corners (bends) of the lines are C′s, and
- each C has four bonds, so the bonds which are not shown by the lines are the bonds to H′s.

With this knowledge, it is clear that a line or lines connected zigzag way are the structural formulas that represent structures of n-alkanes without showing C′s and H′s in the formula.

Table 1: Names, molecular formulas, condensed formulas, models, and skeletal formulas of n-alkanes containing 2 to 12 C′s in a chain. (Note: Methane has no structural formula as it has no C−C bond. The C′s are black and H′s are white color in the models))

# of C′s	Name	Molecular formula	Model, Condensed formula, and Structural formula
1	Methane	CH_4	 CH_4
2	Ethane	C_2H_6	 CH_3CH_3 
3	Propane	C_3H_8	 $CH_3CH_2CH_3$ 
4	Butane	C_4H_{10}	 $CH_3CH_2CH_2CH_3$ 
5	Pentane	C_5H_{12}	 $CH_3CH_2CH_2CH_2CH_3$ 
6	Hexane	C_6H_{14}	 $CH_3CH_2CH_2CH_2CH_2CH_3$ 
7	Heptane	C_7H_{16}	 $CH_3CH_2CH_2CH_2CH_2CH_2CH_3$ 
8	Octane	C_8H_{18}	 $CH_3CH_2CH_2CH_2CH_2CH_2CH_2CH_3$ 

9	Nonane	C_9H_{20}	
10	Decane	$C_{10}H_{22}$	
11	Undecane	$C_{11}H_{24}$	
12	Dodeane	$C_{12}H_{26}$	

Stem names of organic compounds

Names of n-alkanes without the last syllable, i.e., without -ane, are the **stem names** that represent the number of C′s in the organic compound. For example, meth- from methane represents one C, eth- from ethane represents two C′s, prop- from propane represents three C′s, and so on. These stem names will be described later in reference to name organic compounds.

Homologous series

It is evident from comparing the molecular formals of n-alkanes shown in Table 1 that n-alkanes differ from each other by a CH_2 or a multiple of CH_2 units. For example, add a CH_2 to methane (CH_4) to convert it to propane (C_2H_6), and two CH_2 units to convert it to butane (C_4H_{10}) and so on. Series of organic compounds that differ from each other by a CH_2 or multiple of CH_2's are **homologous series**. All of the n-Alkanes shown in Table 1 are members of a homologous series.

General formula of n-alkanes

The general formula of n-alkanes is $C_nH_{2\,n+2}$ where n is a counting number, i.e., 1, 2, 3,... For example, when n = 1, it is methane $C_1H_{2\times1+2}$ = CH_4 (Recall: that when the subscript to the element symbol in the formal is 1, it is not written.). When n = 2, it is ethane $C_2H_{2\times2+2}$ = C_2H_6, and so on

Variations in the structural formula to represent different configurations

The structural formulas in Table 1 represent the linear conformation of n-alkanes which is the most stable confirmation. Remember: rotation is possible around any single bond. For example, n-hexane shown in Figure 1.4.2a is rotated around the middle C−C bond to acquire new confirmation shown in Figure 1.4.2b and rotated further along the 2nd C−C from left to acquire another confirmation shown in Figure 1.4.2c. All three structural formulas are shown in Figure 1.4.2a, b, and c represent the same molecule. Remember: different shapes of the same molecule obtained by rotation around a single bond are different **configurations** of the same molecule.

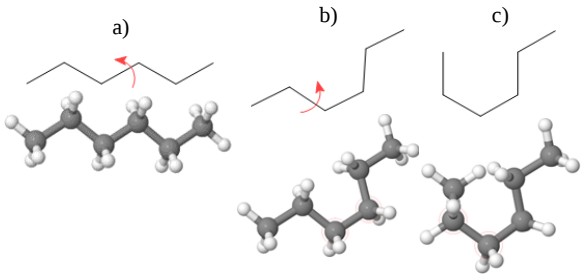
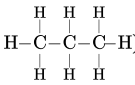
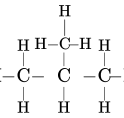
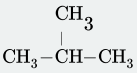


Figure 1.4.2: a) n-hexane in a linear configuration and its model in the same configuration, b) the n-hexane in a bent configuration rotated ~120° around the middle C−C bond indicated by the arrow, i.e., between C′s in red circles and the model in the same configuration, and c) the n-hexane in an other bent configuration rotated ~120° again around the 2nd C−C bond indicated by the arrow, from the left, i.e., between C′s in red circles and the model in the same configuration. (Copyright; Public domain).

Condensed and skeletal formulas of branched-alkanes

Replacing one or both H′s of a non-terminal carbon of straight chain alkane (n-alkane) with a C or a C chain results in a branched-alkane. Replacing hydrogen on the middle carbon of propane () with a C results in isopropane () which is a branched chain alkane.

There are two ways to show the condensed formula of the branched alkane:

- show the condensed formula of the branch within small brackets next to the carbon it is attached to, e.g., $CH_3CH(CH_3)CH_3$ is the condensed formula of isopropane;
- show the condensed formula of the branch hanging above or below the carbon it is attached, e.g.,  is the condensed formula of isopropane.

The skeletal formula of the branched alkane shows the skeletal formula of the branch with the terminal connected to the carbon in the main branch to which it is attached. For example,  is the skeletal formula of isopropane. Figure 1.4.3 shows another example of branched alkane and its condensed and skeletal formulas.

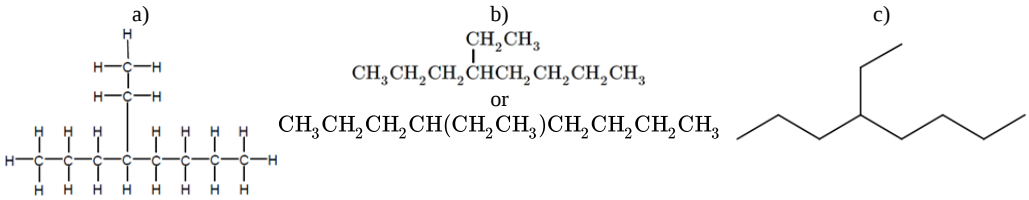


Figure 1.4.3: 4-Ethyloctane -a branched alkane: a) Lewis formula, b) two ways of writing condensed formulas, and c) its skeletal formula. (Copyright; Public domain).

Skeletal formulas of alkenes

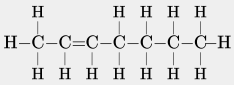
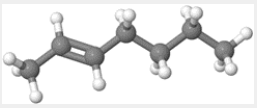
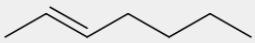
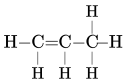
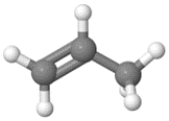
Hydrocarbons containing at least one double bond (C=C bond) are called alkene.

Since the two sp^2 C′s at the double bond have trigonal planer geometry with bond angles 120°, the double bond between the two sp^2 C′s fit in the zigzag skeletal structure, like alkanes, except that two lines are drawn where there is a double bond in the chain.

Table 2 shows the names, Lewis structures, models, condensed formulas, and skeletal formulas of two alkene examples. Their nomenclature will be explained later.

Table 2: Names, Lewis structures, models, condensed formulas, and skeletal formulas of some alkene examples. (Note: C′s are black and H′s are white color in the models)

Name	Lewis structure	Model, Condensed formula, and Structural formula

Hept-2-ene		 $CH_3CHCHCH_2CH_2CH_2CH_3$ 
Propene		 CH_2CHCH_3 

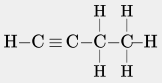
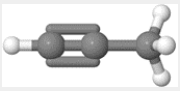
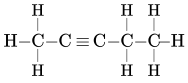
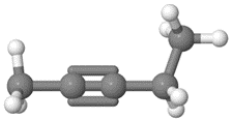
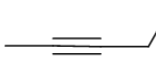
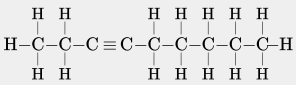
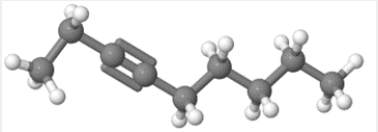
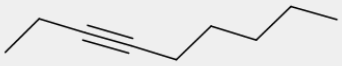
Skeletal formulas of alkynes

Hydrocarbons containing at least one triple bond ($C \equiv C$) bond are called alkyne.

> Since the two sp C's at the triple-bond have linear geometry, three lines show the triple bond and the two single bonds to the other atoms attached to them are drawn in line with the triple bond. The zigzag skeletal structure shows the rest of the structure as usual.

Table 3 shows some alkyne examples' names, Lewis structures, models, condensed formulas, and skeletal formulas. Their nomenclature will be explained later.

Table 3: Names, Lewis structures, models, condensed formulas, and skeletal formulas of some alkyne examples. (Note: C's are black and H's are white color in the models)

Name	Lewis structure	Model, Condensed formula, and Structural formula
Propyne		 $CHCCH_3$ 
Pent-2-yne		 $CH_3CCCH_2CCH_3$ 
Non-3-yne		 $CH_3CH_2CCCH_2CH_2CH_2CH_2CH_3$ 

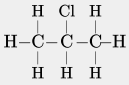
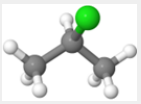
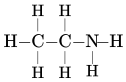
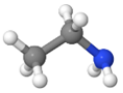
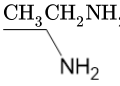
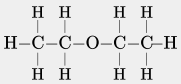
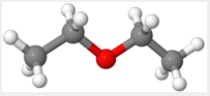
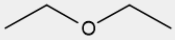
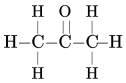
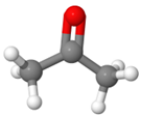
Skeletal formulas of organic compounds containing heteroatoms

> If any heteroatom, i.e., an atom other than C or H, is present in an organic compound:
>
> - its symbol in the skeletal formula shows it and

- any H present on the heteroatom is also shown next to the heteroatom as in the condensed formulas.

The skeletal structure shows the rest of the structure as usual. Table 4 shows the names, Lewis structures, models, condensed formulas, and skeletal formulas of some example organic compounds containing heteroatoms. Their nomenclature will be explained later.

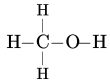
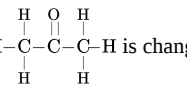
Table 4: Names, Lewis structures, models, condensed formulas, and skeletal formulas of some organic compounds containing heteroatoms. (Note: C′s are black, H′s are white, and heteroatom are in other colors in the models)

Name	Lewis structure	Model, Condensed formula, and Structural formula
2-Chloropropane		 $CH_3CHClCH_3$ 
Ethanamine		 $CH_3CH_2NH_2$ 
Diethyl ether		 $CH_3CH_2OCH_2CH_3$ 
Acetone		 CH_3COCH_3 or $CH_3\overset{O}{\overset{\|}{C}}CH_3$ 

Variations in the formulas representing organic compounds

The condensed and the skeletal formulas are usually presented in organic chemistry, as described in the previous sections. However, they may be varied slightly according to the need.

If needed, mixed Lewis, condensed, and skeletal structures emphasize a particular part of the molecule.

For example, the H′s are usually written to the right of C or to the right of heteroatoms in the condensed formula, but sometimes this order may be reversed. For example, Lewis formula for methanol  is presented in condensed form as CH_3OH, but it may be condensed as H_3COH, or as H_3C-OH to emphasize the C−O-bond. Similarly, Lewis formula of acetone  is changed to skeletal formula as  , but it may also be shown as  to emphasize the C−H bond.

Sometimes the formula is further condensed by placing the repeating units within a bracket and subscript outside the bract to show the number of repeat units. For example, the condensed formula of isopropane is $CH_3-\underset{\displaystyle CH_3}{\underset{|}{CH}}-CH_3$ that has three CH_3 units attached to the central C. It can be further condensed as $(CH_3)_3CH$. Similarly, the condensed formula of decane $CH_3CH_2CH_2CH_2CH_2CH_2CH_2CH_2CH_2CH_3$ contains eight CH_2 units and can be condensed further as $CH_3(CH_2)_8CH_3$.

Examples of drawing Lewis, condensed, and skeletal formulae

Example 1.4.1

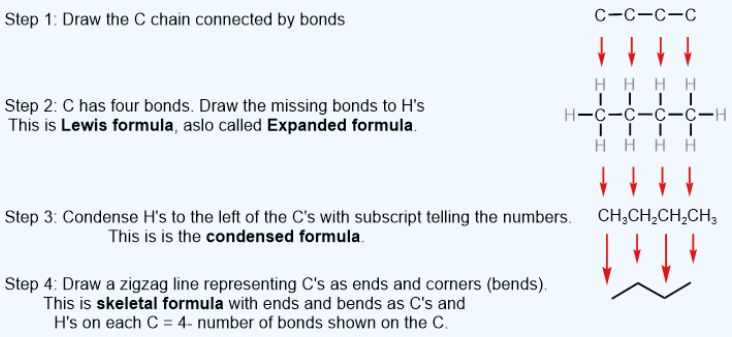
Draw molecular, Lewis, condensed, and skeletal formulas of butane.

Solution

But- in butane means four C′s and -ane means it is an alkane. Substituting n = 4 in the general formula of alkanes ($C_nH_{n\times 2+2}$) gives:

molecular formula = $(C_4H_{4\times 2+2}=C_4H_{10}$.

For Lewis, condensed, and skeletal formulas follow the steps shown in the figure below.

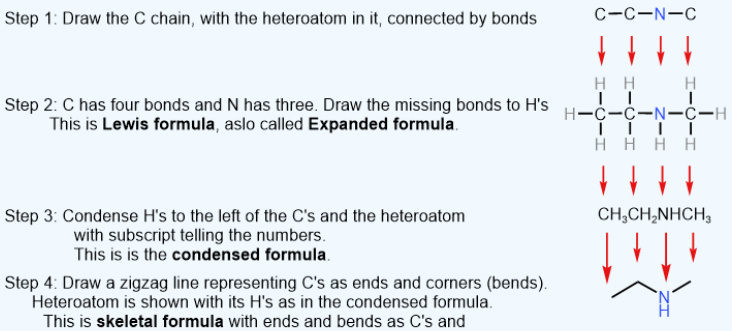


Example 1.4.2

Draw Lewis, condensed, and skeletal formulas for methylethylamine

Solution

Meth- means one C, eth-means two C′s, and aminie indicates a N two which the carbon chains are attached. With this information, follow the steps shown in the figure below.



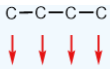
Example 1.4.3

Draw Lewis, condensed, and skeletal formula of 1-chlorobutane

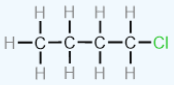
Solution

But- means four C′s chain and 1-chloro- means there is chlorine at the terminal C. The nomenclature will be explained later. With this information, follow the steps shown in the figure below.

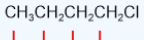
Step 1: Draw the C chain, with the heteroatom in it, connected by bonds



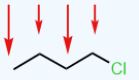
Step 2: C has four bonds. Draw the missing bonds to H's.
Terminal C has one missing bond with Cl, add one Cl.
This is **Lewis formula**, aslo called **Expanded formula**.



Step 3: Condense H's to the left of the C's and trad Cl as H.
with subscript telling the numbers.
This is is the **condensed formula**.



Step 4: Draw a zigzag line representing C's as ends and corners (bends).
Heteroatom is shown with its H's (if any) as in the condensed formula.
This is **skeletal formula** with ends and bends as C's and
H's on each C = 4- number of bonds shown on the C.



1.5: Formal Charge

Learning Objectives

- Calculate the formal charge on atoms in organic molecules.
- Understand the geometry of and the location of valence electrons in the hybrid orbitals of species with formal positive charge, formal negative charge, and organic radicals.

Prelude to a formal charge

Isolated atoms are neutral because they have equal protons and electrons. When an atom loses a a valence electron it becomes +ve species or cation, e.g., Na after loosing an electron becomes Na^+, and Ca after loosing two electron becomes Ca^{2+}. When an atom gains a valence electron, it becomes -ve species or anion, e.g., Cl after gaining an electron becomes Cl^-, and O after gaining two electrons becomes O^{2-}. Covalently bonded species can also be ions, e.g., NH_3^+, and H_3O^+ are cations, and HO^-, and CO_2^{2-} are anions. Organic compounds often become ions as intermediates in reactions.

What is the formal charge?

A formal charge is assigned to an atom in a molecule based on the assumption that bonding electrons are shared equally.

How to calculate the formal charge?

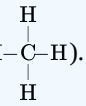
The formal charge is calculated by subtracting an atom's valence electrons in a molecule from the valence electron of the isolated atom. Valence electrons of an isolated atom are equal to the first digit of the group number in the periodic table. For example, H is in group 1 and has one valence electron shown as: $\dot{H}$ in Lewis symbol. C is in group 14 and has four valence electrons shown as: $\cdot\dot{C}\cdot$, i.e., four unpaired dots in its Lewis symbol. N is in group 15 and has five valence electrons shown as: $\cdot\ddot{\dot{N}}\cdot$, i.e., five dots (one paired and three unpaired) in its Lewis symbol. Halogens are in group 17 and have seven valence electrons, e.g. $\cdot\ddot{\dot{C}}l:$, i.e., shown as three paired and one unpaired dots. Valence electrons in a molecule are assigned to an atom, assuming that the bonding electrons are equally shared, i.e., one electron to the atom per one bond. All the nonbonding electrons are assigned to the atom they are on.

The formula for calculating formal charge is:

$$Fc = Ve - (B + Nb)$$

, where Fc is the formal charge, Ve is the valence electrons in an isolated atom, B is the number of bonds attached to the atom, and Nb is nonbonding electrons on the atom in the molecule.

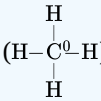
Example 1.5.1

What is the formal charge on C in methane ().

Solution

Isolated C has four valance electrons, there are four bonds and no nonbonding electron in the molecule, so: Ve = 4, B = 4, Nb = 0

$Fc = Ve - (B + Nb)$ = 4 - (4 + 0) = 0

Answer: 0 formal charge, i.e., C in methane (), where superscript over the atom (C in this case) show the formal charge

Reactive intermediates and their formal charges

If one of the C−H bonds in an organic compound breaks, three situations may arise, e.g., for $H-\overset{H}{\underset{H}{C}}-H$ the situations are:

1. one of the bonding electrons may remain on the C, i.e., $(H-\overset{H}{\underset{}{\dot{C}}}-H)$,
2. none of the bonding electrons may remain on the C, i.e, $(H-\overset{H}{\underset{}{C}}-H)$, and
3. both the bonding electrons may remain on the C, i.e., $(H-\overset{H}{\underset{}{\ddot{C}}}-H)$.

Each of these situations results in a C with three bonds which are reactive intermediates. They are short-lived and tend to react to re-establish regular four bonds around the C. Figure 1.5.1 shows the structure and hybridization of the reactive carbon species. A specie with a single unshared electron on an atom is a free radical that tends to make a covalent bond by sharing the electron. A $H-\overset{H}{\underset{}{\dot{C}}}-H$ is an example of a free. The $(H-\overset{H}{\underset{}{\dot{C}}}-H)$, is sp^2 hybridized with three bonds and an unshared electron in its p-orbital, as illustrated in Figure 1.5.1a.

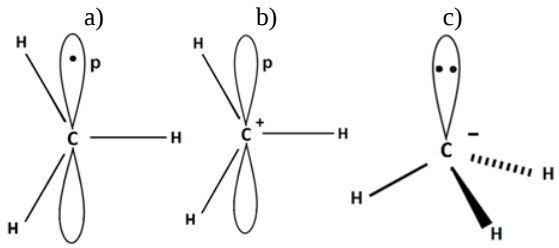


Figure 1.5.1: Illustration of a) a free radical $H-\overset{H}{\underset{}{\dot{C}}}-H$ that is sp^2 hybridized with three $\sigma_{sp^2\text{-}s}$ bonds, and a p-orbital occupied by a single electron, b) a carbocation $H-\overset{+}{\underset{H}{C}}-H$ that is sp^2 hybridized with three $\sigma_{sp3\text{-}s}$ bonds, and a vacant p-orbital, and c) a cabanion $H-\overset{H}{\underset{}{\ddot{C}}}-H$ this is sp^3 hybridized with three $\sigma_{sp^3\text{-}s}$ bond and one sp^3 orbital occupied by a lone pair. (Copyright; Public domain)

Example 1.5.2

What is the formal charge on C in $H-\overset{H}{\underset{}{\dot{C}}}-H$.

Solution

Isolated C has four valance electrons, there are three bonds and one nonbonding electron on C in the molecule, so: Ve = 4, B = 3, Nb = 1

$Fc = Ve - (B + Nb)$ = 4 - (3 + 1) = 0

Answer: 0 formal charge, i.e. C in $H-\overset{H}{\underset{}{\dot{C}^{0}}}-H$.

Example 1.5.3

What is the formal charge on C in $H-\overset{H}{\underset{}{C}}-H$.

Solution

Isolated C has four valance electrons, there are three bonds and no nonbonding electron on C in the molecule, so: Ve = 4, B = 3, Nb = 0

$Fc = Ve - (B + Nb)$ = 4 - (3 + 1) = +1

Answer: +1 formal charge, i.e. (H–$\overset{+1}{\underset{\text{H}}{\text{C}}}$–H).

Example 1.5.4

What is the formal charge on C in H–$\underset{\text{H}}{\ddot{\text{C}}}$–H.

Solution

Isolated C has four valance electrons, and there are three bonds and two nonbonding electrons on C in the molecule, so: V_e = 4, B = 3, N_b = 2

$Fc = Ve - (B + Nb)$ = 4 - (3 + 2) = -1

Answer: 0 formal charge, i.e. H–$\overset{-1}{\underset{\text{H}}{\ddot{\text{C}}}}$–H.

- A radical is a C with three bonds and one unshared electron. C in H–$\underset{\text{H}}{\dot{\text{C}}}$–H is an example of a radical that is sp^2 hybridized with three bonds and a p-orbital is partially filled by a single electron, as illustrated in Figure 1.5.1a. A radical tends to make a bond by accepting an electron in its partially filled p-orbital.
- A carbocation is a C with three bonds and no unshared electrons. A H–$\overset{+1}{\underset{\text{H}}{\text{C}}}$–H is an example of carbonation that is sp^2 hybridized with three bonds and a vacant p-orbital, as illustrated in Figure 1.5.1b. A carbocation tends to make a bond by accepting a lone pair in its vacant p-orbital.
- A carbanion is a C with three bonds and two unshared electrons. H–$\overset{-1}{\underset{\text{H}}{\ddot{\text{C}}}}$–H is an example of a carbanion that is sp^3 hybridized with three bonds and the fourth sp^3 orbital occupied by a lone pair, as illustrated in Figure 1.5.1c. A carbanion tends to make a bond by donating the lone pair.

Species containing N and O with formal charges

N has three bonds and one lone pair, e.g., in ammonia H–$\underset{\text{H}}{\ddot{\text{N}}}$–H and O has two bonds and two lone pairs, e.g., water H–$\underset{\text{H}}{\ddot{\text{O}}}$: molecule. One more bond results in a cation, and one fewer bond results in anion species, as shown by solving the formal charges in the following examples.

Example 1.5.5

What is the formal charge on a) N in H–$\underset{\text{H}}{\ddot{\text{N}}}$–H and b) on O in H–$\underset{\text{H}}{\ddot{\text{O}}}$:

Solution

a) Isolated N has five valance electrons, there are three bonds and two nonbonding electrons (a lone pair) in the molecule, so: Ve = 5, B = 3, Nb = 2

$Fc = Ve - (B + Nb)$ = 5 - (3 + 2) = 0

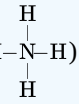
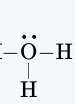
Answer: 0 formal charge, i.e., N in H–$\overset{0}{\underset{\text{H}}{\ddot{\text{N}}}}$–H

a) Isolated O has six valance electrons, there are two bonds and four nonbonding electrons (two lone pairs) in the molecule, so: Ve = 6, B = 2, Nb = 4

$Fc = Ve - (B + Nb)$ = 6 - (2 + 4) = 0

Answer: 0 formal charge, i.e., N in O in

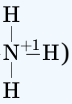
Example 1.5.6

What is the formal charge on a) N in  and b) on O in 

Solution

a) Isolated N has five valance electrons, there are four bonds and no nonbonding electrons in the molecule, so: Ve = 5, B = 4, Nb = 0

$Fc = Ve - (B + Nb)$ = 5 - (4 + 0) = +1

Answer: +1 formal charge, i.e., N in N in 

a) Isolated O has six valance electrons, there are three bonds and two nonbonding electrons (one lone pair) in the molecule, so: Ve = 6, B = 3, Nb = 2

$Fc = Ve - (B + Nb)$ = 6 - (3 + 2) = +1

Answer: +1 formal charge, i.e., O in 

Example 1.5.7

What is the formal charge on a) N in  and b) on O in 

Solution

a) Isolated N has five valance electrons, there are two bonds and four nonbonding electrons (two lone pairs) in the molecule, so: Ve = 5, B = 2, Nb = 4

$Fc = Ve - (B + Nb)$ = 5 - (2 + 4) = -1

Answer: -1 formal charge, i.e. N in 

a) Isolated O has six valance electrons, there is one bond and six nonbonding electrons (three lone pair) in the molecule, so: Ve = 6, B = 1, Nb = 6

$Fc = Ve - (B + Nb)$ = 6 - (1 + 6) = -1

Answer: +1 formal charge, i.e., O in

1.6: Resonance

Learning Objectives

- Understand the concept of resonance and how it stabilizes molecules.
- Draw resonance contributors and predict equal, major, and minor resonance contributors.
- Predict resonance stabilization and factors that affect it.
- Predict the relative strength of a π-bond based on orientation and sizes of p-orbitals.

Prelude to resonance

C not only makes straight chains and branched chains but makes cyclic chains, as illustrated in Figure 1.6.1. Benzene is a member of a class of hydrocarbons called aromatic hydrocarbons. A set of characteristics of aromatic compounds will be described later. Here only the nature of bonds in benzene is described to introduce the concept of resonance.

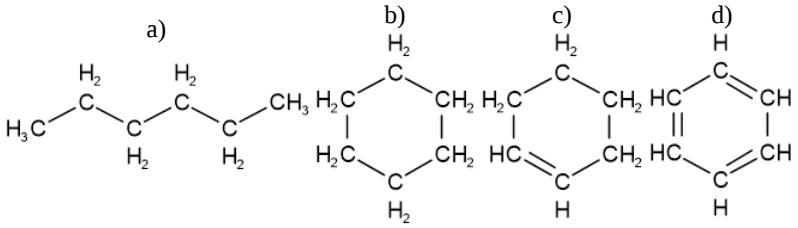


Figure 1.6.1: Semi-condensed formulas of a) a six C straight chain with all single bonds called n-hexane, b) a six C cycle with all single bonds called cyclohexane, c) six C cycle with one double bond called cyclohexene, and d) a six C cycle with all C′s sp^2 hybridized called benzene. (Copyright; Public domain)

More than one correct Lewis structure can be drawn for some compounds or polyatomic ions.

If more than one Lewis structure can be drawn for a compound, these are called **contributing structures** or **resonance contributors**.

The contributing structures of a compound are separated from each other by double-headed arrows, as shown in Figure 1.6.2, for the case of benzene molecule and carbonate (CO_3^{2-}) polyatomic anion.

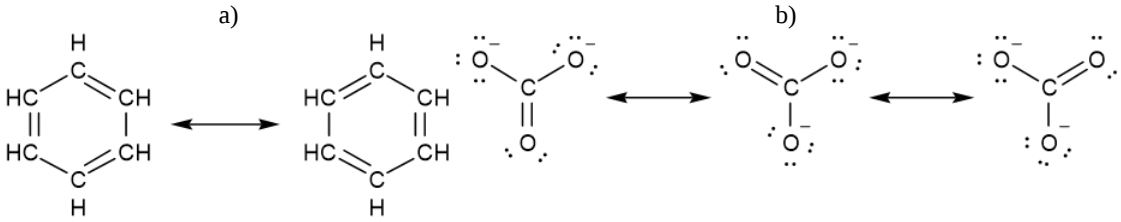


Figure 1.6.2: The resonace contributors of benzene (C_6H_6) and carbonate (CO_3^{2-})ion. (Copyright; Public domain)

Are the compounds or polyatomic ions with more than one contributing structure a mixture of these structures? The answer is no; none of the contributing structures exist. For example, both the contributing structures of benzene imply that three are three C−C single bonds and three C=C double bonds. Experimental results show that all six C−C bonds in benzene are equal. Similarly, all three contributing structures of CO_3^{2-} imply one double bond and two single bonds. Experimental results show that all three C−O bonds in CO_3^{2-} are equal. The concept of resonance was introduced to deduce the actual structure in the cases of compounds having two or more contributing structures.

What is resonance?

- **Resonance** is a way of describing bonding in molecules and polyatomic ions by a combination of contributing structures into the actual structure.
- The hybrid of the contributing structures -the actual structure, is called a **resonance hybrid**.

Consider the structure of a benzene molecule: every C in benzene is sp^2 hybridized with three σ-bonds in a trigonal planar geometry and a p-orbital perpendicular to the plane, as illustrated in Figure 1.6.3a. All C′s and H′s are in one plane, and all the p-orbitals are perpendicular to them and parallel to each other. Assuming p-orbitals in benzene simultaneously make a half-π-bond with two adjacent p-orbitals, it can be concluded that all of the C-to-C bonds are equal, i.e., each one is composed of one σ and a half π-bond, as illustrated in Figure 1.6.3b. The same structure is assumed by a fifty%-fifty% hybrid of its two contributing structures. The bonding in the resonance hybrid of benzene agrees with the experimental results, i.e., all six C−C bonds in benzene are equal. Similarly, the resonance hybrid of CO_3^{2-}, arrived at by assuming it is a hybrid of its three equal contributing structures, is illustrated in Figure 1.6.3c. The bonding in the resonance hybrid of CO_3^{2-} agrees with the experimental results, i.e., all three C−O bonds in CO_3^{2-} are equal.

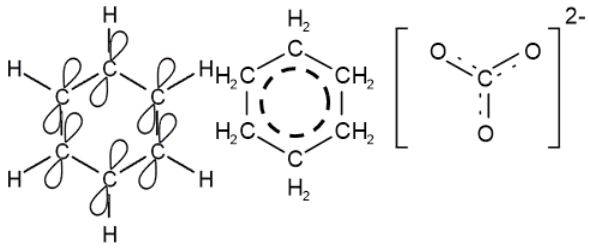


Figure 1.6.3: Illustration of a) sp^2 hybridization of each carbon of a benzene ring, b) a resonance hybrid of benzene, and c) a resonance hybrid of CO_3^{2-} ion. (Copyright; Public domain)

Note that the π bond is not localized between two adjacent atoms in resonance hybrids; it is spread over three or more adjacent atoms.

> The π-bond spread over three or more adjacent atoms is called a **delocalized π-bond** and the elections in the delocalied π-bond are called **delocalized electrons**.

For example, the delocalized π-bonds are represented by a dashed-lined circle inside the benzene ring or by dashed lines in the case of CO_3^{2-} resonance hybrids.

How is an sp^3 atom bearing a lone pair involved in resonance?

An sp^3-hybridized atom with lone pair changes its hybridization from sp^3 to sp^2 and places its lone pair in the p-orbital for resonance to happen when it is adjacent to an sp^2- or sp-hybridized atom.

Delocalization of π- or nonbonding electrons results in more bonding, i.e., delocalized electrons are spread over more than two nuclei. It results in lowering the potential energy, called resonance stabilization. The resonance hybrid is always more stable than the predicted stability of any of its contributing structures.

Summary of facts about resonance

1. When more than one correct Lewis structure can be drawn for a compound, these are called **resonance contributing structures**.
2. The actual structure of the compound that is deduced by assuming it is a hybrid of the contributing structures is called the **resonance hybrid**.
3. The extra stability of the resonance hybrid due to the resonance is called **resonance stabilization**.
4. The **contributing structures do not exist**; they are just the correct Lewis structures that help deduce the resonance hybrid. Consequently, the double-headed arrow does not show a chemical reaction; it is used only to separate the contributing structures of a compound.

How to draw a resonance contributor?

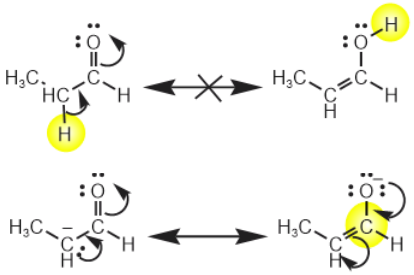
If one Lewis structure is already drawn, its resonance contributor can be drawn by the following rules, as illustrated with examples:

- Move π-electrons or lone pair only (do not move σ-electrons) to create a new π-bond or lowne pair.

- Curved arrows show the movement of electrons, the arrow starts where the electrons are located, and the arrowhead points where they are going to create a new π bond or a new nonbinding electron (do not make a new σ-bond).
- The electron movement often creates formal charges. Calculate the formal charge and show it if it is not zero. For example, π-electrons are moved to create a new lone pair and formal charges in the following example.

- For resonance to happen, there must be at least two adjacent sp^2 or sp atoms. Resonance can not happen involving sp^3 atom.
  - An exception is an sp^3 atom carrying a lone pair; it changes its hybridization to sp^2 and places its lone pair in the p-orbital for resonance, as shown in the following example.

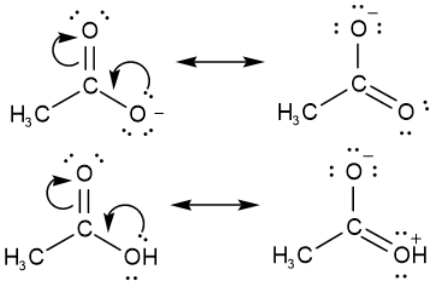
- Atoms do not move, i.e., a σ-bond is not broken or created during regular resonance. The first example below is not resonance because a $\mathrm{C-H}$ σ-bond from a $-\mathrm{CH_2}-$,is broken and a $\mathrm{O-H}$ σ-bond is formed. It is a chemical reaction, not a resonance. The second example is resonance, as only a lone pair and a π-bond pair are moved.



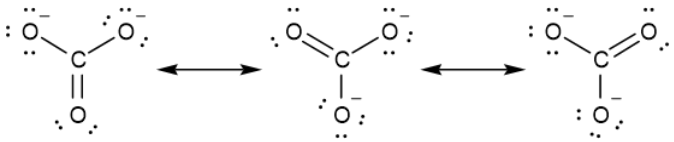
- A H can not have more than two valence electrons (duet rule), and second-row elements, like C, N, O, F, etc. can not have more than eight valence electrons (octet rule) in any of the resonance contributors. (Third-row and higher-row elements are exceptions and may have up to twelve valence electrons in molecules). For example, the movement of a lone pair to create a π-bond in the following example is not allowed as it makes a C with ten valence electrons (five bonds). Simultaneous movement of the $\mathrm{C=O}$ π-bond avoids this problem, as in the previous example.

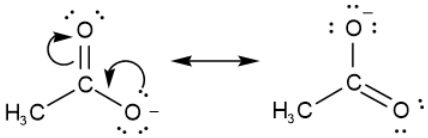
The relative importance of the contributing structures

- There is more resonance stabilization for equal resonance contributors than for unequal resonance contributors. The first example below has more resonance stabilization due to equal contributors than the second one with unequal contributors.



- More the number of equal contributors means more resonance stabilization. For example, resonance stabilization is more for $\mathrm{CO_3^{2-}}$ with three equal contributors than for $\mathrm{CH_3COO^-}$ with two equal contributors.

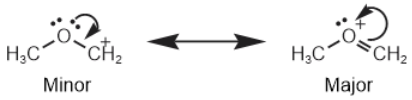




In the cases of unequal contributors, the following factors determine the major contributor:

- The negative charge on a more electronegative atom or positive charge on a less electronegative atom is more stable than otherwise. The contributor on the right in the following example with a negative charge on more electronegative O is a major contributor, and the other is minor.

- A contributor with an octet incomplete on an atom is significantly less stable than the contributor with all atoms octet complete.
- A contributor with more bonds is significantly more stable than one with fewer bonds. In the following example, the contributor on the left with octet incomplete on one C and fewer bonds is a negligible contributor relative to the other.



- A contributor with more formal charges is significantly less stable than a contributor without formal charges. The contributor on the left, with two formal charges and fewer bonds, is a negligible contributor compared to the other.

Strength of the π-bond and resonance in relation to size and orientation of the p-orbitals

Both the π-bond and, consequently, the resonance are strong when the two p-orbitals are of the same size and parallel. The strength of the π-bond and the resonances become weaker when

- the two p-orbitals are out of parallel orientation and become zero when the two p-orbitals are perpendicular to each other, and
- when the size of the two p-orbitals differ, the more significant the difference, the weaker the overlap.

These two effects are illustrated in the Figure below.

The two p-orbitals are parallel, resulting in the maximum π-bond strength and the maximum resonance

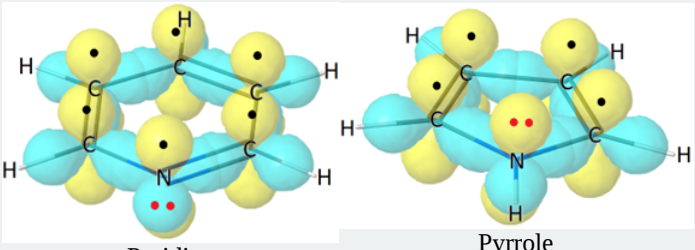
The two p-orbitals are perpendicular, with no π-bond and no resonance

The two p-orbitals are parallel but of different sizes resulting in a weaker π-bond and weaker resonance.

- In addition to the two effects described above, the third factor is electronegativity, i.e., the less electronegative the atom donating a lone pair of electrons for the resonance, the stronger the resonance.

The following examples support the three effects described above.

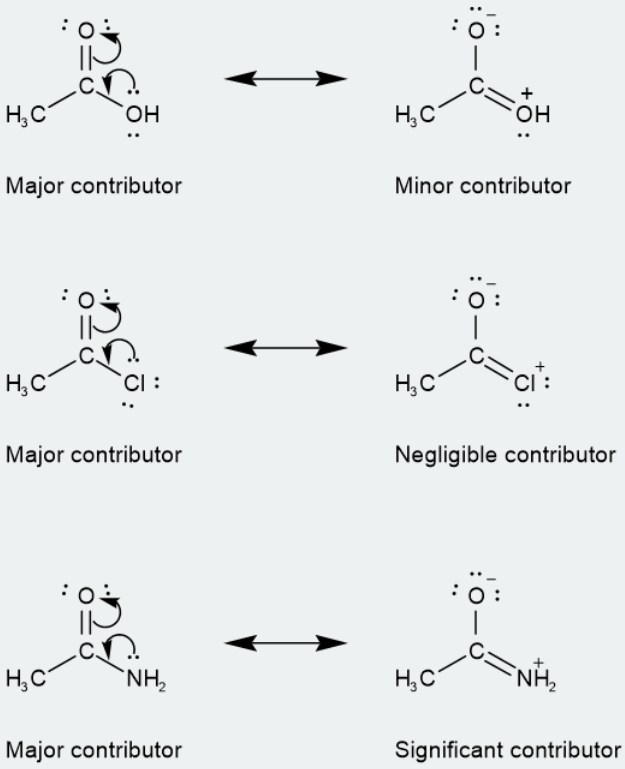
The N in pyridine can donate its lone pair of electrons to a proton, i.e., it is a base (pKb 8.8) because the lone pair is in an sp^2-orbital which is perpendicular to the other p-orbitals and not occupied in resonance. The N in pyrrole is not able to donate its lone pair of electrons to a proton, i.e., it is a very weak base (pKb 17.8) because the lone pair is in a p-orbital parallel to other p-orbitals and occupied in resonance and not that much available to protons, as shown in Figure 1.6.4.



Pyridine Pyrrole

Figure 1.6.4: Illustration of p-orbitals (yellow lobes) and sp^2-orbitals (blue lobes) in pyridine (C_5H_5N) and pyrrole (C_4H_4NH). Note: the lone pair (red dots) in sp^2-orbital is perpendicular to the other p-orbitals in pyridine but in a p-orbital parallel to the other p-orbital in pyrrole. (Copyright; Public domain)

Three examples of resonance are shown below, where the second contributor in each case is minor because there is charge separation, and +ve charge is placed on an electronegative atom. However, there are differences in these three cases, i.e., i) the second contributor is minor in the 1st example where 2p-orbital of O is in resonance with 2p-orbital of C which are similar in sizes, but negligible in the second example because 3p-orbital of Cl is in resonance with 2p-orbital of C which are different in sizes, ii) the second contributor is minor in the 1st example but significant in the 3rd example because the p-orbitals are similar in size but N is less electronegative and more willing to donate its lone pair than the O in the 1st example.



CHAPTER OVERVIEW

2: Nomenclature and physical properties of organic compounds

2.1: Hydrocarbons

Learning Objectives

- Identify hydrocarbons and their subclasses: alkane, alkene, alkyne, and aromatic.
- Write IUPAC name for a skeletal structure and draw the skeletal structure from the IUPAC name for simple hydrocarbons.
- Understand some physical properties like melting points, boiling points, and solubilities of hydrocarbons.
- Predict the reactive sites, i.e., δ- (nucleophilic) and δ+ (electrophilic) regions in the structures of simple hydrocarbons.
- Identify primary, secondary, tertiary, or quaternary C's and H's.

Classification of hydrocarbons

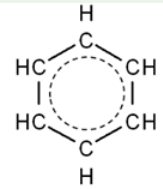
What are hydrocarbons

Organic compounds composed of C's and H's only are called **hydrocarbons**.

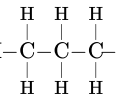
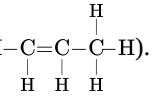
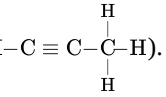
The hydrocarbons are divided into two major classes based on the absence or presence of an aromatic ring.

Aromatic ring

An **aromatic ring** is a cyclic chain of only sp^2-hybridized atoms in planer geometry having π-electrons in the ring equal to 4n+2, where n is a positive integer, i.e., 0, 1, 2, 3, ...,. Benzene is an example of an aromatic ring that is a cycle of six sp^2-hybridized C's having six π-electrons, which is a number equal to 4n+2 for n = 1.



Benzene

- Hydrocarbons with at least one aromatic ring in their structure are called **aromatic hydrocarbons**.
- Hydrocarbons that do not have an aromatic ring in their structure are called **aliphatic hydrocarbons**. The aliphatic hydrocarbons are subdivided into i) alkanes, ii) alkenes, and iii) alkynes. Figure $2.1.1$ illustrates the classification of hydrocarbons.
  - **Alkanes** have all $\mathrm{C}-\mathrm{C}$ single bonds, i.e., all the C's are sp^3-hybridized, e.g., propane ().
  - **Alkenes** have at least one $\mathrm{C}=\mathrm{C}$ double bond, e.g., propene ().
  - **Alkynes** have at least one $\mathrm{C} \equiv \mathrm{C}$ triple bond, e.g., propyne ().

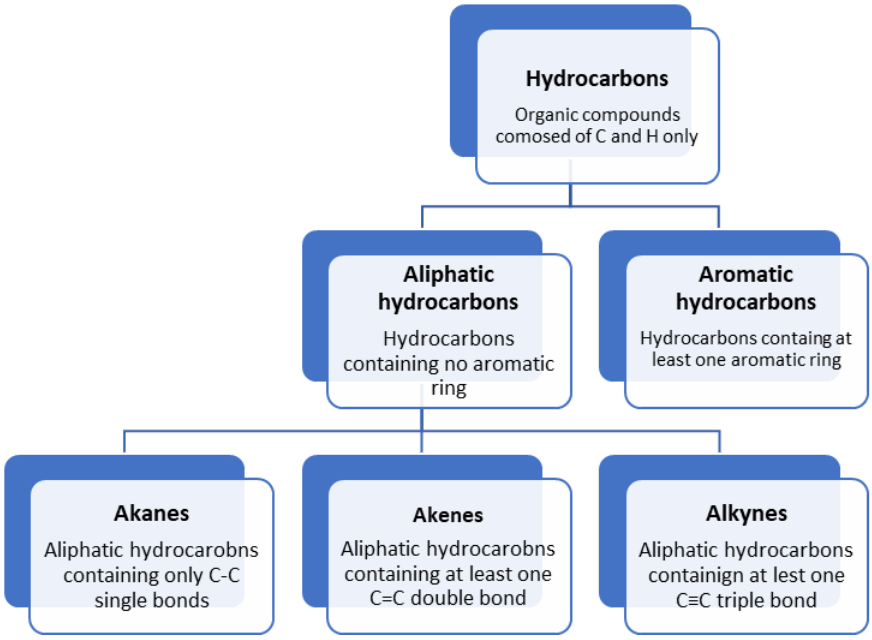


Figure 2.1.1: Classification of hydrocarbons. (Copyright; Public domain)

Alkanes

Nomenclature of alkanes

Organic compounds have trivial names as well as systematic names. Systematic names are based on the International Union of Pure and Applied Chemistry (IUPAC) recommendations, referred to as **IUPAC nomenclature**. Table 1 lists stem names that represent the number of C′s in a continuous chain of C′s (straight chain) or a cyclic chain part of an organic compound.

Table 1: Stem names representing from one to hundred carbons

# of C′s	Stem name	# of C′s	Stem name	# of C′s	Stem name
1	Meth	11	Undec	30	Triacont
2	Eth	12	Dodec	40	Tetracont
3	Prop	13	Tridec	50	Pentacont
4	But	14	Tetradec	60	Hexacont
5	Pent	15	Pentadec	70	Heptacont
6	Hex	16	Hexadec	80	Octacont
7	Hept	17	Heptadec	90	Nonacont
8	Oct	18	Octadec	100	hect
9	Non	19	Nonadec		
10	Dec	20	Icos		

Straight chain alkanes

> To name a straight chain alkane, the first syllable, i.e., alk- is replaced with the stem name that tells the number of C′s in the chain, and the last syllable, i.e., -ane is retained as a suffix in which 'an' tells all bonds are single bonds, and 'e' tells it is a hydrocarbon, i.e., -ane means an alkane.

For example, CH_4 is called methane, CH_3-CH_3 is called ethane, and $CH_3-CH_2-CH_3$ is called propane, where meth-, eth-, and prop- are the stem names representing one, two, and three C′s. Table 2 shows the formulas, IUPAC names of some straight-chain alkanes, and some of their physical properties. The physical properties will be discussed in a later section.

Table 2: Formulas, IUPAC names, melting points, boiling points, densities, and physical states of some straight-chain alkanes.

Formula	IUPAC Name	Melting point (°C)	Boiling point (°C)	Density (g/ml)	Physical state at 20 °C
CH_4	Methane	−182	−162	0.000656	Gas
C_2H_6	Ethane	−183	−89	0.00126	Gas
C_3H_8	Propane	−188	−42	0.00201	Gas
C_4H_{10}	Butane	−138	0	0.00248	Gas
C_5H_{12}	Pentane	−130	36	0.626	Liquid
C_6H_{14}	Hexane	−95	69	0.659	Liquid
C_7H_{16}	Heptane	−91	98	0.684	Liquid
C_8H_{18}	Octane	−57	126	0.703	Liquid
C_9H_{20}	Nonane	−54	151	0.718	Liquid
$C_{10}H_{22}$	Decane	−30	174	0.730	Liquid
$C_{11}H_{24}$	Undecane		196	0.740	Liquid
$C_{12}H_{26}$	Dodecane	−10	216	0.749	Liquid
$C_{13}H_{28}$	Tridecane	−5.4	235	0.756	Liquid
$C_{14}H_{30}$	Tetradecane	5.9	253	0.763	Liquid
$C_{15}H_{32}$	Pentadecane	10	270	0.769	Liquid
$C_{16}H_{34}$	Hexadecane	18	287	0.773	Liquid
$C_{17}H_{36}$	Heptadecane	22	303	0.777	Liquid
$C_{18}H_{38}$	Octadecane	28	317	0.781	Solid
$C_{19}H_{40}$	Nonadecane	32	330	0.785	Solid
$C_{20}H_{42}$	Icosane	37	343	0.789	Solid

Continuous C chain can be straight or wiggle waggle

A regular zigzag line, also called a straight chain, usually represents the hydrocarbons, but it may be a wiggle waggle because of rotation allowed around each C−C single bond. Figure 2.1.2 shows the semi-condensed formulas of decane presented in five different configurations and the corresponding skeletal formulas. The terminal C in a chain is the carbon at the ends connected with only one other C. The internal carbon in a chain is the C connected with at least two other C′s.

- A **continuous C chain** starts from a terminal carbon and reaches the other terminal C through the connected continuous C−C single bonds.
- A **straight-chain hydrocarbon** has all the internal C′s connected with only two other C′s, no matter how much wiggle waggle the chain may be.

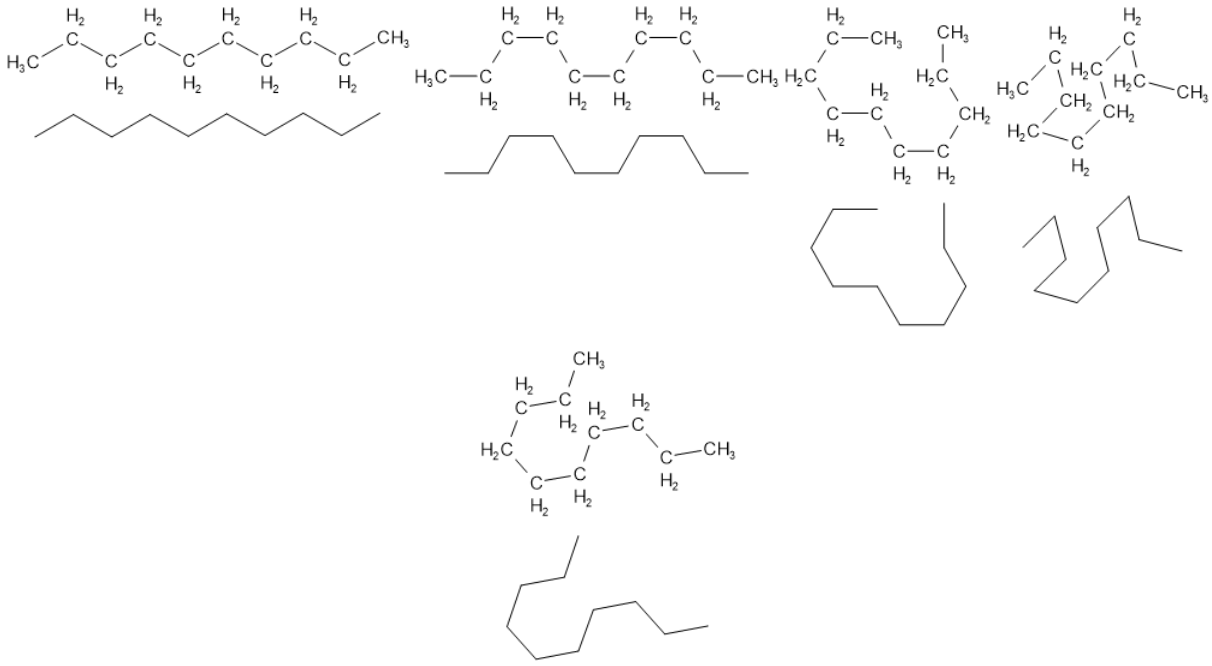
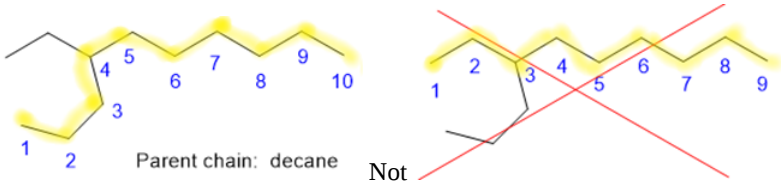


Figure 2.1.2: Five configurations of decane ($C_{10}H_{22}$) caused by rotation around C−C single bonds are drawn: semi-condensed formulas (top) and the corresponding skeletal formulas (bottom). (Copyright; Public domain)

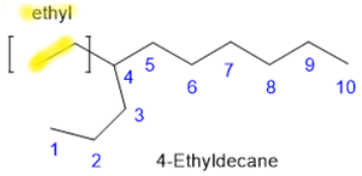
Branched alkanes

If at least one C in an alkane is connected with three or four other C′s, it is a **branched alkane**. The parent chain is the longest continues C chain. The smaller chains branching from the parent chain are called **branches, substituents**, or **alkyl groups**. R- represents a general alkyl group. Rules for naming the branched alkanes are the following.

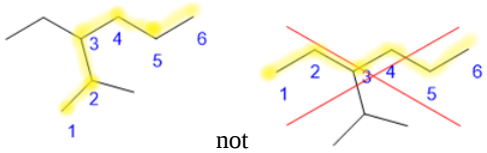
- Find the most extended continuous C chain, called the **parent chain.** Name it by using the stem name that tells the number of C′s in it with the suffix -ane that means it is an alkane.



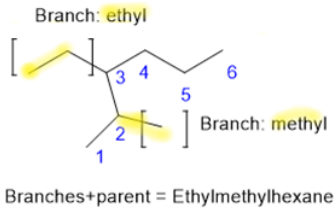
- Name the branch chain by using the stem name that tells the number of C′s in the continues C chain of the branch with the suffix -yl that tells it is a branch. Use the name of the branch as a prefix to the name of the parent chain.



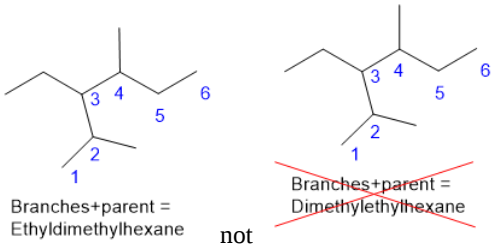
- If there are two or more longest chains of equal length, choose the parent chain with the highest number of branches on it.



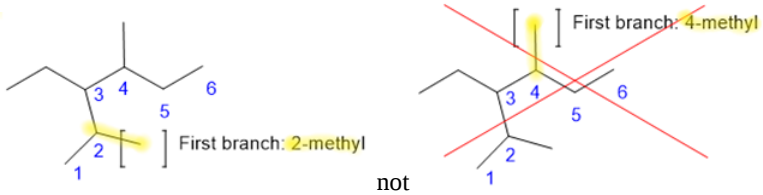
- If there is more than one branch, arrange their names alphabetically as prefixes to the name of the parent chain.



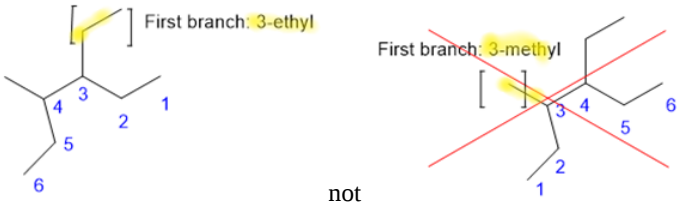
- If there is more than one branch with the same name, list it once with the prefix di-, tri-, tetra-, etc., added to the branch name to represent two, three, four, etc. of the same branches. The alphabetization of the branch names is based on the branch name without the prefixes di-, tri-, tetra-, etc. For example, dimethyl, trimethyl, and tetramethyl are all alphabetized based on the 'm' of methyl, not based on the 'd' or 't' first alphabets of the prefixes.



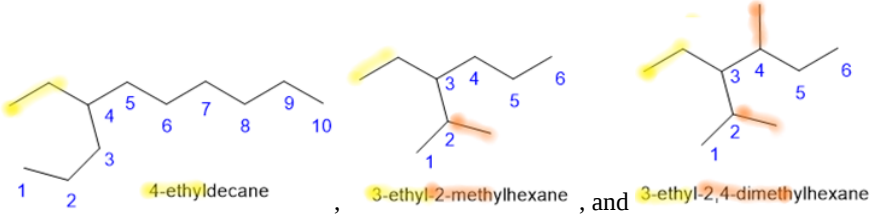
- Number the C′s of the parent chain starting from the end that gives the lowest number to whichever branch appears first on the parent chain.



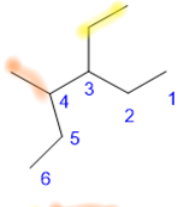
  - If numbering from left or right has the same number for the first branch, start the numbering from the side that gives a lower number to the branch that comes alphabetically first.

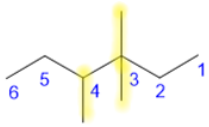


- List the number to which the branch is attached to the parent chain before the branch name, separated by a hyphen. This is the IUPAC name of the alkane.



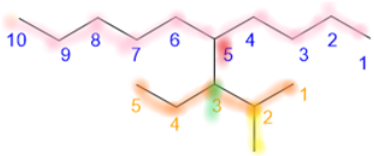
  - If two or more branches have the same name, list their numbers in increasing order, separated by commas, and the final number separated by a hyphen from the branch name.





- If two branches are on the same C, use the same number twice, i.e., once for each branch. 3,3,4-trimethylhexane

- If a branch is further branched, name it as if it is a parent chain following the above rules, except that i) the suffix of the brach is -yl instead of -ane, ii) insert a number from which the beach is attached to the parent chain before -yl, iii) Enclose the branch name in small brackets and place it as a prefix to the name of the parent chain, iv) insert a number before the bracket to which the beach is attached on the parent change, as shown in the following example.

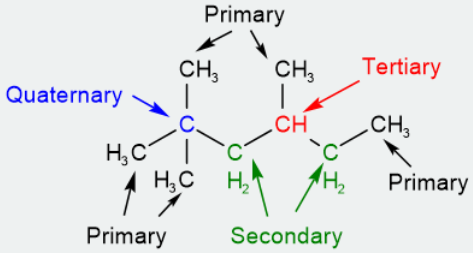


5-(2-methylpentan-3-yl)decane

Primary, secondary, and tertiary C′s and H′s

- The terminal C′s, i.e., the C′s with only one single bond with another C are **primary carbon**.
- The H′s bonded to primary C′s are **primary hydrogen**.
- The C′s with two single bonds with other C′s are **secondary carbon**.
- The H′s bonded to secondary C′s are **secondary hydrogen**.
- The C′s with three single bonds with other C′s are **tertiary carbon**.
- The H′s bonded to tertiary C′s are **tertiary hydrogen**.
- The C′s with four single bonds with other C′s are **quaternary carbon**.

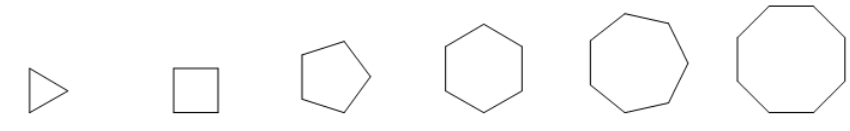
The quaternary C′s do not have any H bonded to them. The following figure labels the primary, secondary, and tertiary carbons and hydrogen with different colors.



Cycloalkanes

An alkane with one ring of C′s in its structure is called cycloalkane. IUPAC rules for naming cycloalkanes are the following.

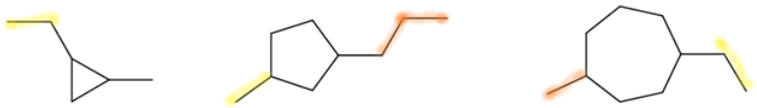
- If there is one ring of C′s without any substituent on it, it is named by using the stem name that tells the number of C′s in the ring with prefix cyclo- and suffix -ane, as shown in the following examples.



- If there is only one branch or a substituent on the ring of C′s, then the ring supplies the parent name, and the branch name is added as a prefix, as for alkanes but without any number preceding the branch name. Some examples of IUPAC names of branched cycloalkanes are shown below.

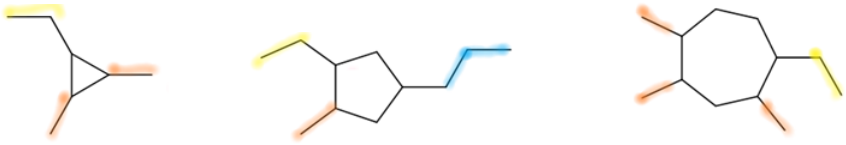
methylcyclopropane, propylcyclopentane, and ethylcycloheptane

- If there are two substitutes, they are listed alphabetically, preceded by location number, as in the case of alkanes. The numbering of C′s in the chain starts from the point of attachment of the substituent that comes alphabetically first. The numbering goes clockwise or counterclockwise, whichever gives the second substitute the lowest number, as shown in the following examples.



1-ethyl-2-methylcyclopropane, 1-methyl-3-propylcyclopentane, and 1-ethyl-4-methylcycloheptane

- If there are more than two substituents, they are listed as prefixes in alphabetic order as for alkanes. The ring C′s attached to one of the substituents is assigned number 1, and the numbering is continued clockwise or counterclockwise so that the addition of numbers assigned to the substituents is the lowest possible. Some examples are shown below.



1-ethyl-2,3-dimethylcyclopropane, 1-ethyl-2-methyl-4-propylcyclopentane, and 1-ethyl-2,4,5-trimethylcycloheptane

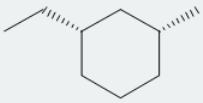
Sometimes the ring name is treated as a branch with the last syllable changed from -ane to -yl, e.g., cyclopropane becomes cyclopropyl as a branch. This change is applied particularly when the chain attached to the ring has more number of C′s than the number of C′s in the ring.

Cis, trans isomerism in cycloalkanes

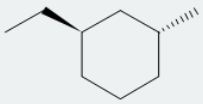
Cycloalkanes with two alkyl groups on different carbons have two configurations: i) both alkyl groups pointing in the same direction are called 'cis' orientation, and ii) the two bulky groups pointing in the opposite direction are called 'trans' orientation, as shown below for the case of 1,2-dimethyl cyclopropane and 1-ethyl-3-methylcyclohexane examples.

cis-1,2-dimethylcyclopropane

trans-1,2-dimethylcyclopropane



cis-1-ethyl-3-methylcyclohexane



trans-1-ethyl-3-methylcyclohexane

The cis- and trans- configuration can not inter-convert by rotation around C−C bond as the rotation is restricted by the carbon bridge in the case of cycloalkanes. Therefore, the cis- and trans-configuration of cycloalkane are isomers of each other, i.e., the same formula but different compounds. For example, cis-1,2-dimethylcyclopropane has a boiling point of 37 C, and its isomer, i.e., trans-1,2-dimethylcylopropane has a boiling point of 28 ºC. This phenomenon, i.e., cis-, trans- isomerism, is common in all cyclic compounds.

Bonding and physical properties of alkanes

Bonding

The general formula of straight chain and branched alkanes is $C_nH_{(2n+2)}$, and that of cycloalkanes having one ring is $C_nH_{(2n)}$ where n is a positive integer. All C's in alkanes are sp^3 hybridized, i.e., tetrahedral geometry with all bond angles about 109.5°, as illustrated in Figure 2.1.3 for the case of ethane molecule. The bond energy for $C-C$ bond is ~3.8 eV and for $C-H$ bond is ~4.4 eV.

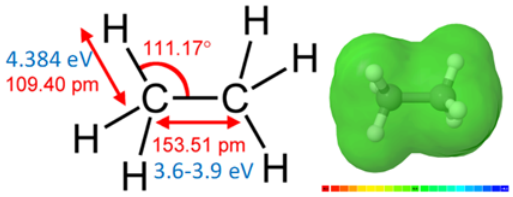


Figure 2.1.3: Illustration of geometry (right) and electrostatic potential map of ethane. (Copyright; the geometry modified from Benjah-bmm27, Public domain via Wikimedia Commons, and electrostatic potential map drawing using https://chemagic.org/molecules/amini.html, Public domain)

Polarity

The electronegative difference for the $C-H$ bond is 0.4, which falls in the nonpolar bond category. The electrostatic potential map of a molecule shows $\delta+$ region in blue, $\delta-$ region in red, and neutral or nonpolar region in green color. Alkanes are nonpolar, as shown in green in the electrostatic map of ethane in Figure 2.1.3. The strong $C-C$ and $C-H$ bonds make the alkanes stable and less reactive. The nonpolar nature of alkanes makes them more inert chemically.

Intermolecular forces and boiling points

Alkanes are nonpolar molecules where the only intermolecular interaction is London dispersion forces. Straight-chain alkanes are a homologous series.

> As the molar mass, and, consequently, the surface area and the number of electrons in the electron cloud, gradually increase, the London dispersion forces increase, resulting in a gradual increase in melting and boiling points, as shown in Table 2.

The first four members for the alkane series, i.e., CH_4 to C_4H_{10} are gases, the next from C_5H_{12} to $C_{17}H_{36}$ are liquid and the larger alkanes are wax or soft solids.

Natural gas is the primary source of methane and ethane. Petroleum, which in the crude oil form is a viscous liquid composed of thousands of organic compounds, primarily hydrocarbons, is the primary source of alkanes and also the main source of organic raw materials. The differences in the boiling points allow the segregation of crude oil into different fractions, as illustrated in Figure 2.1.4.

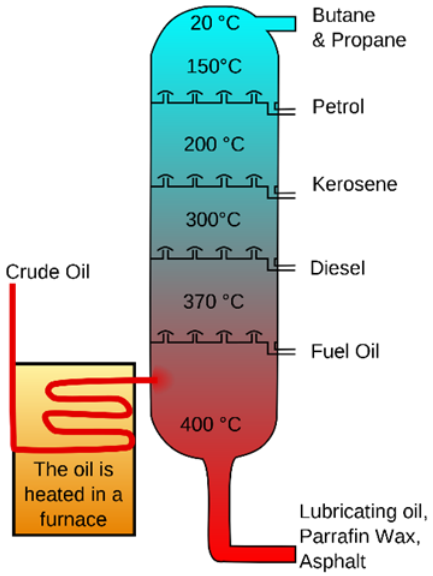


Figure 2.1.4: Illustration of the separation process of components of curd oil based on their boiling points by fractional distillation (Copyright; Crude_Oil_Distillation-fr.svg: Image originale:Psarianos, Theresa knott ; image vectorielle:Rogilbertderivative work: Utain (), CC BY-SA 3.0, via Wikimedia Commons)

Solubility

Alkanes are not soluble in water because alkanes are nonpolar, and water is a polar solvent. Alkanes are soluble in each other or other nonpolar solvents. Alkanes float on water because they are less dense than water. The density of liquid alkanes varies in the range of 0.6 g/ml to 0.8 g/ml, which is less than the density of water, which is ~1 g/ml at room temperature. Petroleum and crude oil are mainly composed of alkanes, which is why oil spills make a layer of oil float on the water's surface.

Changes in physical properties with branching

The physical properties of branched alkanes are almost the same as that of straight chain alkanes, except that:

- the polling point increases as the molar mass increases, and
- for the same molar mass, the boiling point decreases as the branching increases, as shown in Table 1.

Table 1: Examples of hydrocarbons with increasing molar mass and increasing branches and their corresponding boiling points

Number of C′s	Name	Condensed formula	Boiling point
5	Pentane	$CH_3CH_2CH_2CH_2CH_3$	36°C
6	Hexane	$CH_3CH_2CH_2CH_2CH_2CH_3$	69°C
6	2-Methylpentane	$(CH_3)_2CHCH_2CH_2CH_3$	60°C
6	2,2-Dimethylbutane	$(CH_3)_3CCH_2CH_3$	50°C
7	Heptane	$CH_3CH_2CH_2CH_2CH_2CH_2CH_3$	98°C

Straight-chain alkanes are long and cylindrical, as illustrated in Figure 2.1.5. Branching makes them spherical, resulting in less contact area between molecules, fewer London dispersion forces, and lower boiling points. Note that the effect of molar mass is dominant over that of branching.

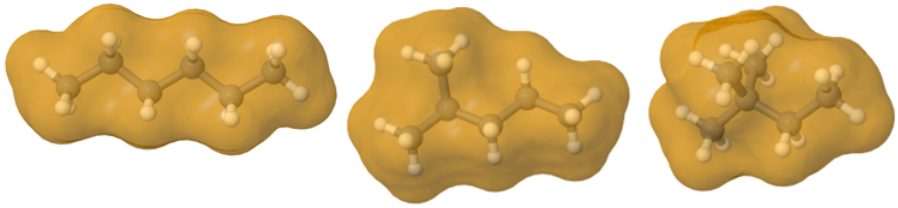


Figure 2.1.5: Comparison of the electron cloud of isomers of hexane, illustrating that shape of the molecule changes from cylindrical to spherical with an increase in branching. The electron cloud of hexane, 2-methylpentane, and 2,2-dimethylbutane, approximately on the same scale using https://chemagic.org/molecules/amini.html (Copyright; Public domain)

Properties of cycloalkanes

The physical properties of cycloalkanes are almost similar to alkanes, except that bonds have to bend to acquire ring-shape that causes angle strain. Further, the molecule can not acquire the most stable configuration due to no or limited rotation around $C-C$ single bond. Therefore, the cycloalkanes, mainly the smaller three or four-member rings, are relatively unstable. Except for cyclopropane, there is a limited rotation possible around $C-C$ bond that allows the molecule to acquire bent-shape, rather than a planer shape that releases some of the strains, as shown by models in Figure 2.1.6. The strain decreases, and the stability increases in this order: cyclopropane > cyclobutane > cyclopentane > cyclohexane < cycloheptane > cyclooctane. Notably, the five and six-membered rings, e.g., cyclopentane and cyclohexane, have little or no strain, so they are the most common cyclic compounds found in nature.

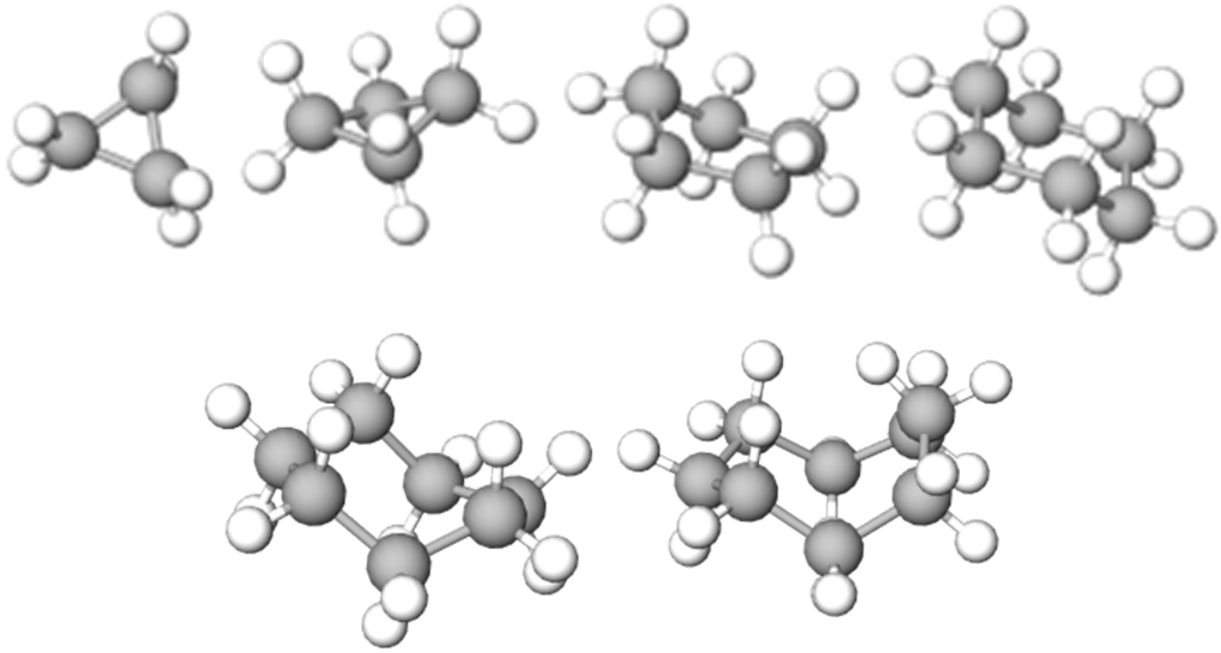


Figure 2.1.6: Models of cycloalkanes, from left top to right bottom: cyclopropane, cyclobutane, cyclopentane, cyclohexane, cycloheptane, and cyclooctane. All of them except cyclopropane are bent-shaped. (Copyright; Public domain)

Some uses of alkanes

The first four members, i.e., methane, ethane, propane, and butane, are gases used as fuels. The next four are components of gasoline, and higher than that are components of kerosene and diesel, used as fuels.

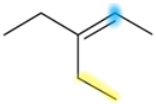
Alkanes and other hydrocarbons are part of almost all organic compounds but are not so common in pure form in living things. Some examples of uses in biological systems are the following. Long-alkanes having eighteen or more C, s are semisolids found in waxes as in vaseline used in ointments and cosmetics. The waxes are found on surfaces of leaves and fruits to protect against water loss, prevent the leaching of essential minerals by rain, and protect against bacteria, fungi, and harmful insects. Alkanes are also found in some pheromones, e.g., sand bees use tricosane ($C_{23}H_{48}$), pentacosane ($C_{25}H_{53}$), and heptacosane ($C_{27}H_{56}$) to identify a mate.

Alkenes

Nomenclature of alkenes

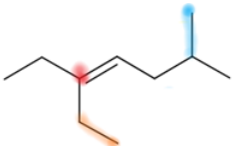
The procedure for IUPAC naming alkenes is the same as alkanes, except for the following changes.

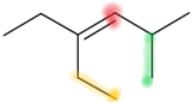
- The suffix to the stem name of the parent chain is -ene, where 'en' tells its an alkene, and 'e' means it is a hydrocarbon, e.g., $CH_2{=}CH_2$ is ethene.
- A location number of the first C of C=C bond is inserted between the stem- name and -ene suffix, separated by hyphens, as in the following example, except for the propene that does not need to show the location number.



3-ethylpent-2-ene

- The parent chain numbering starts from the end that gives the lowest number to the first C of C=C bond.
  - If the first C of C=C bond receives the same number from either side, then the rule for alkane applies, i.e., start numbering from the side that gives the lower number to the branch that appears first, as shown in the following examples.

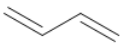




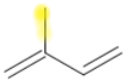
4-ethyl-2-methylhex-3-ene, ~~3-ethyl-5-methylhept-3-ene~~

3-ethyl-6-methylhept-3-ene, ~~5-ethyl-2-methylhept-4-ene~~

- If two or more double bonds exist, the suffix -ene is changed to -diene, triene-, etc. The rest of the nomenclature remains the same.

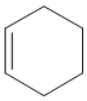


buta-1,3-diene

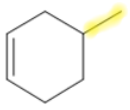


2-methylbuta-1,3-diene

- In cycloalkenes, the C′s of C=C are #1 and #2 and do not need to show the location number.
  - If there is a branch on cycloalkene, the C′ of C=C bond are #1 and #2, and the numbering is continued clockwise or counterclockwise to give the lowest number to the branch, as shown in the following examples.

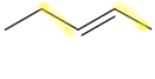


cyclohexene

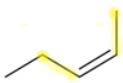


4-methylcyclohex-1-ene

- there are two alkyl groups, one on each C of C=C bond (or third or fourth group different from the other group on the same C) then a descriptor E indicating cis, or Z indicating the trans orientation of the bulkier groups, preceded by the location number of the first C of the C=C bond, is added as a prefix enclosed within brackets.



(2E)-pent-2-ene



(2Z)-pent-2-ene

- if the double bond is in a branch, the -ene suffix is replaced with -enyl, where 'en' tells it is an alkene and 'yl' means it is a branch.

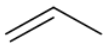
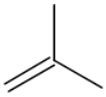
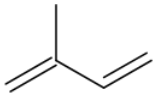
Caution

In the older system:

- The parent chain may not be the longest, but the parent chain is the longest chain counting the double bond, and
- The location number of the double bond is inserted before the stem name, e.g., but-2-ene ($CH_3CH{=}CHCH_3$) is written as 2-butene in the old IUPAC nomenclature.

Some common names of alkenes

Some common names are still being used in the chemical literature and industries. For example, ethylene, propylene, isobutylene, and isoprene are often used in place of IUPAC names ethene, propene, 2-methylpropene, and 2-methylbuta-1,3-diene.

Structure and IUPAC name	 ethene	 prop-1-ene	 2-methylprop-1-ene	 2-methylbuta-1,3-diene
Common name	ethylene	propylene	isobutylene	isoprene

Bonding and physical properties of alkenes

Alkenes are nonpolar compounds with only London dispersion forces as intermolecular forces, like alkanes. Therefore, the physical properties of alkenes are similar to alkanes with the same C skeleton, i.e., they are insoluble in water, soluble in nonpolar solvents, and have densities lower than that of water. The significant difference between alkanes and alkenes is the C=C bond in alkenes described below.

bonding

The general formula of alkenes with one double bond is $C_nH_{(2n)}$, where n is a positive integer. The C′s in C=C bond are sp^2 hybridized with trigonal planer geometry around each of the sp^2 C′s with bond angles around 120°, as illustrated in Figure 2.1.7. Except for C=C, all other C′s in alkenes are sp^3 hybridized, i.e., tetrahedral geometry with all bond angles about 109.5°. Although alkenes are nonpolar overall, the Electrostatic potential map in Figure 2.1.7 shows a red region where π-electrons are located above and below the σ bond and slightly blush region where H′s of sp2 C′s are located. These $\delta-$ and $\delta+$ regions impart reactivity to alkenes that will be described in later sections.

Degree of unsaturation or index of hydrogen deficiency (IHD)

The general formula of an alkane is $C_nH_{(2n+2)}$, where n is a positive integer. One degree of unsaturation is two less H′s than in an alkane of the same carbon structure. For example, one ring in cycloalkane or one double bond in alkenes reduces two H′s, one unsaturation degree. The **degree of unsaturation** is also known as the **index of hydrogen deficiency (IHD)**. Remember, if a halogen with one bond is added to an alkane, it replaces one H, an O with two bonds does not change the number of H′s, and a N with three bonds add one H. Based on this information, the degree of unsaturation, i.e., number of double bonds + rings, can be calculated by the following formula.

$$IHD = (C's + 1) - \frac{H's - N's + X's}{2}$$

, where IHD is an index of hydrogen deficiency, which is equivalent to the number of π bonds + rings, C′s is the number of C atoms, H′s is the number of H atoms, N′s is the number of N atoms, and X′s is number of halogen atoms in the molecular formula.

If oxygen is in the molecule, it is not included in the IHD calculation.

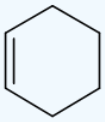
Example 2.1.1

What is the degree of saturation in cyclohexane (C_6H_{10}?

Solution

C′s = 6, H′s = 10, N′s = 0, X′s = 0.

$IHD = (C's + 1) - \frac{H's - N's + X's}{2} = (6 + 1) - \frac{10 - 0 + 0}{2}$ = 2, which is one π-bond and one ring as shown below.



Substituents attached to sp^2 hybridized C′s of alkenes are also called vinylic. The bond energy for C=C bond is ~7.4 eV which is less than two times of ~3.8 eV of C−C bond, and the vinylic =C−H bond is ~4.8 eV, which is stronger than alkane −C−H~4.4 eV. Comparison ~7.4 eV for C=C bond vs 3.8 eV for C−C bond shows that, although a double bond ((σ + π) is stronger than a single bond, a π bond is weaker than a σ bond.

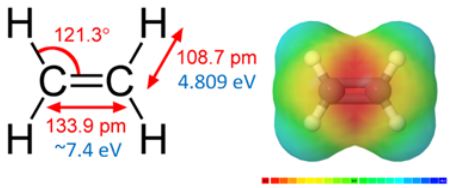
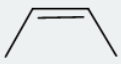


Figure 2.1.7: Illustration of geometry (right) and electrostatic potential map of ethene. (Copyright; the geometry modified from Benjah-bmm27, Public domain via Wikimedia Commons, and electrostatic potential map drawing using https://chemagic.org/molecules/amini.html, Public domain)

Cis, trans isomerism in alkenes

If there are two alkyl groups, one on each C of a double bond, they can be pointing in the same direction, called cis orientation, or in the opposite direction, called trans orientation, as shown in the Figure below for the case of but-2-ne. The cis and trans alkenes of the same molecular formula are different compounds related as isomers of each other. This is because three is no rotation around C=C bond without breaking and re-making the π bond.



cis-but-2-ene, IUPAC name: (2Z)-but-2-3n3

trans-but-2-ene, IUPAC name: (2E)-but-2-3n3

Cis trans isomerism plays a vital role in biological systems. For example, fatty acids are significant components of fats and vegetable oils. Fatty acid contains a long alkyl chain attached to a carboxylic acid group. The alkyl chain is usually alkane in the cases of animal fats or includes one or more double bonds, almost always in the cis configuration in the cases of vegetable oils and fish oil. For example, models of steric acid with no double bonds, oleic acid with one cis double bond, and linoleic acid with two cis double bonds are shown below.

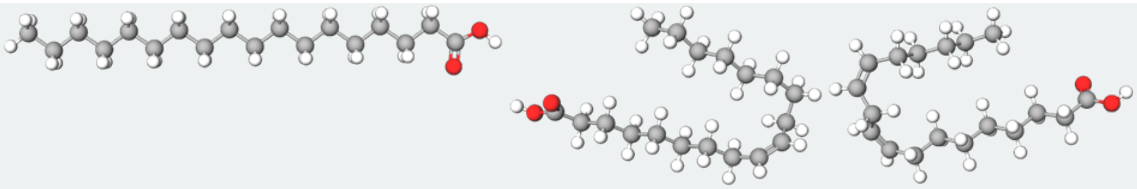
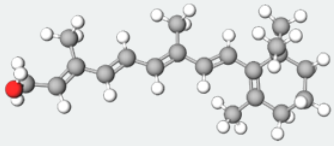


Figure 2.1.8: Model of three fatty acids contains 18 C′s each: steric acid with no double bond, linoleic acid with one cis double bond, and oleic acid with two cis double bonds. (Copyright; Public domain)

The cis-double bonds change the long cylindrical shape of fatty acids to a bent or more spherical shape with less contact area with the neighboring molecules. It results in lower intermolecular forces and melting points than the same chain with a trans or no double bond. That is why vegetable oils with more cis-double bonds are liquids, while animal fats with little to no double bonds are solids at room temperature. Another example is vitamin A having five trans double bonds, as shown in the model drawing below. Vitamin A leads to retinol synthesis, which plays a crucial role in our vision based on switching one of its double bonds between cis and trans configurations.



Model of vitamin A molecule with five trans double bonds.

Polarity and chemical reactivity

Although C=C bond is nonpolar, the electrostatic potential map of ethene in Figure 2.1.7 shows there is an electron-rich region (red color in the map) above and below the axis joining the nuclei. This is because the σ bond is along the axis of nuclei, and the π bond places electrons above and below the σ bond. Molecules with negative charge or electron-rich (δ-) regions, as in the case of C=C bond in alkene, are **nucleophiles**. Molecules with positive charge or electron-deficient (δ+) region are electrophiles. Chemical reactions between nucleophiles and electrophiles are facilitated by the attraction between the opposite charges that allow them to come closer to each other for bond-making and bond-breaking to happen.

Electrophile/nucleophile or acid/base?

The term acid is used for electrophile when δ+ part is a proton, and the base is used in the place of nucleophile when δ- region is a lone pair of electrons interacting with a proton.

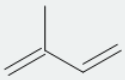
Alkenes uses and importance in biological systems

Alkenes are among the raw materials for the synthesis of several organic chemicals. Mainly, ethene and propene are the raw material for synthesizing commercially important polymers: polyethylene and polypropylene.

Ethene is a ripening agent for fruits. It allows fruit growers to pick fruits while they are green and less susceptible to bruising and then treat them with ethene gas for ripening when ready for sale. Lycopene, -a red pigment in tomatoes, and carotene, -a yellow pigment in carrots, are polyenes, shown below.

Terpenes -a class of natural products built from an alkene

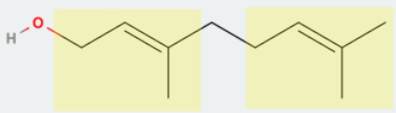
Terpenes are a class of natural products having a general formula $(C_5H_8)n$, where n is a positive integer. Isoprene ((C_5H_8), i.e., 2-methylbuta-1,3-diene is the building block of terpene.



isoprene, IUPAC name: 2-methylbuta-1,3-diene

Terpenoids are derivatives of terpenes in which heteroatoms/functional group is added. Natural rubber, essential oils, and steroids are examples of terpenes or terpenoids. They play important roles in living things, e.g., in defense against diseases.

Some examples of terpenes and terpenoids are shown below, with the isoprene skeleton highlighted.



Geraniol

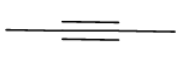
Limonene Camphor

Alkynes

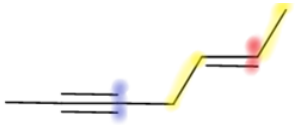
Nomenclature

The IUPAC nomenclature of alkynes followes the same rule as alkenes, except for the following changes.

- The suffix to the stem name is -yne where 'yn' tells it is an alkyne, and 'e' means it is a hydrocarbon, e.g., $\mathrm{CH \equiv CH}$ is ethyne.
- If a double and a triple bond are in the same molecule, they are treated equally. One exception is when the double receives the same number as the triple bond counted from the other end of the parent chain; preference is given to the double bond, as shown in the following examples.



but-2-yne



(2E)-hept-2-en-5-yne

Note: in the second example, the double bond received a lower number than the triple bond. The common name of ethyne is acetylene, which is also accepted as its IUPAC name.

Bonding and physical properties of alkynes

Alkynes are nonpolar compounds with only London dispersion forces as intermolecular forces, like alkenes and alkanes. Therefore, the physical properties of alkynes are similar to alkanes with the same C skeleton. They are insoluble in water, soluble in nonpolar solvents, and have lower densities than water's. The significant difference is the presence of a triple bond in alkynes described below.

Bonding

The general formula of alkyne with one triple bond is $C_nH_{(2n-2)}$, where n is a positive integer. The C′s in $\mathrm{C \equiv C}$ bond are sp hybridized with a linear geometry around each of the sp C′s with bond angles 180º, as illustrated in Figure 2.1.8. Substituents attached to sp hybridized C′s of alkynes are also called **acetylenic**. The bond energy for $\mathrm{C \equiv C}$ bond is ~10 eV which is about 2.5 times stronger than that of $\mathrm{C-C}$ bond. Due to the cylindrical shape of two π-bond clouds, there is free rotation possible around a $\mathrm{C \equiv C}$ bond, but due to the linear geometry, it means less.

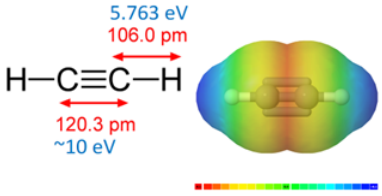


Figure 2.1.8: Illustration of geometry (right) and electrostatic potential map of ethyne. (Copyright; the geometry modified from Benjah-bmm27, Public domain via Wikimedia Commons, and electrostatic potential map drawing using https://chemagic.org/molecules/amini.html, Public domain)

polarity and chemical reactivity

Although $\mathrm{C \equiv C}$ bond is nonpolar, the electrostatic potential map of ethyne in Figure 2.1.8 shows there is an electron-rich region (red color in the map) around the axis joining the nuclei. This is due to π-bonding electrons around the σ bond. The π bonds

provide δ- region in alkynes, making them nucleophiles, like alkenes. The H′s are also blue in the electrostatic potential map of ethyne, showing that the C−H bond is polar with $\delta+$ charge on H.

Dependence of electronegativity of C on hybridization

The C−H bond is nonpolar with an electronegativity difference of 0.4 between C and H. The green color in the electrostatic potential map of ethane reflects it. However, the C−H bond becomes shorter, stronger, and polar with a change in the hybridization of the C from sp^3 to sp^2 and sp. It means the electronegativity of C increases with a change in the hybridization from sp^3 to sp^2 and sp. The reason is that the sp orbital is wider and shorter due to more s-orbital character (50% s-character) than sp^2 orbital (33% s-character), which, in turn is wider and shorter the sp^3 orbital (25% s-character), as illustrated in Fig. 7. The wide and shorter orbitals place the electron closer to the C nucleus for a stronger attraction to the nucleus, which means higher electronegativity.

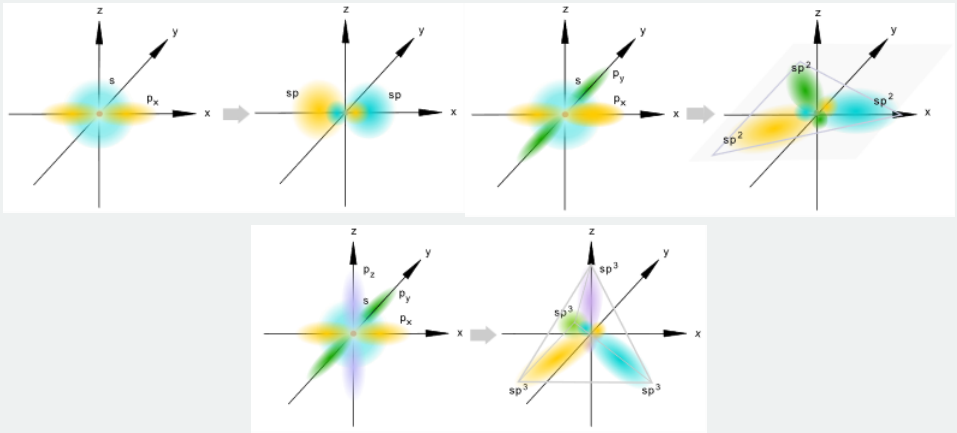


Figure 2.1.1: Hybridization of one s and one p orbitals to produce two sp-orbitals wider and shorter than hybridization of one s and two p orbitals to produce three sp^2 orbitals and one s and three p orbitals to produce four sp^3 orbitals. (Copyright: Joanna Kośmider, Public domain, via Wikimedia Commons)

Uses of alkynes

Alkynes are less common in biological systems or the chemical industry than alkenes and alkanes. The first member, i.e., acetylene, is used in oxyacetylene torches for cutting and welding metals, as illustrated in the figure on the right. Acetylene is also a starting material for several chemicals used in the chemical industry.

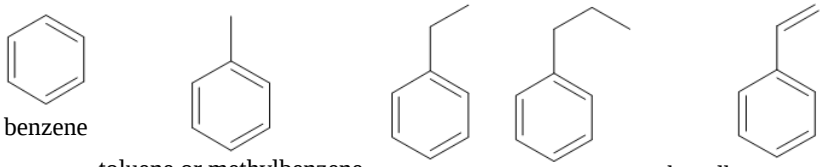
Aromatic hydrocarbons

Aromatic hydrocarbons contain a planer cycle or ring (not bent or puckered) in their structure with all atoms sp^2-hybridized and π-electrons in the ring $= 4n + 2$, where n is a positive integer. Benzene is an example of an aromatic ring that will be described here. Although aromatic hydrocarbons have systematic names like other hydrocarbons, some alternate and familiar names are so widely used that they are accepted as equivalent to IUPAC names.

Nomenclature of benzene derivatives

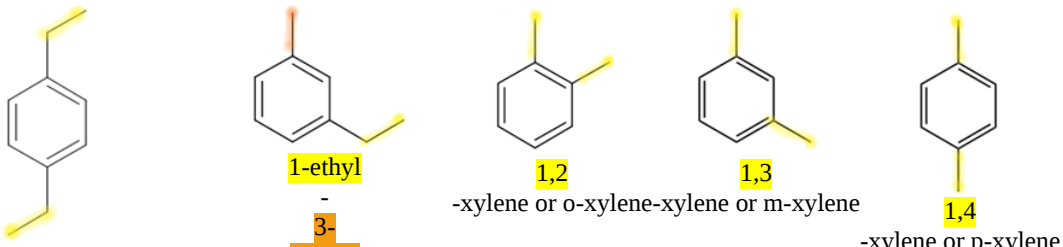
- The name of the six sp^2 hybridized C-chain is benzene. It is used as the parent name for the substituted benzene rings.
- If there is only one substituent, i.e., mono-substituted benzene, use the branch name as a prefix to the parent name 'benzene' without numbering.

- If the substituent is methyl ($-CH_3$), the parent name 'toluene' is often used for it, and if the substituent is ethenyl ($-CH=CH_2$), it is also called styrene. A few examples are given below.



benzene toluene or methylbenzene ethylbenzene propylbenzene ethenylbenzne or styrene

- If there are two substituents (di-substituted benzene), start numbering the chain from the first substituent and go clockwise or counterclockwise so that the second substituent receives the lower number. List the substitutes alphabetically, preceded by their number, as prefixes to the parent name 'benzene.'
  - If one of the two substitutes is methyl ($-CH_3$), start numbering the chain from the point of attachment of methyl and go clockwise or counterclockwise so that the second substitute receives the lower number. List the second substituent preceded by its number as a prefix to the parent name 'toluene.'
  - If there are two methyls ($-CH_3$) substituents, the parent name 'xylene' is often used. Numbering starts from one of the methyls and goes clockwise or counterclockwise, so the second methyl receives the lower number. List the numbers as prefixes to the parent name 'xylene,' numbers separated by commas from each other and by a hyphen from the parent name, as usual.
  - Alternate to the numbers, the word 'ortho' or 'o' for a 1,2 relationship, 'meta' or 'm' for a 1,3 relationship, and 'para' or 'p' for a 1,3 relationship of two substituents on a benzene ring are also accepted. A few examples are the following.



1,4-diethyl benzene or p-diethylbenzene

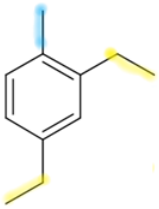
1-ethyl-3-methyl benzne or 3-ethyltoluene

1,2-xylene or o-xylene

1,3-xylene or m-xylene

1,4-xylene or p-xylene

- If there are more than two substituents (poly-substituted benzene), start numbering the chain from the first substituent and go clockwise or counterclockwise so that the substituents receive an overall lower number. List the substitutes alphabetically, preceded by their number, as prefixes to the parent name 'benzene.'
  - If one substituent is methyl ($-CH_3$), start numbering from the point of attachment of methyl ($-CH_3$) and go as usual to assign the lower overall number to the rest of the substituents, and use the parent name 'toluene.'



2,4-diethyl-1-methyl benzne or 2,4-diethyltoluene

Bonding and physical properties of aromatic hydrocarbons

Aromatic hydrocarbons are nonpolar compounds with only London dispersion forces as intermolecular forces, like other hydrocarbons. Benzene has low solubility in water (~1.8 g/L at room temperature), melts at 5.53 ^{o}C, and boils at 80.1 ^{o}C, which are

comparable to that of cyclohexane (immiscible in water, 6.5 ^{o}C, and boiling point is 80.7 ^{o}C). The primary difference is resonance resulting in significant stabilization of benzene compared to any other hydrocarbon, as described below.

Bonding

All C′s in the benzene ring are sp^2 hybridized, i.e., trigonal planer geometry with all bond angles about 120^{o}, as illustrated in Figure 2.1.10. All C′s in benzene ring and H′s or other substituents attached to them are in one plane. Substituents attached to the C′s of the benzene ring are also called **phenylic**. The bond energy for C$\overset{\cdots}{-}$C bond in benzene ring is ~5.4 eV which is between a C−C and a C=C bond. the phenylic C−H bond is ~4.9 eV, almost the same as in alkenes. The C$\overset{\cdots}{-}$C bond in the benzene ring is (140 pm), i.e., about in the middle of a single and a double bond. This is explained by benzene having two equal resonance contributors in which a single and double bond alternates and the resonance hybrid has a 50% π-bond.

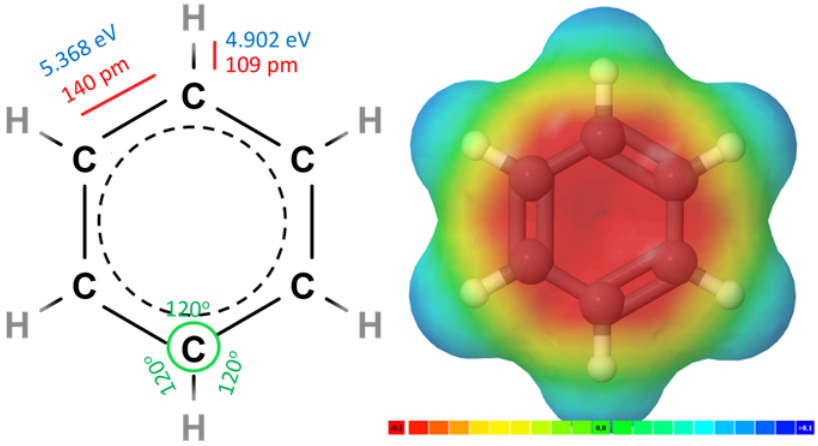


Figure 2.1.10: Illustration of geometry (right) and electrostatic potential map of benzene. (Copyright; Public domain)

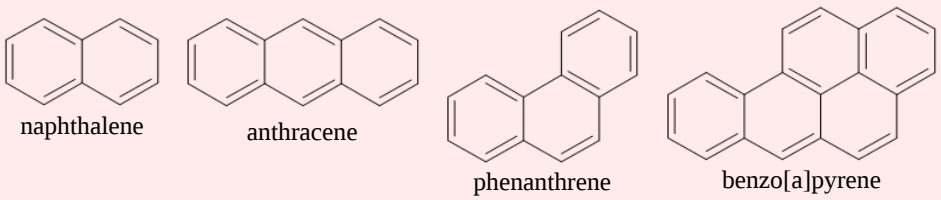
polarity and chemical reactivity

Although benzene is nonpolar, the electrostatic potential map of benzene in Figure 2.1.10 shows there is a circle of an electron-rich region (red color in the map) above and below the ring. This is due to the delocalized π-bond extending over all C′s of the benzene ring. The resonance in a cycle of aromatic rings imparts significantly more resonance stabilization, called **aromatic stabilization,** than in any other hydrocarbon.

> Although the δ^- region of delocalized elections makes benzene a nucleophile, its nucleophilic character is weak because donating these electrons causes a loss of aromatic stabilization. Therefore benzene needs stronger electrophiles to react with than an alkene and follows a different reaction route than an alkene does.

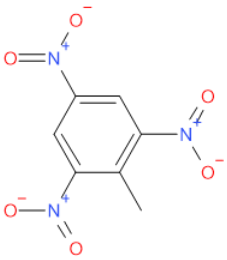
Cancer risk from polycyclic aromatic hydrocarbons (PAHs)

Benzene increases the risk of leukemia and other blood disorders. Polycyclic **aromatic hydrocarbons** (PAHs) contain two or more benzene rings fused edge to edge. Naphthalene has two benzene rings fused. Naphthalene is used in mothballs. Anthracene has there benzene rings linked together. Anthracene is used in the manufacture of dyes. Phenanthrene is also three benzene rings fused. Phenanthrene is known to cause cancer, i.e., it is a **carcinogen**. Benzo[a]pyrene has five benzene rings connected. Benzo[a]pyrene is a potent carcinogen found in tobacco smoke, barbecued meat, and automobile exhaust.

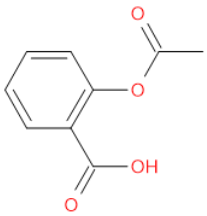


Uses of aromatic hydrocarbons

Benzene is used as a raw material for synthesizing several chemicals. Some familiar aromatic ring-containing compounds include explosives, e.g., trinitrotoluene (TNT), common medicines, e.g., aspirin, acetaminophen, ibuprofen, sulfanilamide, and flavoring agents, e.g., vanillin, shown below.

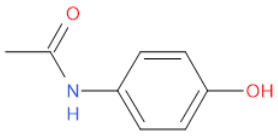


2,4,6-trinitrotoluene, TNT



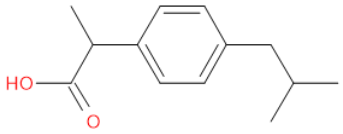
2-acetyloxybenzoic acid

(Aspirin)



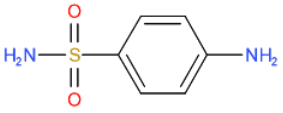
N-(4-hydroxyphenyl)acetamide

(Acetaminophene)



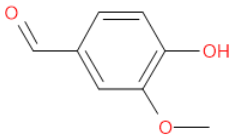
2-[4-(2-methylpropyl)phenyl]propanoic acid

(Ibuprofen)



4-aminobenzenesulfonamide

(Sulfanilamide)



4-hydroxy-3-methoxybenzaldehyde

(Vanilin)

2.2: Functional group

Learning Objectives

- Recognize the two basic mechanisms of bond breaking and making and draw curly arrows to show the electron movement in these steps.
- Define organic functional group and recognize it in an organic compound.

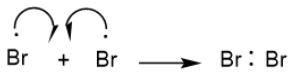
What is a chemical reaction?

A chemical reaction involves making and breaking chemical bonds. Often the elemental composition of the compound changes during a chemical reaction. Sometimes the composition remains the same; atoms rearrange, resulting in a new combination that is an isomer of the initial compound. Often making and breaking of a covalent bond is involved in organic reactions. Recall that a covalent bond is a shared pair of electrons. The bond breaking or making can happen in one of two modes: homolytic or heterolytic, as described below.

Homolytic breaking and bond-making

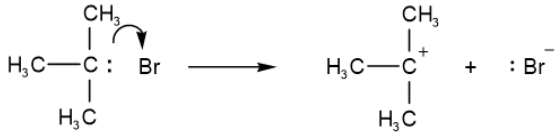
In homolytic bond breaking, each bonded atom receives one of the two electrons in a covalent bond, e.g.:

, where half-headed arrows show the movement of a single electron. The above reaction produces two atoms with one unpaired electron, i.e., free radicals. The reverse of it is hemolytic bond making, e.g.:

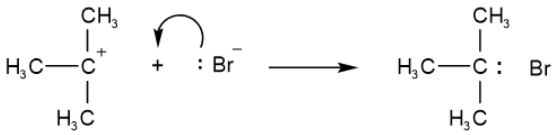


Heterolytic bond-making and breaking

In heterolytic bond breaking, the bonding electrons move to one of the bonded atoms, usually to a more electronegative atom, making a lone pair, e.g.:



, where the regular double-headed arrows show the movement of an electron pair. The formal charge changes in heterolytic bond-making and breaking. The bonded atom that receives the electrons usually becomes an anion, and the other becomes a cation in the above reaction. Revers of it is heterolytic bond making, e.g.:



What is a functional group?

> A functional group is an atom or a group of atoms that imparts a characteristic set of physical and chemical properties to the compound.

For example, alkanes with no functional group are the least reactive classes of organic compounds. Alkanes have only $\mathrm{C-C}$ and $\mathrm{C-H}$ single bonds, which are nonpolar and strong bonds. Alkenes, alkynes, and aromatic compounds are relatively reactive classes of organic compounds because they have a $\mathrm{C=C}$ bond, $\mathrm{C \equiv C}$ bond, or a benzene ring as functional groups. Although alkenes, alkynes, and aromatic hydrocarbons are nonpolar, there is a partial negative (δ^-) character in the π-bond region that attracts electrophilic reagents having δ^+ regions. Further, the π-bond is weaker than a σ-bond making it easier to break during chemical reactions.

Classification of organic compounds based on the functional groups

One primary class of organic compounds is hydrocarbons that contain only C′s and H′s, listed in Table 1, and the others are organic compounds with heteroatom/s, in their functional group/s. Hydrocarbons have been introduced in the previous section. Other classes of organic compounds containing heteratom\s in their functional groups are described next.

The heteroatom/s in the functional group can be a single bonded, i.e., sp^3-hybridized halogen, O, N, S, P, etc., listed in Table 2; a double bonded, i.e., sp^2-hybridized O, N, S, P, etc., listed in Table 3; or a combination of these, listed in Table 4.

Table 1: Classes of hydrocarbons based on their functional groups. (R-, R_1-, R_2, R_3, and R_4- represents a general alkyl group or hydrogen)

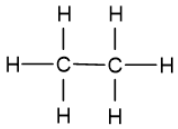
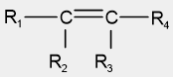
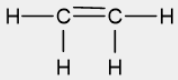
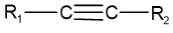
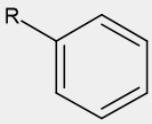
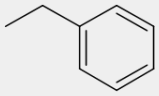
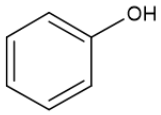
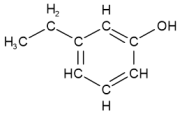
Class name	Group name	General structural formula	General condensed formula	Prefix	Suffix	Example
Alkane	Alkyl	R	R	alkyl-	-ane	 ethane
Alkene	Alkenyl		$R_2C{=}CR_2$	alkenyl-	-ene	 ethene
Alkyne	Alkynyl		$RC \equiv CR_2$	alkynyl-	-yne	 ethyne
Aromatic (benzene derivatives)	Phenyl		$R{-}C_6H_5$	phenyl-	-benzene	 ethylbenzene

Table 2: Classes of organic compounds containing an sp^3-hybridized heteroatom in functional groups. (R-, R_1-, R_2, R_3, and R_4- represent a general alkyl group or hydrogen.)

Class name	Group name	General structural formula	General condensed formula	Prefix	Suffix	Example
Haloalkane	Halo	$R{-}X$, i.e., $R{-}F$, $R{-}Cl$, $R{-}Br$, or $R{-}I$	$R{-}X$ i.e., $R{-}F$, $R{-}Cl$, $R{-}Br$, or $R{-}I$	halo- i.e., fluoro-, chloro-, bromo-, or iodo-)	-	$CH_3{-}CH_2{-}Cl$ (chloroethane)
Alcohol	Alcohol	$R{-}OH$	ROH	hydroxy-	-ol	$CH_3{-}CH_2{-}OH$ (ethanol)
Phenol	Phenol		C_6H_5OH	-	-phenol	 3-ethylphenol
Ether	Ether	$R{-}O{-}R'$	ROR'	Alkoxy	-	$CH_3{-}CH_2{-}O{-}CH_2$ (ethoxyethane)

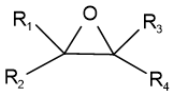
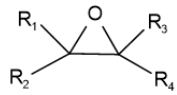
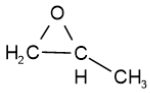
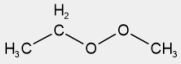
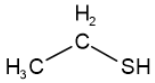
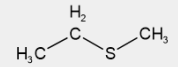
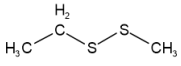
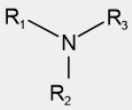
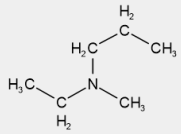
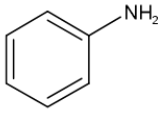
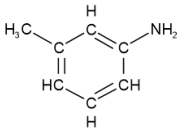
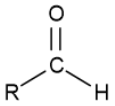
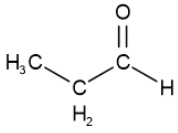
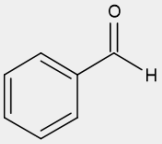
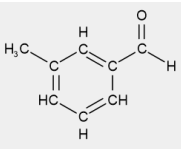
Class name	Group name	General structural formula	General condensed formula	Prefix	Suffix	Example
Epoxide	Epoxide			epoxy-	-oxirane	 2-methyloxirane, or 1,2-epoxypropane
Peroxide	Peroxy	$R-O-O-R'$	$ROOR'$	peroxy-	alkyl peroxide	 (methylperoxy)ethane
Thiol	Sulfhydryl	$R-S-H$	RSH	sulfanyl-	-thiol	 ethanethiol
Sulfide (Thioether)	Sulfide	$R-S-R'$	RSR'	sulfanyl-	-sulfide	 (methylsulfanyl)ethane
Disulfide	Disulfide	$R-S-S-R'$	$RSSR'$	disulfanyl-	-disulfide	 (methyldisulfanyl)ethane
Amine	Amine		$R_1R_2R_3N$	amino-	-amine	 ethyl(methyl)propylamine
Aniline	Aniline		$C_6H_5NH_2$	-	-aniline	 3-methylaniline

Table 3: Classes of organic compounds containing an sp^2-hybridized heteroatom in functional groups. (R-, R_1-, R_2, R_3, and R_4- represents a general alkyl group or hydrogen)

Class name	Group name	General structural formula	General condensed formula	Prefix	Suffix	Example
Aldehyde	Aldehyde		$RCHO$	oxo-	-al	 propanal
Benzaldehyde	Benzaldehyde		C_6H_5COH	-	-benzaldehyde	 3-methylbenzaldehyde

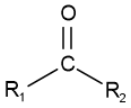
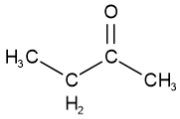
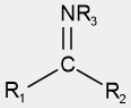
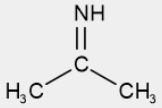
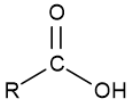
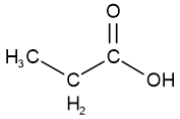
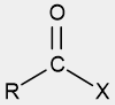
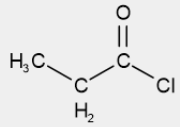
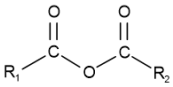
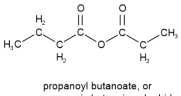
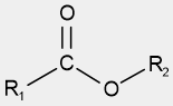
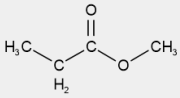
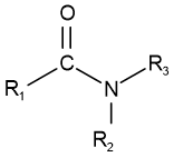
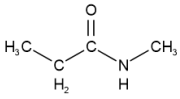
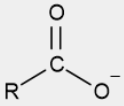
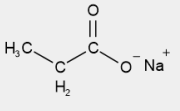
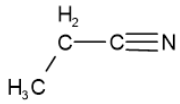
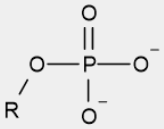
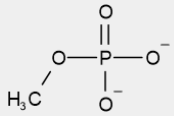
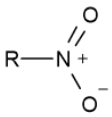
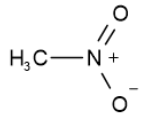
Class name	Group name	General structural formula	General condensed formula	Prefix	Suffix	Example
Ketone	Ketone		$RCOR'$	oxo-	-one	 butan-2-one
Imine	Imine		$R_1C(=NR_3)R_2$	imino-	-imine	 propan-2-imine

Table 4: Classes of organic compounds containing a mix of sp^3-, sp^2-, and sp-hybridized heteroatoms in functional groups. (R-, R_1-, R_2, R_3, and R_4-represents a general alkyl group or hydrogen)

Class name	Group name	General structural formula	General condensed formula	Prefix	Suffix	Example
Carboxylic acid	Carboxyl		$R-COOH$	carboxy-	oic acid	 propanoic acid
Acyl halide	Acyl halide		$R-COX$	-	oil halide	 propanoyl chloride
Acid anhydride	Acid anhydride		$R_1-(CO)O(CO)R_2$		R1-oil R2-oat, or R1-oic R2-oic anhydride	 propanoyl butanoate, or propanoic butanoic anhydride
Easter	Easter		R_1COORR'	-	alkyl -alkanoate	 methyl propanoate
Amide	Amide		$RCONR'R''$	alkyl carbamoyl	-amide	 N-methylpropanamide
Carboxylate	Carboxylate		$RCOO^-$	-carboxy	-oate	 sodium propanoate

Class name	Group name	General structural formula	General condensed formula	Prefix	Suffix	Example
Nitrile	Nitrile	$\mathrm{R{-}C{\equiv}N}$	RCN	cyano-	-nitrile	 propanenitrile
Phosphate	Phosphate	$\mathrm{ROP(=O)(OH)_2}$		-	phosphate	 methyl phosphate
Nitro	Nitro	$\mathrm{RNO_2}$		nitro-	-	 nitromethane

2.3: Functional groups containing sp3-hybridized heteroatom

Learning Objectives

- Recognize and assign IUPAC name to functional groups containing sp^3-hybridized heteroatoms, including haloalkanes, alcohols, ethers, and amines.
- Predict the polarity of the bonds in the functional groups of haloalkanes, alcohols, ethers, and amines and recognize the creative sites based on the partial charges in the functional group regions.
- Identify primary, secondary, tertiary, and quaternary designation of haloalkanes, alcohols, ethers, and amines.

Halogenated hydrocarbons

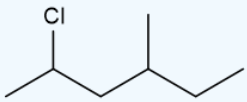
When a H in a hydrocarbon is replaced an X, where X is a halogen, i.e., fluorine (F), chlorine (Cl), bromine (Br), or iodine (I), it is a halogenated hydrocarbon. A halogenated alkane is called **haloalkane**.

Nomenclature of halogenated hydrocarbons

Halogen is treated as a branch of a hydrocarbon. The parent chain is numbers following the same rules as for branched hydrocarbons, and the halogens are listed in alphabetic order, preceded by location number, like other alkyl branches.

Example 2.3.1

What is the IUPAC name of the following compound?



Solution

- The longest chain is the parent name: the longest chain is 6-C long, so the parent name is hexane.
- List the branches before the parent name in alphabetic order: the branches are chloro- and methyl-, so: chloromethylhexane
- Number the parent chain from the end that gives the lowest number to the first branch: Cl is at #2 from the right, and CH_3 is at #3 from the left, so number the parent chain from the right side that gives #2 to Cl and #4 to methyl.

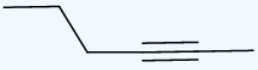
Answer: 2-chloro-4-methyl hexane

Example 2.3.2

Wright the skeletal structure of 5-chlorohex-2-yne?

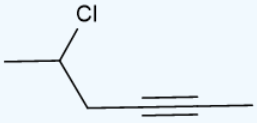
Solution

Find the parent name: hex-2-yne. Hex tells six C chain and 2-yne tells there is a triple bond at C#2, i.e.



.

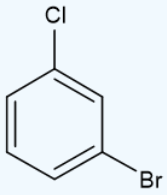
Attach the branches to the parent chain according to their location numbers: the only branch in this chase is chloro- at C#5.



Answer: 5-chlorohex-2-yne

Example 2.3.3

What is the IUPAC name of the following compound?



Solution

- The longest chain is the parent name: the longest chain is 6-C cycle with alternate double bonds, i.e., all the C′s are sp2-hybridized, i.e., an aromatic hydrocarbon with parent name benzene.
- List the branches before the parent name in alphabetic order: the branches are chloro- and bromo-, so: bromochlorobenzene
- Number the benzene ring starting from the 1st branch and going in the direction that gives the lowest number to the 2nd branch: Br is at #1, and Cl is at #3.

Answer: 1-bromo-3-chlorobenzene

Physical properties of halogenated hydrocarbons

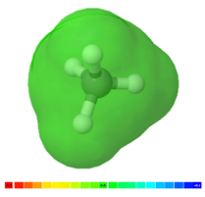
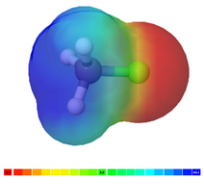
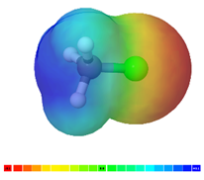
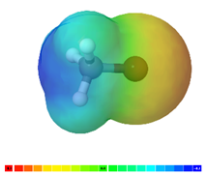
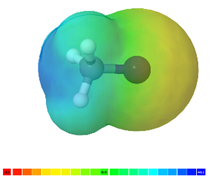
The halogen atom is sp^3-hybridized with three sp^3 orbitals occupied by lone pairs, and one makes bonds with the C, i.e., $R-\ddot{\underset{\cdot\cdot}{X}}:$, where X is a halogen: F, Cl, Br, or I. The lone pairs are usually not shown except when needed. The C−X bond is polar, i.e., $\overset{\delta+}{C}-\overset{\delta-}{X}$, because halogens are more electronegative than C. It makes the C a δ^+, i.e., an electrophile in reactivity. The lone pairs also add to the polarity because these electrons are localized on one side of the nucleus in sp^3 orbitals. The bond polarity can be observed from electrostatic potential maps, shown in Table 5. Recall that green is neutral, red is δ^- and blue is δ^+ in electrostatic potential map. Note that the listed electronegativity values of C and I are the same, but the bond is still polar. This is because the listed values are slightly different from the actual values. Further, the polarizability increases in this order: F<Cl<Br<I, i.e., I is the most polarizable.

What is polarizability

Polarizability is the tendency of a matter to acquire dipole-moment when subjected to an electric field. Larger atoms have more loosely held outermost electrons, which makes them more polarizable. Therefore, the polarizability increases in going from top to bottom in a group, e.g., the polarizability order of halogen is: F<Cl<Br<I.

Bond length increases in this order: F<Cl<Br<I and electronegativity has the opposite trend, i.e., decreases in this order: F>Cl>Br>I. The dipole moment depends on both the electronegativity difference of the bonded atoms and the bond length; the two effects contradict each other. As a result, the dipole moment does not change much by changing the halogen. The bond dissociation energy decrease in this order: F>Cl>Br>I, which makes C−F the strongest and the least reactive and C−I the weakest and the most reactive group.

Table 5: Physical characteristics of C−X bond in comparison with C−H bond.

Halogen	H for reference	F	Cl	Br	I
Electrostatic potential map					
C−X bond length (pm)	109	139	178	193	214

Halogen	H for reference	F	Cl	Br	I
C−X bond dissociation energy (kJ/mol)	440	464	355	309	228
Electronegativity of X or H	2.2	4.0	3.0	2.8	2.5
Dipole moment (D)	0.33	1.85	1.87	1.81	1.62

Solubility in water and boiling points of fluoro- and chloro-alkanes are comparable to alkanes of the same molar mass. Although the $C-F$ and $C-Cl$ are polar that tend to increase intermolecular forces due to dipole-dipole interaction, the volume of the molecule is small compared to the alkane of similar molar mass due to F or Cl being heavier atoms. Smaller volume means less contact area and fewer intermolecular forces. The opposing factors cancel each other.

Haloalkanes are denser than alkanes. Mono-fluoroalkanes and mono-chloroalkanes are less dense than water, but bromo- and iodo-alkanes are denser than water. For example, CH_3CH_2Cl and CH_3CH_2Br are liquids having densities of 0.891 g/mL and 1.354 g/mL at 25 ^oC. Although mono-chloroalkanes are less dense than water, di-, tri-, and tetra-chloroalkanes are denser than water. For example, CH_3Cl is a gas at room temperature with a density of 1.003 g/mL at its boiling point -23.8 ^oC, while CH_2Cl_2, $CHCl_3$, and CCl_4 are liquids with density higher than water, i.e., 1.327 g/mL, 1.483 g/mL, and 1.594 g/mL, respectively. So, hydrocarbons float on water while the denser haloalkanes sink in water.

Some uses of halogenated hydrocarbons

$C-I$, $C-Br$, and $C-Cl$ bond are polar and weaker than $C-H$ that makes them reactive functional groups. The raw organic material, i.e., alkanes, is usually first converted to haloalkanes intermediates to synthesize other organic compounds. The $C-F$ bond is stable, which is why Teflon, composed of $-(CF_2)_n-$ chains, is one of the most inert polymers. Haloalkenes with all of the $C-H$ bonds replaced with $C-F$ or $C-Cl$, called chlorofluorocarbons (CFCs) are usually nontoxic, nonflammable, odorless, and noncorrosive liquids that make them ideal as solvents, degreasing agents, and heat transfer fluids in air condition and refrigerator systems. For example, CCl_2F_2 (Freon-12) and CCl_3F (Freon-11) were used as refrigerants, and CCl_4 common name carbon tetrachloride was used as a solvent, cleaning and degreasing agent. They are being phased out because Freons destroy the stratosphere's ozone layer, and CCl_4 is toxic and carcinogenic.

Role of CFCs in the ozone hole

The CFCs like CCl_2F_2 and CCl_3F, being stable molecules, survive when released in the atmosphere, reach the stratosphere, and decompose when UV light shines on them in the stratosphere. Their by-products, particularly Cl_2, catalyze the destruction of the ozone layer in the stratosphere that filters out UV light. It resulted in an ozone-depleted region over the southern hemisphere called the ozone hole. Without the ozone layer, UV light reaches the earth's surface and causes damage to living things. Therefore, the CFCs are being replaced with haloalkanes with some $C-H$ bonds left to make them less stable, so they may decompose before reaching the stratosphere. As a result of this worldwide effort, the ozone hole is recovering over time, as shown in Figure $2.3.1$.

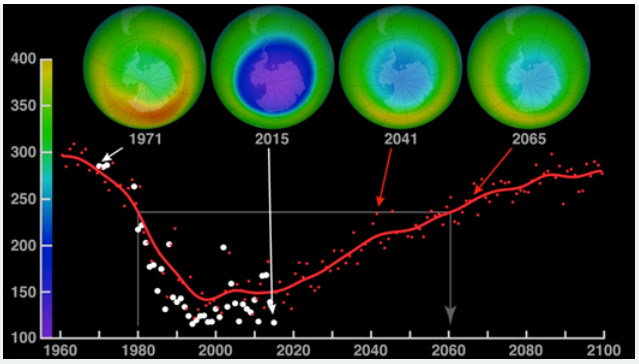


Figure 2.3.1: Ozone hole recovery projection, 1960-2100, the x-axis is time in years, and the y-axis shows the ozone concentration in Dobson units, where purple is the lowest and reddish is the highest concentration ration. (Copyright; NASA, Public domain, via Wikimedia Commons)

The most promising replacements include hydrofluorocarbons (HFCs) and hydrochlorofluorocarbons (HCFCs), like CF_3-CH_2F called HFC-134a and CH_3-CCl_2F called HCFC-141b. Carbon tetrachloride CCl_4, which is toxic and carcinogenic, is also being replaced with dichloromethane CH_2Cl_2 solvent.

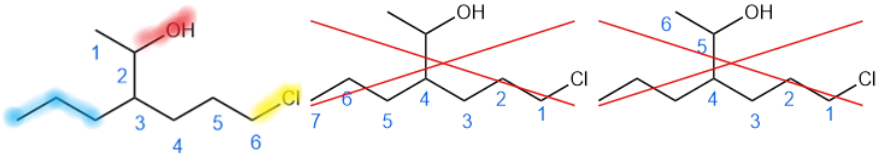
Alcohols

An alkyl group attached with an alcohol ($-OH$) group, i.e., $R-OH$, is an alcohol. A benzene ring attached with an alcohol group is a phenol.

Nomenclature of alcohols and phenols

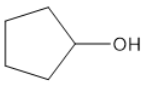
The nomenclature of hydrocarbons is followed for IUPAC names of alcohols and phenols with the following changes.

- The longest chain containing the $-OH$ group is chosen for the parent name, and the 'e' in the suffix -ane, -ene, or -yne is replaced with -ol. For example, CH_3-OH is methanol, and CH_3CH_2-OH is ethanol.
- Numbering starts from the end, giving the $-OH$ group the lowest number in preference over hydrocarbon or halogen branches. This is demonstrated in the following example, where the parent chain is not heptane; it is hexane that contains $-OH$ group, and the number starts from the end that gives $-OH$ lowest number 2 not from the other end that gives $-Cl$ the lowest number 1.

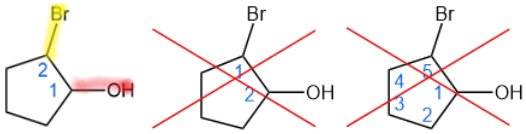


6-chloro-3-propyl hexan-2-ol

- In a cyclic compound containing only $-OH$ group, numbering is not needed, e.g, number is not used in naming cyclohexanone:

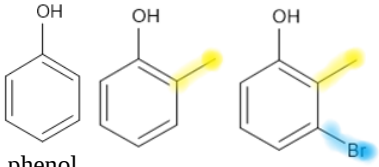


- In a cyclic compound containing $-OH$ group and a hydrocarbon or halogen group, numbering begins from $-OH$ group, e.g.:



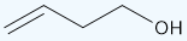
6-bromo cyclopentan-1-ol

- A benzene ring containing an $-OH$ group is given the parent name phenol. If there is another group present, the numbering starts from the $-OH$ group, and phenol is used as the parent name, e.g.:



phenol | 2-methyl phenol | 3-bromo - 2-methyl phenol

Example 2.3.1

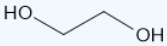
What is the IUPAC name of this compound? 

Solution

- Find the parent name: four C chain is but, and double bond makes it an alkene, i.e., butene.
- Change the 'e' of the suffix with -ol that tells it is an alcohol, i.e., butenol.
- Start number from the end that gives $-OH$ group the lowest number: $-OH$ receives #1 and $C=C$ receives #3.

Answer: but-3-en-1-ol

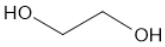
Example 2.3.2

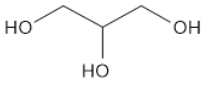
What is the IUPAC name of the following alcohol? 

Solution

- Find the parent name: two C chain is ethane.
- Replace the last 'e' of the suffix with -ol, which tells it is an alcohol; in this case, there are two $-OH$ groups, so add di, i.e., ethandiol
- The number is the same from either end, i.e., #1 and #2 for the two $-OH$ groups.

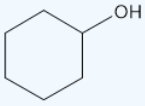
Answer: ethane-1,2-diol

Ethane-1,2-diol (

) is known by its common name ethylene glycol. Similarly, propane-1,2,3-triol (

) is known by its common name, glycerol or glycerine.

Example 2.3.3



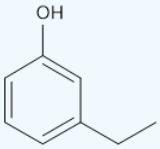
What is the IUPAC name of this compound?

Solution

- Find the parent name: six C cycle is cyclohexane.
- Replace the last 'e' of the suffix with -ol which means it is an alcohol, i.e., cyclohexanol.
- A cycle with only $-\mathrm{OH}$ group do not need numbering.

Answer: cyclohexanol

Example 2.3.4



What is the IUPAC name of this compound?

Solution

- Find the parent name: six C chain with all sp2-hybridized C′s is benzene, and benzene with an $-\mathrm{OH}$ group is phenol.
- List the substituent other than OH as prefix: ethylphenol.
- Start numbering form the $-\mathrm{OH}$ group and continue in the direction that gives the lowest overall number to the other substituents: ethyl receives #3.

Answer: 3-ethylphenol

Physical properties of alcohol and phenols

Alcohols contain a sp^3-hybridized O bonded to a C and a H and the remaining two sp^3-orbitals occupied by lone pairs, i.e.,: $\mathrm{R}-\ddot{\mathrm{O}}-\mathrm{H}$. The O atom is more electronegative than C and H, which means both the bonds are polar, i.e., $\overset{\delta+}{\mathrm{C}}-\overset{\delta-}{\mathrm{O}}-\overset{\delta+}{\mathrm{H}}$, as observed in the electrostatic maps of some alcohols in Figure 2.3.2.

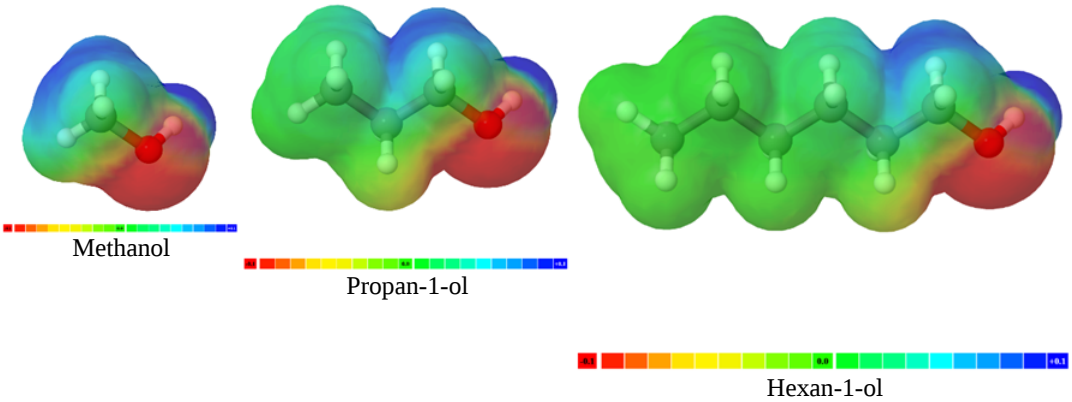


Figure 2.3.2: Electrostatic potential map of three alcohol examples, where O is red, H is white, and C is gray in the model spheres inside the electron cloud. $\delta-$ on O (red color), $\delta+$ on C (blue color), and $\delta-$ on H (blue color), and neutral on the alkyl chain (green color) can be observed. (Copyright; public domain).

An alcohol group has three reactive points: an electrophilic $\overset{\delta+}{\mathrm{C}}$, an acidic $\overset{\delta+}{\mathrm{H}}$, and nucleophylic/basic $\overset{\delta-}{\mathrm{O}}$. The acidity of an alcoholic H is obvious from the fact that it has a pKa ~16 that is about the same acid strength as $\mathrm{H_2O}$.

Nucleophile/base and electrophile/acid?

Recall that $\overset{\delta+}{\mathrm{C}}$ is called an electrophile. At the same time, proton exchange reactions are acid/base reactions, i.e., $\overset{\delta+}{\mathrm{H}}$ is an acidic proton, and a $\delta-$ site when reacts with a C is called nucleophile, and when it donates its lone pair to a H, it is called a base.

Alcohols have significantly higher boiling points compared to alkanes of comparable molar mass.

For example, the boiling point of methanol (molar mass 32 g/mol) is 65 ^{o}C which is significantly higher than the -89 ^{o}C boiling point of ethane of comparable molar mass (30 g/mol). It is explained by the fact that, in addition to London dispersion forces, alcohols have hydrogen bonding between $\overset{\delta+}{\mathrm{H}}$ of one molecule with $\overset{\delta-}{\mathrm{O}}$ of the neighboring molecule as illustrated in Figure 2.3.3. It takes more thermal energy to break London dispersion forces + hydrogen bonding in alcohols than London dispersion forces alone in alkanes. The boiling points of alcohols increase as their molar mass increases, as shown in Table 6.

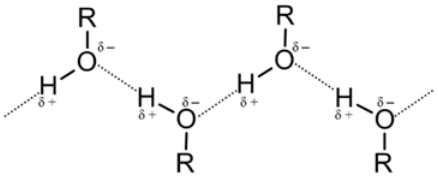


Figure 2.3.3: Illustration of hydrogen bonding in alcohol. (Copyright; Secalinum, Public domain via Wikimedia Commons)

Smaller alcohols having up to three C' are entirely soluble in water, i.e., miscible, one with four C' is partially soluble, and those with longer than four C' are almost insoluble as listed in Table 6. This trend of water solubility change is explained by a balance tween hydrophilic and hydrophobic components in an alcohol molecule, explained later.

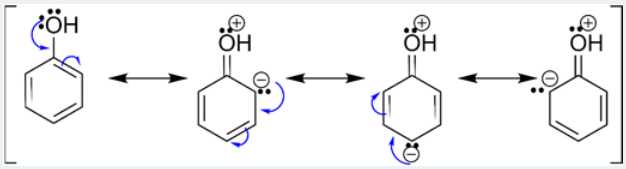
Table 6: Molecular weight, boiling points, and solubility in water of some alcohols

IUPAC name	Molecular formula	Molecular weight (g/mol)	Boiling point (^{o}C)	Solubility in water
Methanol	$\mathrm{CH_3OH}$	32	65	Miscible
Ethanol	$\mathrm{CH_3CH_2OH}$	46	78	Miscible
Propan-1-ol	$\mathrm{CH_3CH_2CH_2OH}$	60	97	Miscible
Butan-1-ol	$\mathrm{CH_3CH_2CH_2CH_2OH}$	74	117	Slightly soluble
Pentan-1-ol	$\mathrm{CH_3CH_2CH_2CH_2CH_2OH}$	88	138	Insoluble
Hexan-1-ol	$\mathrm{CH_3CH_2CH_2CH_2CH_2CH_2OH}$	102	157	Insoluble

Phenol has an $-\mathrm{OH}$ attached to an sp^2-hybridized C of a benzene ring. Phenol is a volatile white crystalline solid that melts at 41 ^{o}C, and boils at 182 ^{o}C. Phenol is partially soluble in water.

$-\mathrm{OH}$ -An activating group on a benzene ring

When an atom with a lone pair of electrons on it is bonded to an sp^2-hybridized C, it changes its hybridization from sp^3 to sp^2 for resonance to happen. The driving force is the resonance stabilization effect. The same happens to O in phenol resulting in resonance that sends electrons from O into the benzene ring as illustrated below. Consequently, the benzene ring of phenol is more electron-rich, i.e., more nucleophilic than benzene without such a group attached. The groups like $-\mathrm{OH}$ that enhance the reactivity of the benzene ring by donating electrons to the ring are called **activating groups**.



Another consequence of this electron donation by O is that the O is more able to hold on to the electrons of O−H bond when the proton leaves acting as an acid, i.e., phenolic O−H is a stronger acid (pKa ~10) compared to an alcoholic O−H (pKa ~16).

The balance between the hydrophilic and hydrophobic character of organic compounds containing a polar group

Polar functional groups like −OH are attracted to water, i.e., water-loving or **hydrophilic**. This is based on the general principle "like dissolves like." On the other hand, Alkyl groups are nonpolar and repel water molecules, i.e., **hydrophobic**. Alkyl groups are more soluble in lipids and fats, which are also called **lipophilic or nonpolar**. Alcohols have both groups in them, and their properties depend on which group dominates. The hydrophilic character dominates when the alkyl chain is small, three (C′s or small, and the hydrophobic character dominates when the alkyl chain is long, i.e., more than four (C′s long making the compound slightly soluble or insoluble. In summary,

- increasing the alkyl-group increases hydrophobic character, e.g., propan-1-ol ($CH_3CH_2CH_2OH$) is water soluble, but butan-1-ol ($CH_3CH_2CH_2CH_2OH$) is partially soluble, and
- increasing the polar group increases the hydrophilic character, e.g., butan-1-ol ($CH_3CH_2CH_2CH_2OH$) with one ($-OH$) group is partially soluble in water, but butane-1,2-diol ($CH_3CH_2CH(OH)CH_2OH$) with two ($-OH$) groups is soluble.

This trend is generally applicable to all polar groups containing organic compounds.

Some examples of important alcohols and phenols

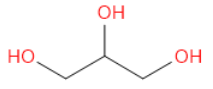
Methanol (CH_3-OH) is used as a solvent, paint remover, and fuel and converted to formaldehyde which is then used to make plastics. If methanol is ingested, it converts to formaldehyde, which causes headache and blindness and may cause death. That is why methanol is mixed with ethanol not meant for drinking.

Ethanol (CH_3-OH) is the active ingredient in alcoholic beverages. It is also used as a solvent for perfumes, varnishes, and medicines, e.g., in tincture iodine. The fermentation process obtains ethanol, but these days, it is mainly derived from ethene.

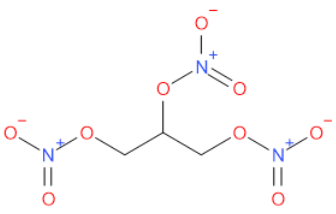
Isopropanol. 2-methylpropanol, commonly known as isopropanol, kills bacteria and viruses. Ethanol also has the same property. Ethanol or isopropanol are 60% to 80% fraction of hand sanitizers; the remaining is usually ethylene or glycerin to keep the skin soft. If ethanol is the active ingredient in the hand sanitizer, care is needed as ethanol is volatile and flammable.

Ethylene glycol, i.e., ethane-1,2-diol ($HO-CH_2CH_2-OH$) is used as antifreeze in cooling systems of vehicles and as a solvent for paints, inks, and plastics. Ethylene glycol is poisonous because, if ingested, it converts to oxalic acid, which causes kidney stones.

Glycerin or glycerol, i.e., propane-1,2,3-triol ($HO-CH_2CH(OH)CH_2-OH$) is obtained as a byproduct during the soap making from fats and oils. It is a viscous liquid used in skin lotions, cosmetics, shaving creams, and liquid soaps. Nitroglycerin, the active component of explosive dynamite, is derived from glycerol.



Glycerol


Nitroglycerin

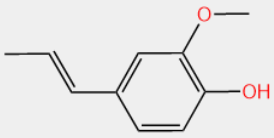
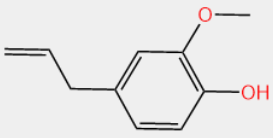
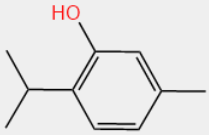
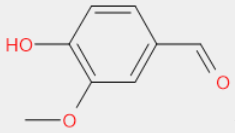
Bisphenol a

Bisphenol a (BPA) is used used in making polycarbonate that is used to manufacture beverage bottles. There is a concern that leaching out BPA from beverage bottles may cause harmful health effects.

Phenol is primarily used for the production of precursors of plastics. Phenol is used as an antiseptic and disinfectant. Phenol is also a part of the structure of essential oils responsible for plants' odor and flavor, e.g., isoeugenol from nutmeg, eugenol from clover, thymol from thyme, and vanillin from vanilla, shown in Table 1.

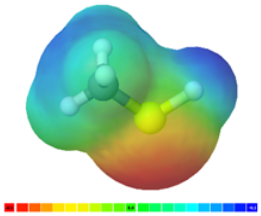
Table 3: Examples of phenol present in essential oils

Isoeugenol from nutmeg	Eugenol from clover	Thymol from thyme	Vanillin from vanilla

Isoeugenol from nutmeg	Eugenol from clover	Thymol from thyme	Vanillin from vanilla
			
 Isoeugenol	 Eugenol	 Thymol	 Vanillin

Thiol

Thiols are sulfur analogs of alcohols, i.e., they contain the thiol ($-SH$) group. Like alcohols, thiols contain an electrophilic C and acidic H attached to S atom as illustrated by the electrostatic potential map of methanethiol CH_3-SH in the figure on the right, where S is shown as yellow, H as white, and C as gray sphere in the model.



The thiols are named following the rules of alcohols, except that suffix -thiol is added to the name of alkane, e.g., CH_3-SH is methanethiol and (\ce{CH3CH2-SH}\)) is ethanethiol.

One characteristic of thiols is that they have a strong odor, e.g., methanthiol (CH_3-SH) is an odor-causing compound present in oysters, cheddar cheese, onions, and garlic. The garlic odor also contains prop-2-ene-1-thiol (\ce{CH2=CH-CH2-SH}\). Onion odor is also due to propane-1-thiol ((\ce{CH3CH2CH2-SH}\), which is a lachrymator, i.e., a substance that makes eyes tear.

Ethers

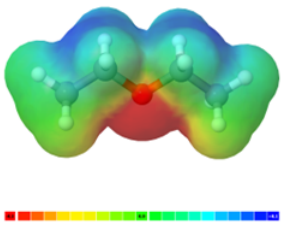
Ethers have O with two single bonds with two alkyl groups, i.e., $R-O-R'$ group.

Nomenclature of ethers

IUPAC naming of ether follows the rules of branched alkanes. The large alkyl group is named as parent chain, and the smaller alkyl groups with oxygen are named as a branch, i.e., an alkoxy ($R-O-$) group. For example, $CH_3CH_2-O-CH_2CH_3$ is ethoxyethane and $CH_3CH_2CH_2-O-CH_3$ is 1-methoxypropane . However, trivial names for ethers are often used. A common name is formed by listing the two alkyl groups in alphabetic order, followed by the word ether. For example, common name of $CH_3CH_2-O-CH_2CH_3$ is diethylether and of $CH_3CH_2CH_2-O-CH_3$ is propyl methyl ether.

Physical properties of ethers

Ethers contain a sp^3-hybridized O bonded to two C′s and the remaining two sp^3-orbitals occupied by lone pairs, i.e.,: $R-\ddot{O}-R$. The O atom is more electronegative than C, which means both the bonds are polar, i.e., $\overset{\delta+}{C}-\overset{\delta-}{O}-\overset{\delta+}{C}$, as observed in the electrostatic maps of diethylether shown on in figure on the right.



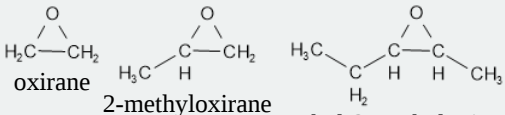
Since there is no $O-H$ bond, pure ethers do not have hydrogen bonding; they have a weak dipole-dipole interaction and London dispersion forces. Therefore, the boiling points of ether are comparable to alkanes of the same molecular mass, e.g., diethyl ether (molar mass 74 g/mol) has a boiling point of 35 °C, which is almost the same as the boiling point of pentane (molar mass 72 g/mol, boiling point 36 °C). However, ethers do have $\overset{\delta-}{O}$ and that can establish hydrogen bonding with $\overset{\delta+}{H}$ of water molecule. So, ethers have solubility in water comparable to alcohols of the same

molar mass, i.e., ethers with three or less $C's$ are miscible in water, and those with four or more $C's$ are slightly soluble or insoluble in water.

Ethers are less reactive than alcohols because alcohols have an acidic $\overset{\delta+}{H}$, which is missing in ethers. Because of relatively low chemical reactivity, ethers are used as solvents to conduct other compounds' reactions.

Epoxides -reactive ethers and sterilizing agents

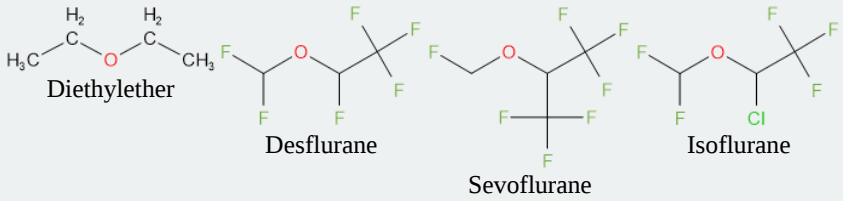
Epoxides are cyclic ethers in which one atom in a three-membered ring is O. Epoxides take the parent name oxirane, and numbering, if needed, starts from the O atom. Three examples are shown below.



Epoxides are unstable and highly reactive due to angle strain as the bonds bent from a regular angle of 109.5° to 60°. They are used as intermediates in organic synthesis. Oxirane is a room-temperature gas that reacts fast with the compounds in microorganisms, causing their death. So, it is used as a fumigant in foodstuffs and textiles and as a hospital sterilizer for surgical instruments.

Ethers as anesthetics

Diethyl ether has been used as an anesthetic agent for a long time. Its use in anesthesia has been discontinued because it is a volatile and flammable gas posing a fire or explosion risk in the surgery room. Further, it has an irritating effect on the respiratory passages and causes nausea. Fluorinated or chlorofluorinated ethers, such as desflurane, sevoflurane, and isoflurane listed below, are used as anesthetic agents because they are less volatile, less flammable, and have fewer other undesirable side effects.



Peroxides, sulfides, and disulfides

Peroxides have $O-O$ bond, which is one of the weak bonds with bond dissociation energy ~200 kJ/mol, about half of the strength of a $C-C$ or a $C-H$ bond. This bond is found in **hydroperoxy group** ($R-O-O-H$) and **peroxy group** ($R-O-O-R'$. The peroxides are less common but more reactive due to weak $O-O$ bond.

Sulfides are S analogs of ethers, i.e, they have sulfide ($R-\ddot{S}-R$) group. The names of sulfides are similar to those of ethers, i.e., list the two alkyl groups alphabetically, followed by the word sulfide. For example, $CH_3CH_2-S-CH_2CH_3$ is diethyl sulfide and $CH_3CH_2CH_2-S-CH_3$ is propyl methyl sulfide. Sulfides are less common in nature. **Disulfide** $R-S-S-R$ group, i.e., a sulfur analog of peroxides found in proteins. The $S-S$ is a relatively weak bond with a bond dissociation energy of ~250 kJ/mol.

Primary (*p*), secondary (*s*), tertiary (*t*), and quaternary classification

The *p*, s, *t*- classification of $C's$ and $H's$

In an organic compound:

- if a C is not bonded with any other C or bonded with only one other C, it is a **primary** C;
- if bonded with two other $C's$, it is a **secondary** C;
- if bonded with three other \(\ce{C's}\), it is a **tertiary** C; and
- if bonded with four other $C's$, it is a **quaternary** C.

The H′s on a primary C′s are **primary** H′s, those on a secondary C′s are **secondary** H′s, and those on tertiary C′s are tertiary H′s. Quaternary C′s do not have any H on them. The figure on the right illustrates the primary, secondary, tertiary, and quaternary C′s and H′s marked in different colors and pointed by arrows. Table 1 illustrates the primary, secondary, and tertiary haloalkanes, alcohols, alkyl groups in ethers, and amine.

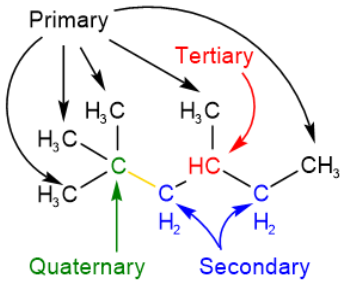
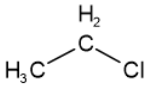
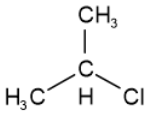
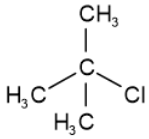
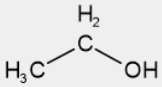
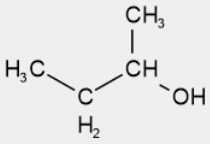
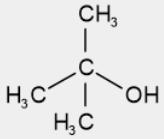
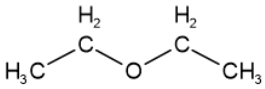
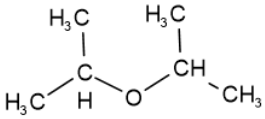
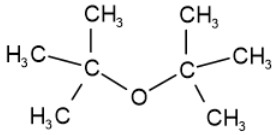
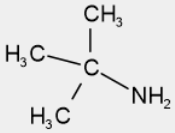
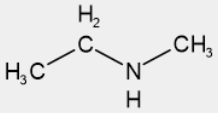
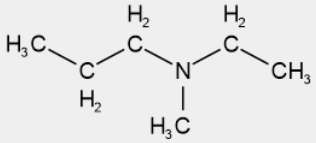


Table 1: Primary, secondary, and tertiary classifications of haloalkanes, alcohols, alkyl groups, and amines.
(Common names using secondary (*s*) and tertiary (*t*) designation or *iso* (for the case of secondary propyl) alkyl groups are used.)

Classification	Primary	Secondary	Tertiary
Haloalkanes	 ethyl chloride	 *iso*propyl chloride	 *t*-butyl chloride
Alcohols	 ethanol	 *sec*-butanol	 *t*-butanol
Alkyl groups in ethers	 *di*ethyl ether	 *diiso*propyl ether	 *ditertiary* butyl ether
Amines	 *t*-butyl amine	 ethyl(methyl)amine	 ethyl(methyl)propylamine

The *p*, s, *t*- classification of haloalkanes and alcohols

A halogen or an −OH group:

- if bonded with a primary C, it is a **primary haloalkane,** or **primary alcohol**;
- if bonded with a secondary C, it is a **secondary haloalkane** or **secondary alcohol**;
- if bonded with a tertiary \(\ce{C}\), it is a **tertiary haloalkane** or **tertiary alcohol**.

Quaternary C′s do not have any halogen or −OH group on them.

The *p*, s, *t*- classification of alkyl groups

An alkyl group connected through:

- a primary C to a parent chain or a functional group is a **primary alkyl group (p-alkyl)**;
- a secondary C is a **secondary alky group (s-alkyl)**; and
- a tertiary C is a **tertiary alkyl group (t-alkyl)**.

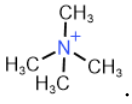
The primary (*p*), secondary (*sec*, or *s*), and tertiary (*t*) designation of alkyl groups (except for *iso*propyl used in the place of sec-propyl) are also accepted in IUPAC nomenclature, and these designations are often used in common names of organic compounds,

as shown by the commons in Table 1.

The *p*, s, *t*- classification of amines

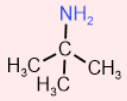
Amines are organic compounds that have one or more H′s of ammonia (NH_3) replaced with an aliphatic hydrocarbon group.

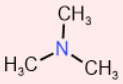
- If there is only one alkyl group bonded with the N, it is a **primary amine group**, i.e., RNH_2;
- if two alkyl groups are bonded with the N, it is a secondary amine group, i.e., $RR'NH$;
- if three alkyl groups are bonded with the N, it is a **tertiary amine group**, i.e., $RR'R''N$, and
- if four alkyl groups are bonded with the N, it is not an amine but an ammonium ion, i.e., $RR'R''R'''N^+$, e.g.,

tetramethylammonium:  .

Caution-*p*, *s*, or *t* designation of amines is based on the N not on the C

Unlike haloalkenes and alcohols, the primary, secondary, and tertiary classification of amines is based on whether there are one, two, or three alkyl groups bonded with the N, irrespective of the C bonded with the N is primary, secondary, or tertiary. For example, as shown below, *t*-butylamine is a primary, and trimethylamine is a tertiary amine.


t-butylamine -a primary amine


trietylamine -a tertiary amine

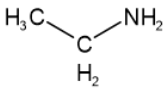
Amines

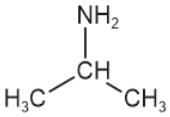
Ammonia has an sp^3-hybridized N having three single bonds with nitrogen and one lone pair, i.e., $:NH_3$. If one or more H′s of ammonia are replaced with an aliphatic hydrocarbon, it is an amine group, and the compound is an amine, i.e., RNH_2, $RR'NH$, or $RR'R''N$ are amine groups. If one or more H′s of ammonia are replaced with a benzene ring, it is an aromatic amine.

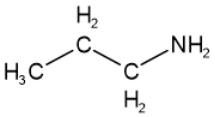
Nomenclature of amines

IUPAC naming of amines follows the rules for naming alcohols with the following changes:

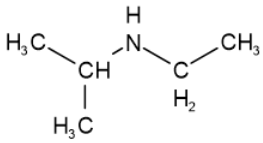
- The suffix's last letter 'e' is replaced with amine, e.g., CH_3NH_2 is methanamine. A few examples are the following.

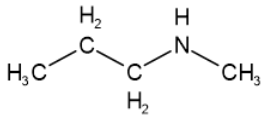

ethanamine

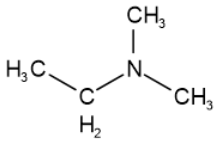

propan-2-amine


propan-1-amine

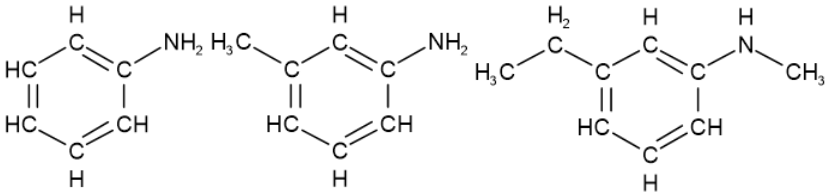
- If more than one hydrocarbon group is attached with the N, the longest one is chosen as the parent name, and the smaller ones are listed alphabetically as substituents preceded by N-, where N tells the branch is attached to N. For example, $CH_3CH_2NHCH_3$ is N-methylethanamine.
- An alternate and easier way is to list the alkyl groups as substituents in alphabetic order (the middle one is enclosed in small brackets for easy reading), followed by the word amine. For example, $CH_3CH_2NHCH_3$ can be named ethyl(methyl)amine. A few examples are listed below.


propan-1-amine


ethyl(methyl)propylamine


methyl(propan-2-yl)amine

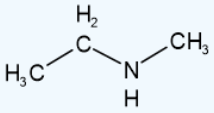
- A benzene ring containing an $-NH_2$ group is given the parent name aniline. If there is another group present, the numbering starts from the $-NH_2$ group, and aniline is used as the parent name, e.g.:



aniline 3-methylaaniline 3-ethyl-N-methylaniline

Example 2.3.1

What is the IUPAC name of the compound shown in the figure on the right?

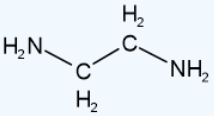


Solution

1. List the alkyl groups in alphabetic order (place the middle name in brackets) and end with the word amine: ethyl(methyl)amine, or
2. Choose the longest substituted as the parent name and the other as substitute preceded by N- to indicate the substituent is on N, i.e.,: N-methylethanamine.

Example 2.3.2

What is the IUPAC name of the compound shown in the figure on the right?

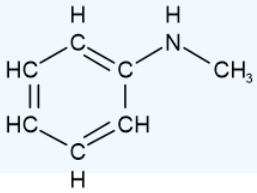


Solution

Ethane is the only alkyl group, so the stem name is ethane. It has two amine groups, so use di-preceded by number to tell the location of the amine groups: ethan-1,2-diamine.

Example 2.3.3

What is the IUPAC name of the compound shown in the figure on the right?

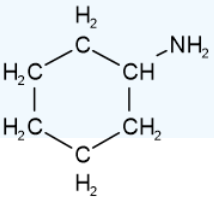


Solution

A benzene ring attached to an amine group has the parent name aniline. Methyl substituent is on N, so list it as a prefix branch name preceded by N-, i.e., N-methylaniline.

Example 2.3.4

What is the IUPAC name of the compound shown in the figure on the right?

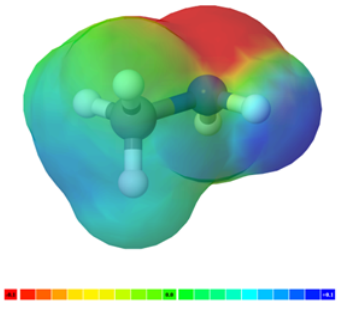


Solution

The alkyl group is cyclohexane; replace the last 'e' with amine: cyclohexanamine.

Physical properties of amines

Amines contain a sp^3-hybridized N bonded to C′s and H′s and one sp^3-orbitals occupied by lone pairs, i.e.,: $RR'R'N$:. The N atom is more electronegative than C, which means the C−N and N−H bonds are polar, i.e., $\overset{\delta+}{C}-\overset{\delta-}{N}-\overset{\delta+}{H}$, as observed in the electrostatic maps of methanamine shown in the figure on the right, where N is blue, C is gray, and H is a white sphere.



The N atom is less electronegative than O, so the C−N bond (3.0-2.1 = 0.9) is less polar and less reactive than C−O (3.5-2.1 = 1.4). Due to the less electronegativity of N, it is more willing to donate its lone pair to any proton around, i.e., amiens are basic compounds. Amines are also classified as organic bases.

Primary and secondary amines have hydrogen bonding, but the $N-H$ bond is less polar than the $O-H$ bond. Therefore, the boiling points of primary and secondary amines are higher than alkanes but lower than alcohol of comparable molar mass, as shown in Table 8.

Table 8: Molar masses and boiling points of some amines

Name	Condensed	Molar mass	Boiling point
Ethane	CH_3CH_3	30.1 g/mol	-88.6 ^oC
Methylamine	CH_3NH_2	31.1 g/mo	-6.3 ^oC
Methanol	CH_3OH	32.0 g/mol	65.0 ^oC
Butane	$CH_3CH_2CH_2CH_3$	58 g/mole	0 ^oC
Ttrimetylamine	$(CH_3)_3N$	59 g/mole	3 ^oC
Propylamine	$CH_3CH_2CH_2NH_2$	59 g/mol	48.0 ^oC

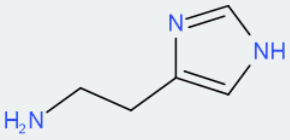
Tertiary amines have boiling points comparable with alkanes as they do not have $N-H$ bond and do not make hydrogen bonding with each other, as shown in Table 8.

The primary, secondary, and tertiary amines make a hydrogen bonds with water. Therefore, amines containing up to five $C's$ are soluble in water, and those with more than five $C's$ are slightly soluble or insoluble.

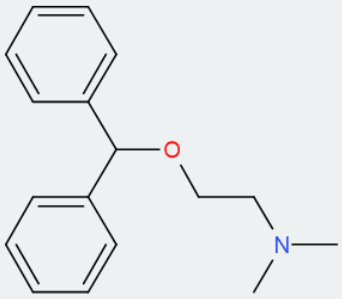
Importance of amines in health and medicine

Amines are present in amino acids, the building blocks of proteins. Amines are present in several physiologically active compounds. The body releases the histamine in response to injury or allergic reactions. Histamine causes blood vessels to dilate, and redness and swelling occur in the area. Antihistamine administered to block the effects of histamine is another amine, diphenhydramine.

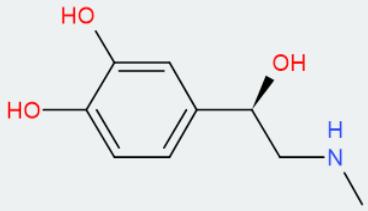
Epinephrine (adrenaline) and norepinephrine (noradrenaline) are released in a "fight-or-flight" situation. They increase blood glucose and move the blood to the muscles. Norepinephrine is a remedy for colds, hay fever, and asthma. It contracts the capillaries in the mucous membranes of the respiratory passage.



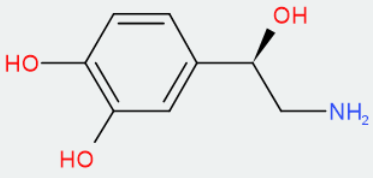
Histamine



Diphenhydramine

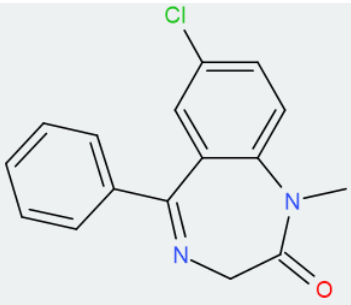


Adrenaline

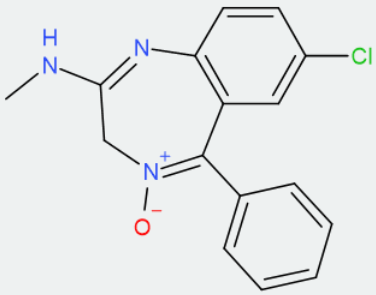


Noradrenaline

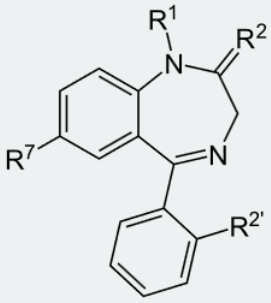
Tranquilizers are drugs that relieve the symptoms of anxiety and tension. These include diazepam (earlier name valium), Chlordiazepoxide (trade name Librium), and other benzodiazepines, which are sedative and hypnotic medicines that cause calming effects and drowsiness.



Diazepam (Valium)



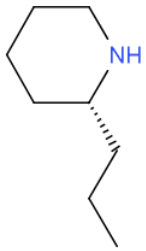
Chlordiazepoxide (Librium)



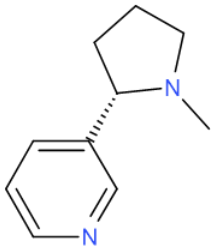
Benzodiazepines

Alkaloids

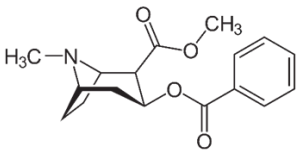
Amines can donate a lone pair to a proton relatively easily because N is less electronegative than O, i.e., amines are basic compounds with pKa value ~10. Alkaloids are nitrogen-containing basic compounds extracted from plants. Most alkaloids are physiologically active and used in anesthetics, antidepressants, and stimulants; many are habit-forming. For example, coniine extracted from "poison hemlock" can cause weakness, fast respiration, paralysis, and death. Nicotine found in tobacco is addictive, but in large doses, it causes depression, nausea, vomiting, or death. The solution of nicotine in water is used as an insecticide. Cocaine extracted from the coca plant is a stimulant for the central nervous system. Piperidine is responsible for the pungent smell and taste of black pepper. The Chemical structures of these alkaloids are shown below.



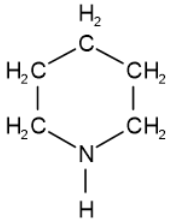
(S)-Conine



Nicotine

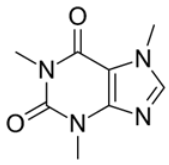


Cocaine

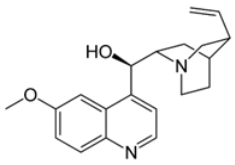


Piperidine

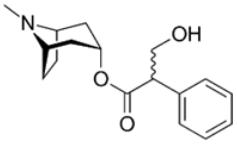
Caffeine is found in coffee and tea. Caffeine increases alertness and is used in some pain relievers to counter the drowsiness caused by antihistamines. Quinine from the bark of the cinchona tree is used to treat malaria. Atropine from belladonna accelerates slow heart rates and is anesthesia for eye exams. Their structures are shown below.



Caffeine



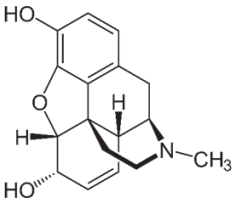
Quinine



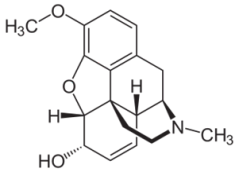
Atropine

Morphine and codeine are obtained from the opium poppy plant and used as painkillers and cough syrup. Heroin, which is strongly addictive, is received by the chemical modification of morphine. OxyContin, which is a prescription drug for the relief of fever pain, is structurally similar to heroin and also has similar physiological effects. The structures of these alkaloids are shown below.

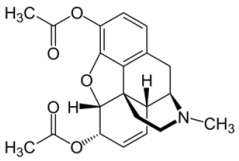
Seedhead of Opium Poppy



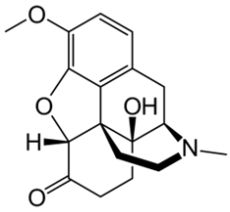
Morphine



Codeine



Heroin



OxyContin

2.4: Functional groups containing sp2-hybridized heteroatom

Learning Objectives

- Identify, given IUPAC names to the structural formula, and draw the structural formula from the IUPAC name of simple aldehydes, ketones, and imines.
- Understand the polarity and predict some physical properties, reactive sites, and relative reactivities of aldehydes, ketones, and imines based on the bond polarity.

Carbonyl group and its subclasses

A (C=O) group is a carbonyl group. It has a σ-bond between sp^2 orbitals and a π between p orbitals of a C and an O. Lone pairs of electrons occupy the remaining two sp^2 orbitals of O as shown here: $-\overset{|}{\underset{|}{C}}=\ddot{O}:$. The lone pairs are usually not shown. The carbonyl group is represented as $-\overset{|}{\underset{|}{C}}=O$, $-\overset{O}{\overset{||}{C}}-$ or simply as C=O. If the O is replaced with N it becomes an imine C=NR group.

The carbonyl group is subdivided into aldehydes, ketones, carboxylic acids, and carboxylic acid derivates based on what is bonded to the carbonyl carbon, i.e., what are the X and Y in this formula: $X-\overset{O}{\overset{||}{C}}-Y$, as listed in Table 1.

Table 1: Sub-groups of carbonyl group based on what is X and Y in the general formula of a carbonyl group: $X-\overset{O}{\overset{||}{C}}-Y$.

X	Y	Group name	General formula
Hydrocrbon (R) or H	H	Aldehyde	$R-\overset{O}{\overset{||}{C}}-H$
Hydrocarbon (R)	Hydrocarbon (R′)	Ketone	$R-\overset{O}{\overset{||}{C}}-R'$
Hydrocrbon (R) or H	$-\overset{NR}{\overset{||}{C}}-R''$	Imine	$R'-\overset{NR}{\overset{||}{C}}-R''$
Hydrocarbon (R) or H	Alcohol group (−OH)	Carboxylic acid	$R-\overset{O}{\overset{||}{C}}-OH$
Hydrocarbon (R) or H	Oxy anion ($-O^-$)	Carboxylate anion	$R-\overset{O}{\overset{||}{C}}-O^-$
Hydrocarbon (R) or H	Halogen (−X, where −X = −F, −Cl, −Br, or −I)	Acid halide	$R-\overset{O}{\overset{||}{C}}-X$
Hydrocarbon (R) or H	Carboxyl $(-O-\overset{O}{\overset{||}{C}}-R')$	Acid anhydride	$R-\overset{O}{\overset{||}{C}}-O-\overset{O}{\overset{||}{C}}-R'$
Hydrocarbon (R) or H	Alkoxy (−OR′)	Ester	$R-\overset{O}{\overset{||}{C}}-OR'$
Hydrocarbon (R) or H	Amine ($-NH_2$, −NHR′, or −NR′R″)	Amide	$R-\overset{O}{\overset{||}{C}}-NH_2$
Hydrocarbon (R)	Cyano ($-C\equiv N$)	Nitrile	$R-C\equiv N$

The first three, i.e., aldehydes, ketones, and imines, are described in this section. The others, i.e., carboxylic acids and their derivatives, including carboxylate anion, acid anhydrides, acid halides, esters, and amides, are described in the next section. The nitrile group is classified as a carboxylic acid derivative.

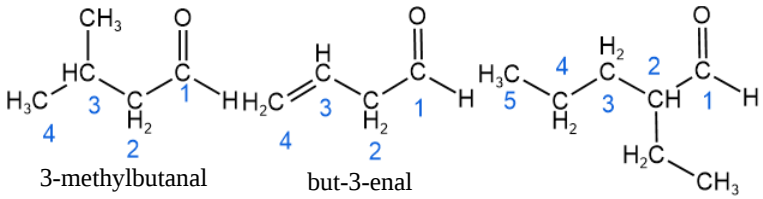
Aldehydes

An aldehyde has either two H's or a H and a hydrocarbon R single bonded with the carbonyl carbon, i.e., $\mathrm{R}-\overset{\overset{\mathrm{O}}{\|}}{\mathrm{C}}-\mathrm{H}$ group. The condensed form of the aldehyde group is $-\mathrm{CHO}$.

Nomenclature of aldehydes

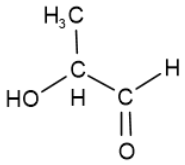
The IUPAC naming of aldehydes follows the rules of naming alcohols, with the following changes.

- The longest chain **containing the aldehyde group** is chosen as the parent name with the last 'e' of the suffix replaced with -al, e.g.,$\mathrm{CH_3CHO}$ is ethanal.
- Start numbering from the aldehyde C. If there is one aldehyde group, it is understood to be #1, i.e., no need to show the number. A few examples are shown below.

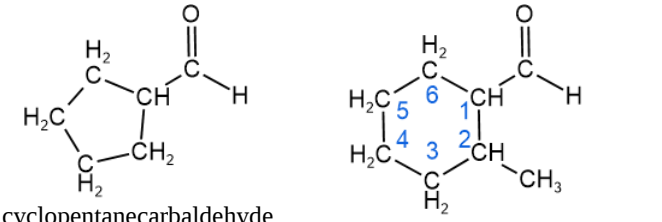


3-methylbutanal but-3-enal 2-ethylpentanal

- If more than one functional groups are present, the order of preference of functional groups is the following: $\overleftarrow{\text{carboxylic acid}>\text{aldehyde}>\text{ketone}>\text{alcohol}>\text{amine}>\text{thiol}}$. A suffix represents the higher priority group, and prefixes represent the other groups. For example, 2-hydroxypropanal shown on the right.

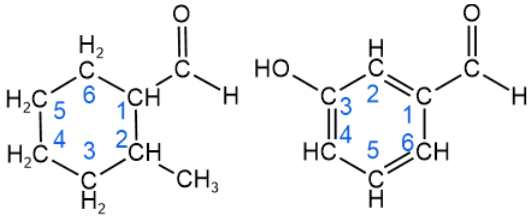


- The suffix "carbaldehyde" represents an aldehyde group bonded to a cycloalkane. Numbering starts from the point of attachment of the aldehyde to the cycloalkane, as shown in the examples below.



cyclopentanecarbaldehyde 2-methylcyclohexane-1-carbaldehyde

- An aldehyde group bonded to a benzene ring has the parent name benzaldehyde, as shown in the following examples. Numbering, if needed, starts from the aldehyde attachment point.

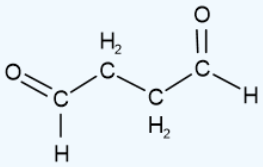


benzaldehyde 3-hydroxybenzaldehyde

The common name of methanal ($\mathrm{H_2C=O}$) is formaldehyde, and ethanal ($\mathrm{CH_3-CHO}$) is acetaldehyde. Aldehydes take their common names from the common names of carboxylic acids, described in a later section.

Example 2.4.1

What is the IUPAC name of this compound shown in the figure on the right?



Solution

the longest chain containing aldehyde groups is four C, i.e., butane.

Change the final 'e' with -al, and since there are two aldehyde groups, use dial.

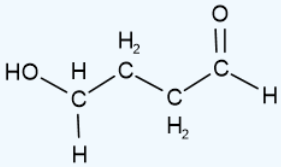
Answer: butanedial

Note 1: the final 'e' is dropped when the following letter added starts with a vowel, e.g., -al; otherwise, it is retained as in butanedial.

Note 2: Location number is not needed even for this dialdehyde because it is understood that the aldehyde groups are at the end of the chain.

Example 2.4.2

What is the IUPAC name of this compound shown in the figure on the right?



Solution

The longest chain is four C, i.e., butane

There is an aldehyde group, so change the final 'e' with -al, i.e., butanal.

There is an alcohol group that is added as a hydroxy- prefix, i.e., hydroxybutanal.

The alcohol group needs the location number, start from aldehyde, and the alcohol group receives #4.

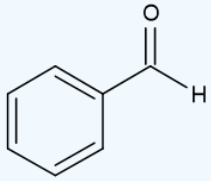
Answer: 4-hydroxybutanal.

Example 2.4.3

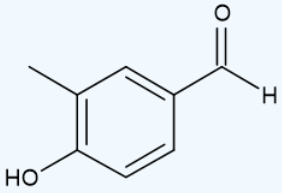
Write the skeletal formula of 4-hydroxy-3-methylbenzaldehyde?

Solution

The parent name is benzaldehyde, i.e., an aromatic ring with an aldehyde group as shown on the right.



Start the number from the point of attachment of aldehyde and go either clockwise or counterclockwise and place a methyl group at #3 and an alcohol group at #4.

Answer: 

Physical properties of aldehydes

The aldehydes have sp^2-hybridized C and O with a double bond (a σ and a π-bond) between them, two lone pairs occupying two sp^2-orbital of an O, and the C bonded with a H and a hydrocarbon R or another H, i.e., $-\underset{|}{C}=\ddot{O}:$. The C=O bond is polar, i.e., $\overset{\delta+}{C}=\overset{\delta-}{O}$, because O are more electronegative than C (3.3-2.1 = 0.9), as shown in Figure 2.4.1. It makes the C a δ^+, i.e., an electrophile in reactivity and the O a δ^-, i.e., a nucleophile or a base in reactivity. The lone pair of electrons on O add to the base character of the carbonyl O.

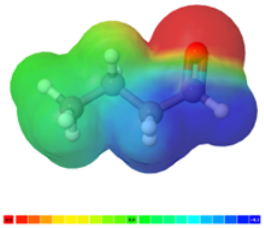
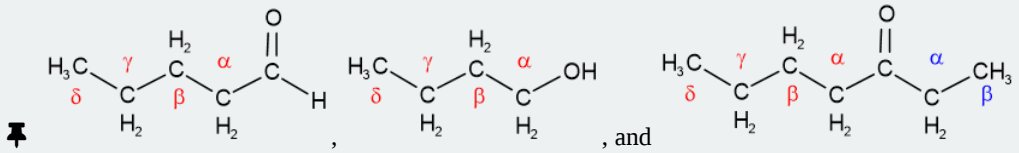


Figure 2.4.1: Electrostatic potential map of butanal ($CH_3CH_2CH_2CHO$). C′s are gray, H′s are white is read and O is red in the model. Blue region is δ^+, red is δ^-, and green is neutral. (Copyright; Public domain)

Note that the δ^+ region extend from carbonyl C to the C attached to it, designated as αC and the H′s on the αC.

α-, β-, γ-, δ- designations of C′s

The C directly bonded to a functional group is designated as α, the next one as β, the third as γ, and so on. A few examples are shown below for explanation.



, , and

Aldehyde group ($\overset{\delta+}{C}=\overset{\delta-}{O}$) is polar. Aldehydes have boiling points higher than the alkanes of comparable molar mass due to the dipole-dipole interaction in addition to London dispersion forces. Aldehydes have boiling points lower than alcohols of the comparable mass because aldehydes do not have hydrogen bonding with each other, as compared below.

Name	Condensed formula	Molar mass	Boiling point
Pentane	$CH_3CH_2CH_2CH_2CH_3$	72 g/mol	36 °C
Butanal	$CH_3CH_2CH_2CHO$	72 g/mol	76 °C
Butanol	$CH_3CH_2CH_2CH_2OH$	74 g/mol	117 °C

Aldehydes can establish hydrogen bonding with water molecules through $\overset{\delta-}{O}$ of carbonyl group with $\overset{\delta+}{H}$ of water molecules. Therefore, aldehydes up to four C′s, i.e., methanal, ethanal, propanal, and butanal, are soluble in water. Pentanal with five C′s is slightly soluble, and hexanal with six C′s is insoluble. Aldehydes, except for formaldehyde, generally smell pleasant and are used in perfumes.

Ketones

A ketone has two hydrocarbon groups bonded with the carbonyl carbon, i.e., $R-\overset{O}{\overset{||}{C}}-R'$ group. The condensed form of the aldehyde group is $RCOR'$.

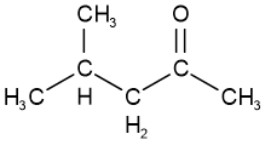
Nomenclature of ketones

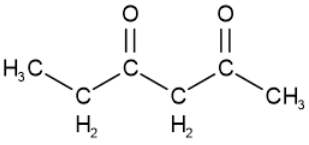
The IUPAC naming of ketones followed the rules of naming alcohols and aldehydes, summarized below.

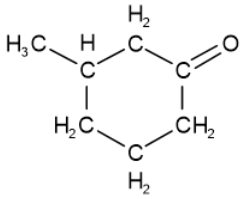
- The longest chain **containing the ketone group** is chosen as the parent name, with the last 'e' of the suffix replaced with -one.
  - The parent chain is numbered starting from the end, which gives the lower number to the ketone group, e.g., CH_3COCH_3 is propan-2-one.
  - For two ketone groups -dione and three -trione suffix is used.
  - For cyclic ketones, the parent name of cycloalkane is used with the last 'e' replaced with -one. Some examples are shown below.

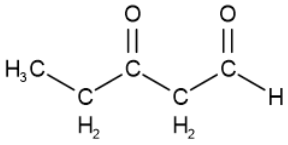
- If a group of higher precedence, e.g., an aldehyde, is also present, then ketone is represented by the prefix -oxo, as shown below.
- A benzene ring is represented by the prefix phenyl-, as shown below.

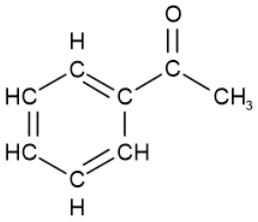
Some examples of the naming using the above rules are shown below.


4-methylpentane-2-one


hexane-2,4-dione

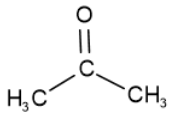

3-methylcyclohexan-1-one

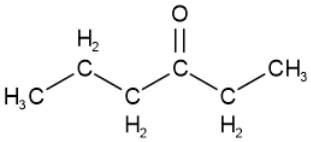

3-oxopentanal

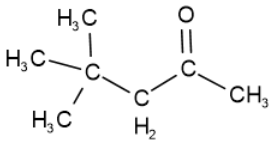

1-phenylethan-1-one

Common names of ketones

Propan-2-one is acetone, and 1-phenylethan-1-one is acetophenone. Common names of other ketones are obtained by listing the alkyl groups bonded to the carbonyl group in the order of group size, followed by the word ketone, as shown below with common names in brackets.

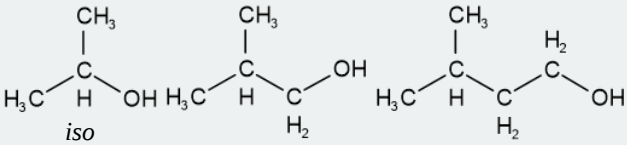

Propan-2-one (acetone)


hexan-3-one (ethyl propyl ketone)


4,4-dimethylpentane-2-one (methyl isobutyl ketone)

Iso-Group

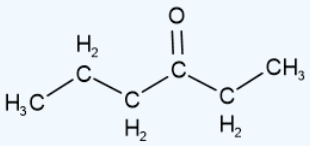
An alkyl group containing $-\mathrm{CH_3}$ attached to the 2nd last ceC takes iso- prefix, as shown in the examples below.


iso propananol — *iso* butanol — *iso* pentanol

The group names with iso-prefix are often used in common names and are also accepted in IUPAC nomenclature.

Example 2.4.1

What is the IUPAC name of this compound shown in the figure on the right?



Solution

the longest chain containing ketone groups is six C, i.e., hexane.

Change the final 'e' with -one, i.e., hexanone.

Add the location of the ketone group. Start numbering the parent chain from the end, giving the ketone a lower number. Ketone receives #3

Answer: hexan-3-one

Example 2.4.2

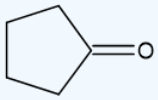
Write the skeletal formula of cyclopentanone.

Solution

The parent name is cyclopentane as shown in the figure on the right.

Add a ketone group to any carbon.

Answer:



Example 2.4.3

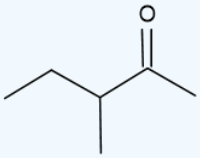
Write the skeletal formula of 3-methylpentane-2-one?

Solution

Parent name is pentane:

Count from either side and add a carbonyl group to C#2 and a methyl group to C#3.

Answer:



Physical properties of Ketones

Ketones have sp^2-hybridized C and O with a double bond (a σ and a π-bond) between them, two lone pairs occupying two sp^2-orbital of an O, and the C bonded with two hydrocarbon groups. i.e., $R-\overset{\overset{:O:}{||}}{C}-R'$. The C=O bond is polar, i.e., $\overset{\delta+}{C}=\overset{\delta-}{O}$ because O are more electronegative than C (3.3-2.1 = 0.9), as shown in Figure 2.4.2. It makes the C a δ^+, i.e., an electrophile in reactivity and the O a δ^-, i.e., a nucleophile or a base in reactivity. The lone pair of electrons on O add to the base character of the carbonyl O.

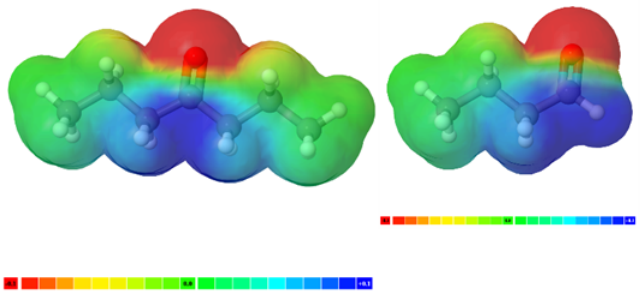


Figure 2.4.2: Electrostatic potential map of a ketone compared with an aldehyde. Haptan-4-one ($CH_3CH_2CH_2COCH_2CH_2CH_3$) and Butanal ($CH_3CH_2CH_2CHO$). C′s are gray, H′s are white, and O is red om the model. Blue region is δ^+, red is δ^-, and green is neutral. (Copyright; Public domain)

Comparison of carbonyl C of an aldehyde and a ketone

Figure 2.4.2 compares the electrostatic potential maps of an aldehyde and a ketone. The following are the points to note:

- The carbonyl C is more $\delta+$ in an aldehyde than in a ketone. The reason is that the alkyl group of ketone partially neutralizes the $\delta+$ of the carbonyl C by hyperconjugation while the H of the aldehyde can not. (Note: hyperconjugation is a special form of resonace which involves σ-bonds. The details of it are out of the scope of this book)
- The carbonyl C of an aldehyde is more accessible to reagents in chemical reactions than the carbonyl C of a ketone. This is because H in an aldehyde is small compared to the hydrocarbon group in a ketone.

These two factors make an aldehyde a more reactive electrophile than a ketone.

Other than the reactivity difference between an aldehyde and a ketone described above, the physical characteristics, i.e., the trend in the boiling points, solubility in water, etc., are the same for aldehydes and ketones.

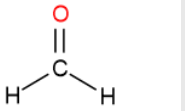
Some important aldehydes and ketones

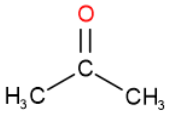
Aldehydes and ketones are often part of the biochemical processes. They are often an intermediate in the conversion of food into energy.

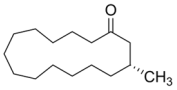
Methanal or **formaldehyde** is a colorless gas with pungent order that has germicidal properties. A 40% mixture of formaldehyde in water called formalin is used to preserve biological specimens. Formaldehyde is a starting material of polymers used in fabrics, insulation, carpeting, and other products.

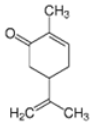
Propan-1-one, or **acetone,** is a colorless liquid with a mild odor used as a solvent in paints, nail polish removers, rubber cement, and cleaning fluids. Care must be taken in handling acetone as it is highly volatile and flammable.

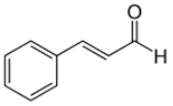
Several natural products are aldehydes or ketones used to flavor foods or as components of fragrances. For example, muscone is a ketone used in musk perfumes. Oil of spearmint contains carvone which is a ketone; cinnamaldehyde is found in cinnamon, almonds contain benzaldehyde, and vanillin is found in vanilla beans. The structures of these aldehydes and ketones are shown below.

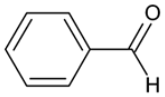

Formaldehyde

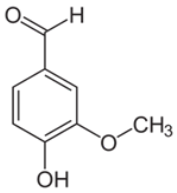

Acetone


Muscone


Carvone


Cinnamaldehyde


Benzaldehyde

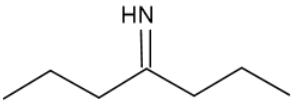

Vanillin

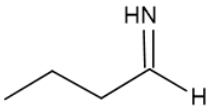
Imines

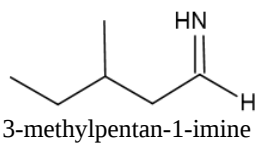
Replacing carbonyl O with a N in a carbonyl (C=O makes an iminie group, i.e., $R'-\overset{\overset{NR}{\|}}{C}-R''$. Note that N has three bonds and one lone pair. The third bond of N is with a H or a hydrocarbon (R) in this case. Imines are less common but important as reactive intermediates.

Nomenclature of imines

IUPAC rules for naming imines are the same as for the corresponding aldehydes or ketones, except that the -al or -one is replaced with the suffix -imine, as shown by the examples below.


heptan-4-imine


butan-1-imine


3-methylpentan-1-imine

Physical properties of imines

The C=N bond is polar, i.e., $\overset{\delta+}{\mathrm{C}}-\overset{\delta-}{\mathrm{N}}$ because N are more electronegative than C (3.0-2.5 = 0.5), as shown in Figure 2.4.4. It makes the C a δ^+, i.e., an electrophile in reactivity and the N a δ^-, i.e., a nucleophile or a base in reactivity. The lone pair of electrons on N add to the base character of the N in an imine group.

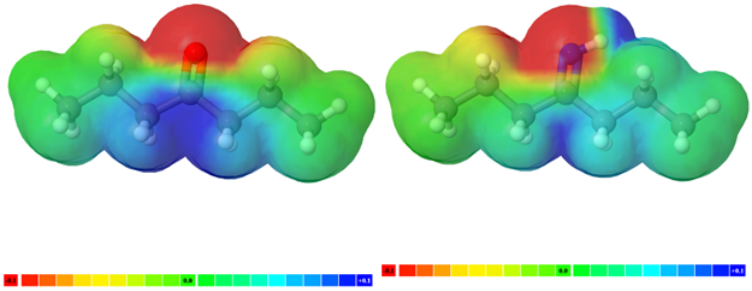


Figure 2.4.3: Electrostatic potential map of a ketone and imine are compared. Haptan-4-one ($CH_3CH_2CH_2COCH_2CH_2CH_3$), and haptan-4-imine ($CH_3CH_2CH_2CNHCH_2CH_2CH_3$). C′s are gray, H′s are white O is red and blue is N in the model. Blue region is δ^+, red is δ^-, and green is neutral. (Copyright; Public domain)

A comparison of the electrostatic potential map in Figure 2.4.3 shows that the C of an imine group is much less $\delta+$ than the corresponding carbonyl C (notice a less blue area in the case of imine compared to the corresponding ketone). This is because C=O bond is more polar (3.5-2.5 =1.0) than a C=N bond (3.0-2.5 = 0.5). Based on the less polar bond, one would expect imines to be more stable in chemical reactions than the corresponding aldehydes or ketones. Since N is less electronegative, it holds on to the lone pair less tightly than an O, which makes imine more basic, i.e., they easily donate their lone pair to a proton. Once the N makes the bond with a proton, it becomes positive charge species that is more reactive than the carbonyl compound it is derived from. Imines are less common but more important as reactive intermediates in converting aldehydes and ketones, particularly in biochemical systems.

2.5: Functional groups containing mix of sp3- and sp2-, or sp-hybridized heteroatom

Learning Objectives

- Identify, assign IUPAC name, and draw structure from the IUPAC name of carboxylic acids and their derivatives, including acid halides, acid anhydrides, esters, amides, and nitriles.
- Predict the changes in the polarity and its effect on the reactivity of carboxylic acids and their derivatives, including acid halides, acid anhydrides, esters, amides, and nitriles.
- Identify phosphoric acid, anhydrides of phosphoric acids, phosphate anions, and esters of phosphoric acids.

What are carboxylic acids and carboxylic acid derivatives?

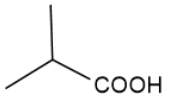
Carboxylic acids have a carbonyl group ($C{=}O$) and a hydroxyl group ($-OH$) on the same carbon, i.e., $-\overset{\overset{O}{\|}}{C}-OH$ group. The carboxylic acid group is represented as $-\overset{\overset{O}{\|}}{C}-OH$, or as $-COOH$. Carboxyl acids have some characteristics of the $C{=}O$ group, some characteristics of the $-OH$ group, and some additional characteristics due to the interaction of the two groups. In **carboxylic acid derivates**, the $-OH$ group is replaced with another group, that includes acid halides ($R-\overset{\overset{O}{\|}}{C}-X$), acid anhydrides ($R-\overset{\overset{O}{\|}}{C}-O-\overset{\overset{O}{\|}}{C}-R'$), easters ($R-\overset{\overset{O}{\|}}{C}-OR'$), and amides ($R-\overset{\overset{O}{\|}}{C}-NH_2$). Nitrile group that has the carbonyl O replaced with a N and the $-OH$ group also replaced with the same N, i.e., $R-C\equiv N$ group is also classified as a carboxylic acid derivative. The nomenclature and physical characteristics of carboxylic acids and their derivatives are described in the following sections.

Carboxylic acids ($R-\overset{\overset{O}{\|}}{C}-OH$)

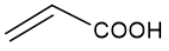
Nomenclature of carboxylic acids

The IUPAC nomenclature of the carboxylic acids follows the following rules.

- The longest hydrocarbon chain containing the carboxylic acid group is chosen as the parent name, with the last letter 'e' of its suffix replaced with -oic acid. For example, $HCOOH$ is methanoic acid, CH_3COOH is ethanoic acid, and CH_3CH_2COOH is propanoic acid.
  - If there are two carboxylic acid groups, the suffix changes to -dioic acid, e.g., $HOOC-COOH$ is ethanedioic acid, and $HOOC-CH_2-COOH$ is propanedioic acid. (note that when the suffix begins with a consonant (the letter 'd' in this case), the last letter 'e' of the parent hydrocarbon name is not dropped)
- Start numbering from the C of the $-COOH$ group. The $-COOH$ group itself does not need a location number, as it is always at the end of the chain.

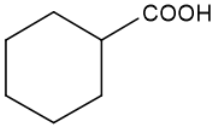


2-methylpropanoic acid

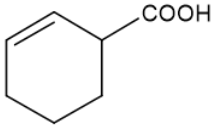


prop-2-enoic acid

- If the $-COOH$ group is bonded to a cyclic chain, the suffix -carboxylic acid is added to the name of the cyclic hydrocarbon. Numbering starts from the point of attachment of $-COOH$ to the ring.

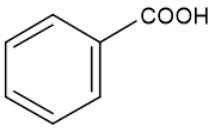


cyclohexanecarboxylic acid

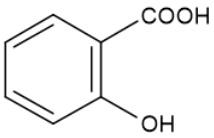


cyclohex-2-ene-1-carboxylic acid

- If the $-\mathrm{COOH}$ group is bonded to a benzene ring, the parent name "benzoic acid is used.
  - Numbering starts from the point of attachment of the $-\mathrm{COOH}$ group to the ring.



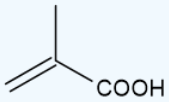
benzoic acid



2-hydroxybenzoic acid

Example 2.5.1

What is the IUPAC name of the compound shown on the right?



Solution

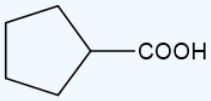
- The longest chain counting the $-\mathrm{COOH}$ group is three C's and a double bond, so the parent name is propene and replace the last letter 'e' with -oic acid, i.e., propenoic acid.
- There is a methyl group attached that becomes a prefix, i.e., methylpropenoic acid.
- A location number is needed for the methyl group and the double bond. Start numbering from the C of the $-\mathrm{COOH}$ group, double bond receives#2 and the methyl group receives #2.

Answer: 2-methylprop-2-enoic acid

The common name of 2-methylprop-2-enoic acid is methacrylic acid, and prop-2-enoic acid is acrylic acid, the monomers (the repeating units) in some polymers.

Example 2.5.2

What is the IUPAC name of the compound shown on the right?



Solution

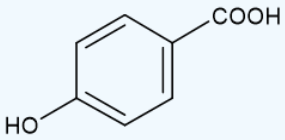
The $-\mathrm{COOH}$ group is attached to a five C cyclic chain, so the name of the cyclic chain becomes the parent name: cyclopentane.

Add the suffix -carboxylic acid to the parent name to indicate the carboxylic acid group attached to a cyclic chain.

Answer: cyclopentanecarboxylic acid

Example 2.5.3

What is the IUPAC name of the compound shown on the right?



Solution

- A carboxylic acid bonded to a benzene ring takes "benzoic acid" as the parent name.
- A $-\mathrm{OH}$ group takes the prefix "hydroxy" in the presence of a $-\mathrm{COOH}$ group, i.e., hydroxybenzoic acid.
- Start numbering from the point of attachment of the $-\mathrm{COOH}$ group: the $-\mathrm{OH}$ group receives #4.

Answer: 4-hydroxybenzoic acid.

Common names of carboxylic acids

Common names of carboxylic acids are derived from the names of natural sources of these acids. Table 1 lists some of the common names of carboxylic acids.

Table 1: Common names and the sources of the common names of some of the carboxylic acids.

Condensed formula	IUPACE name	Common name	Source of the common name
HCOOH	methanoic acid	formic acid	Latin: *formica*, ant

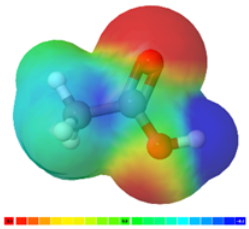
Condensed formula	IUPACE name	Common name	Source of the common name
CH_3COOH	ethanoic acid	acetic acid	Latin: *acetum*, vinegar
CH_3CH_2COOH	propanoic acid	propionic acid	Greek: *propion*, first fat
(\ce{CH3(CH2)2COOH}\)	butanoic acid	butyric acid	Latin: *butyrum*, butter
(\ce{CH3(CH2)4COOH}\)	hexanoic acid	caproic acid	Latin: *caper*, goat
(\ce{CH3(CH2)14COOH}\)	hexadecanoic acid	palmitic acid	Latin: *palma*, palm tree
(\ce{CH3(CH2)16COOH}\)	octadecanoic acid	stearic acid	Greek: *stear*, solid fat
(\ce{CH3(CH2)18COOH}\)	eicosanoic acid	arachidic acid	Greek: *arachis*, peanut

The first syllable of the common names, e.g., form-, acet-, prop-, etc., are also used as the first syllable of the common names of related compounds. For example, $HCOH$ is formaldehyde, CH_3COH is acetaldehyde, etc.

Physical properties of carboxylic acids

The carboxylic acid group has a $C{=}O$ and a $-OH$ groups, i.e., an sp^2- and an sp^3 hybridized O bonded to the same C. Both O's have two lone pairs of electrons on them, i.e., $-\overset{:O:}{\overset{\|}{C}}-\ddot{\underset{\cdot\cdot}{O}}-H$. Lone pair of electrons are usually not shown except when needed, i.e., the carboxylic acid group is represented as $-\overset{O}{\overset{\|}{C}}-OH$ or as $-COOH$.

The carboxylic acid group has three polar bonds, i.e., $\overset{\delta+}{C}{=}\overset{\delta-}{O}$, $\overset{\delta+}{C}-\overset{\delta-}{O}$ and $\overset{\delta-}{O}-\overset{\delta+}{H}$, resulting in a polar group: $-\overset{\delta-}{\overset{O}{\overset{\|}{\overset{\delta+}{C}}}}-\overset{\delta-}{O}-\overset{\delta+}{H}$. This is because O are more electronegative than C (3.3-2.6 = 0.7) and H (3.3-2.2 = 1.17), as shown in Figure 2.5.1. It makes $\overset{\delta+}{C}$ an electrophile, $\overset{\delta-}{O}$ a nucleophile or a base, and $\overset{\delta+}{H}$ an acid in reactivity. Due to the acid protons, carboxylic acids are also classified as organic acids.



$$CH_3-\overset{O}{\overset{\|}{C}}-OH$$

Figure 2.5.1: Electrostatic potential map of ethanoic acid $CH_3-\overset{O}{\overset{\|}{C}}-OH$, C's are gray, O's are red, and H's are wight in the model, red region is $\delta-$, blue is $\delta+$ and green is neutrla in the map. (Copyright; Public domain)

The polar $\overset{\delta+}{C}{=}\overset{\delta-}{O}$ and $\overset{\delta-}{O}-\overset{\delta+}{H}$ bonds allow dipole-dipole interactions and hydrogen bonding in addition to the London dispersion forces. Carboxylic acids have stronger intermolecular forces, higher melting points, higher boiling points, and higher solubilities in water compared to alcohols and aldehydes of comparable molar mass due to more intermolecular forces, as compared in Table 2.

Table 1: Compares boiling points and solubility of a carboxylic acid, alcohol, and aldehyde of comparable molar mass.

Condensed formula	IUPAC name	Molar mass (g/mol)	Boiling point (^oC)	Solubility in water
$CH_3(CH_2)_2COOH$	Butanoic acid	88.1	163	Miscible
$CH_3(CH_2)_3CH_2OH$	Pentan-1-ol	88.1	137	2.3 g/100 mL
$CH_3(CH_2)_3CHO$	Pentanal	86.1	103	Slightly soluble

Two carboxylic acids can make two hydrogen bonds with each other, as illustrated in Figure 2.5.2, behaving as a dimer with two times higher molecular mass. It explains their higher boiling points than alcohols of the same molar mass. Carboxylic acids of up to five C′s, i.e., methanoic acid, ethanoic acid, propanoic acid, butanoic acid, and pentanoic acid, are soluble in water. Hexanoic acid is slightly soluble, and higher acids are insoluble.

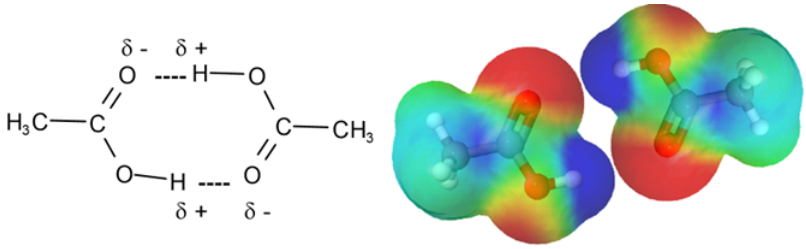
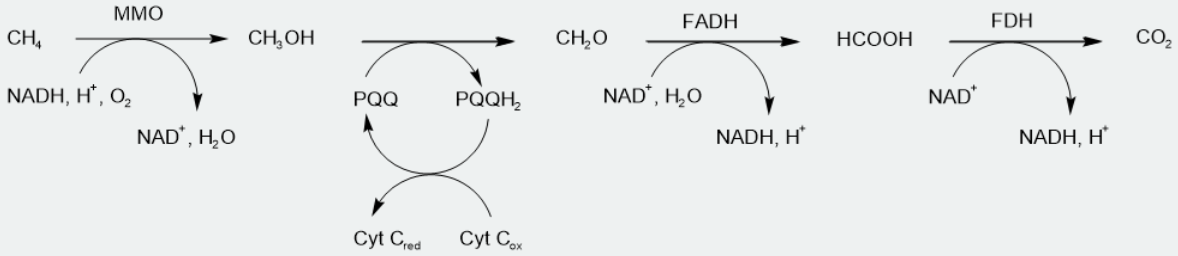


Figure 2.5.2: Illustration of hydrogen bonding and dimer formation using acetic acid as a model carboxylic acid compound. (left: molecular formula showing hydrogen bonds as dotted lines, right: electrostatic potential map showing attraction between oppositely charged regions of two molecules, i.e., hydrogen bonding in this case (Copyright; Public domain)

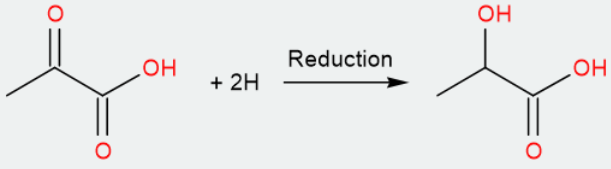
Carboxylic acids have a sour taste because they are acids due to ionizable proton in their $-O-H$ groups. For example, the sour taste of citrus fruits is due to citric acid, and the sour taste of vinegar is due to ethanoic acid.

Oxidation is i) loss of electrons, ii) gain of O, or loss of H; and the **reduction** is the opposite of these. Oxidation and reduction happen together and are collectively called **Redox** reactions. Most chemical reactions in biological systems are redox reactions, e.g., photosynthesis is a reduction of CO_2 to convert solar energy into potential chemical energy, and digestion of food is the opposite, i.e., oxidation to release the energy for the activities of life.

A major portion of organic compounds is hydrocarbon groups, gradually oxidized to alcohols, aldehydes or ketones, carboxylic acids, and finally, carbon dioxide and water, releasing energy. For example, methane (CH_4) oxidizes to methanol (CH_3OH), methanal (CH_2O), methanoic acid ($HCOOH$), and finally to carbon dioxide (CO_2) that is exhaled, as shown below.

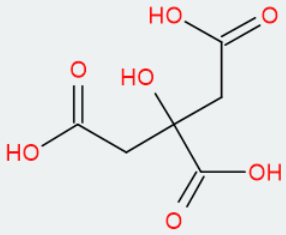


The alcohols, aldehydes, ketones, and carboxylic acids also serve as intermediates for synthesizing compounds the body needs. Carboxylic acids commonly appear in the metabolic process. For example, glucose, a six C compound, is first converted to two pyruvic acid molecules. Under low oxygen conditions (anaerobic), pyruvic acid is reduced to lactic acid.

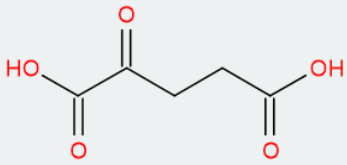


Pyruvic acid Lactic acid

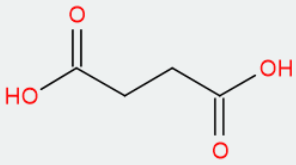
In the presence of oxygen (aerobic), pyruvic acid releases a CO_2 and becomes a two C group that joins a four C compound oxalic acid to make a six C compound citric acid. Citric acid releases a CO_2 and becomes a five C compound α-ketoglutaric acid, which releases another CO_2 and becomes four C compound succinic acid, as shown below. This process goes on through several intermediate carboxylic acids, and either all the C′s of the starting compound convert to CO_2, or the intermediate is utilized to synthesize compounds needed by the body.



Citric acid



, α-Ketoglutaric acid



, and Succinic acid

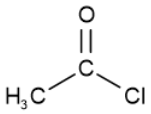
Note: The carboxylic acids are shown as neutral acids in the above example, but in the physiological conditions, they exist as anions, i.e., as a carboxylate group ($-\overset{\overset{O}{\|}}{C}-O^-$). These points will be discussed in a later chapter.

Acid halides and acid anhydrides -the most reactive acid derivatives

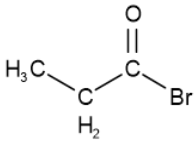
Nomenclature of acid halides

Acid halides contain the ($-\overset{\overset{O}{\|}}{C}-X$) group, where X can be F, Cl, Br, or I.

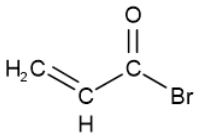
- IUPAC name of the acid halide takes the name of the corresponding carboxylic acid with the suffix -oic acid replaced with -oil halide, as shown in the following examples.



ethanoyl chloride or acetyl chloride



propanoyl bromide



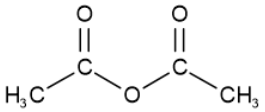
prop-2-enoyl bromide

Acid chlorides and acid bromides are the most common.

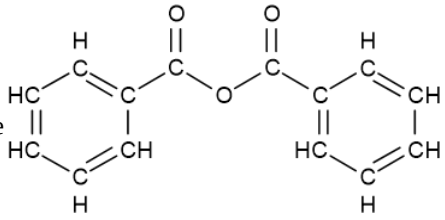
Nomenclature of acid anhydrides

Acid anhydride contains two acyl $R-\overset{\overset{O}{\|}}{C}-$ groups bonded to a common O atom, i.e., ($-\overset{\overset{O}{\|}}{C}-O-\overset{\overset{O}{\|}}{C}-$) group. A carboxylic acid anhydride is derived by condensing two carboxylic acids by losing a H_2O molecule. The acid anhydride is **symmetric anhydride** if both the acids are the same and mixed or **asymmetric anhydride** if two different acids are condensed to form the anhydride,

- Symmetric acid anhydrides are named using the name of the corresponding acid with the last word 'acid' replaced with 'anhydride', as shown in the following examples.

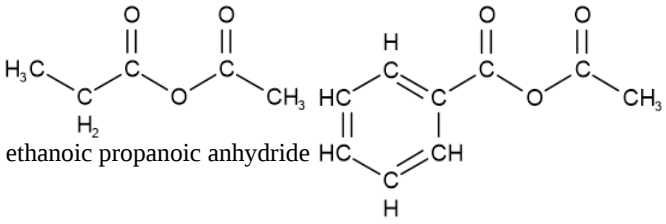


ethanoic anhydride or acetic anhydride



benzoic anhydride

- Mixed or asymmetric anhydrides are named by listing the names of the two acids in alphabetic order without the last word 'acid', followed by the word 'anhydride', as shown in the following examples.



ethanoic propanoic anhydride

accetic benzoic anhydride

Physical properties of acid halides and acid anhydrides

The acid halides group has two polar bonds, i.e., $\overset{\delta+}{\mathrm{C}}{=}\overset{\delta-}{\mathrm{O}}$, and $\overset{\delta+}{\mathrm{C}}{-}\overset{\delta-}{\mathrm{X}}$ resulting in a polar group: $-\overset{\delta+}{\mathrm{C}}(=\overset{\delta-}{\mathrm{O}})-\overset{\delta-}{\mathrm{X}}$. The acid anhydrides have four polar bonds i.e., two $\overset{\delta+}{\mathrm{C}}{=}\overset{\delta-}{\mathrm{O}}$, and two $\overset{\delta+}{\mathrm{C}}{-}\overset{\delta-}{\mathrm{O}}$ resulting in a polar group: $-\overset{\delta+}{\mathrm{C}}(=\overset{\delta-}{\mathrm{O}})-\overset{\delta-}{\mathrm{O}}-\overset{\delta+}{\mathrm{C}}(=\overset{\delta-}{\mathrm{O}})-$. This polarity of the group can be observed in the electrostatic potential maps of acid halide, acid anhydride, and an acid shown in Figure $2.5.3$

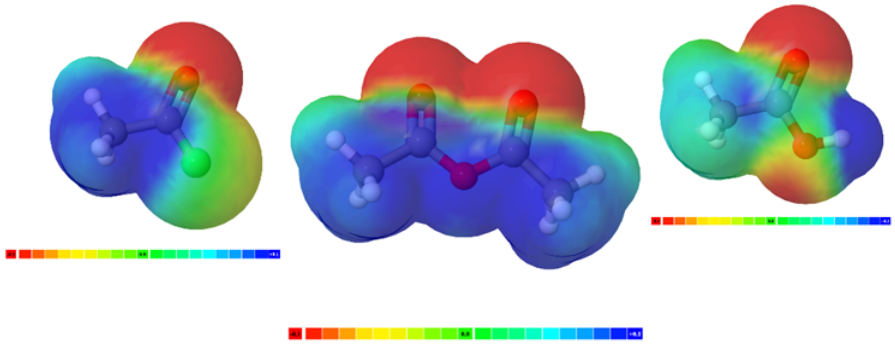


Figure $2.5.3$: Electrostatic potential maps of acetyl chloride ($\mathrm{CH_3COCl}$), acetic anhydride ($\mathrm{CH_3COOCOCH_3}$), and acetic acid ($\mathrm{CH_3COOH}$). C's are gray, H's are white, O's are red, and Cl is green in the model. Blue area is $\delta+$, read is $\delta-$, and green is neutral in the map. (Copyright: Public domain)

It is apparent from the comparison of the electrostatic potential maps shown in Figure $2.5.3$ that the carbonyl C is more bluish, i.e., higher $\delta+$ and stronger nucleophile, in the case of acid halide and acid anhydride than in carboxylic acid. The question is Cl in the acid halide is less electronegative than O in carboxylic acid, then why is the carbonyl C more $\delta+$ in the acid halide than in the acid? The answer is in the fact that the heteroatom not only draws the bonding electron away from the carbonyl C, it also donates its lone pair through resonance that diminishes the $\delta+$ character on the carbonyl C:

- The 2p-orbital of O and 2p-orbital of C of carboxylic acid are of similar size and overlap well for the reasonce to happen and diminish its $\delta+$ character.
- The 3p-orbital of Cl or 4p orbital of Br overlaps poorly with 2p-orbital of carbonyl C due to the size difference and does not diminish its $\delta+$ character.
- The (\ce{O}\) is shared between to $\mathrm{C}{=}\mathrm{O}$ groups in the case of acid anhydride. So it diminish $\delta+$ character of $\mathrm{C}{=}\mathrm{O}$ less than in carboxylic acids.

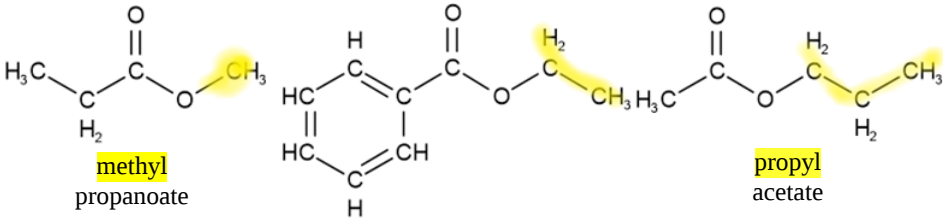
Due to the highly nucleophilic carbonyl C, the acid halides and acid anhydrides are very reactive and primarily used as reactive intermediates in chemical synthesis. Again due to their high reactivity, they can not survive in biological systems. Biochemical systems use carboxylic acid derivatives containing S or phosphate groups as reactive intermediates, which will be described later.

Easters

Esters have an acyl group (($\mathrm{R}-\overset{\mathrm{O}}{\overset{\|}{\mathrm{C}}}-$) bonded with an alkoxy group (easters ($-\mathrm{OR}'$), i.e., an easter ($\mathrm{R}-\overset{\mathrm{O}}{\overset{\|}{\mathrm{C}}}-\mathrm{OR}'$) group.

Nomenclature of esters

- IUPAC name of an ester starts with the name of the alkyl group that is part of the alkoxy ($-\mathrm{OR}'$) group, followed by the name of the acid corresponding to the acyl ($\mathrm{R}-\overset{\overset{\mathrm{O}}{\|}}{\mathrm{C}}-$) group with the suffix -oic acid replaced with the suffix -oate, as shown in the following examples.



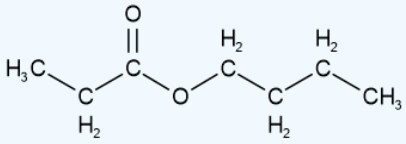
methyl propanoate

ethyl benzoate

propyl acetate

Example 2.5.1

What is the IUPAC name of the compound shown on the right?



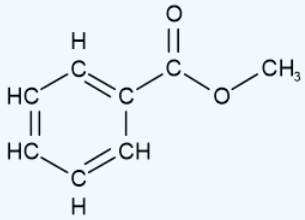
Solution

- Alkyl group in the alkoxy ($-\mathrm{OR}'$) group is four (C's) i.e., butyl.
- The acyl group is three (C's), i.e., propanoic acid; replace acid with -oate, i.e., propanoate.

Answer: butyl propanoate

Example 2.5.2

What is the IUPAC of the compound shown on the right?



Solution

- It is an easter where the alkyl group in the alkoxy ($-\mathrm{OR}'$) group is one (C's), i.e., methyl.
- The acyl group contains benzene, i.e., benzoic acid, and replace the -ic acid with -oate, i.e., benzoate.

Answer: methyl benzoate

Physical properties of esters

Esters group has a polar bonds, i.e., $\overset{\delta+}{\mathrm{C}}=\overset{\delta-}{\mathrm{O}}$, and two polar $\overset{\delta+}{\mathrm{C}}-\overset{\delta-}{\mathrm{O}}$ groups, resulting in a polar group: $-\overset{\overset{\delta-}{\mathrm{O}}}{\overset{\delta+}{\mathrm{C}}\|}-\overset{\delta-}{\mathrm{O}}-\mathrm{R}'$, as shown in Figure 2.5.4. The reactivity of an ester's carbonyl C is almost the same as that of a carboxylic acid.

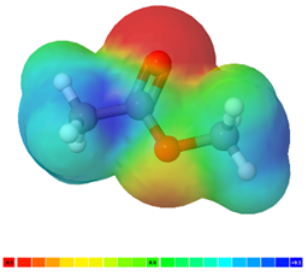
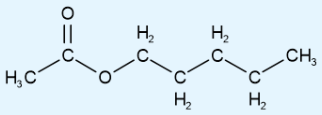
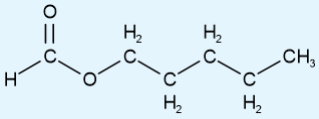
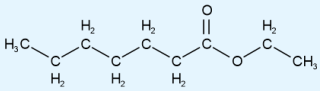
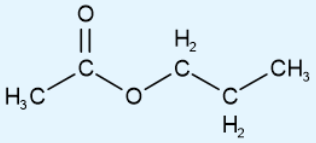
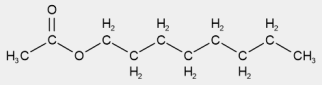
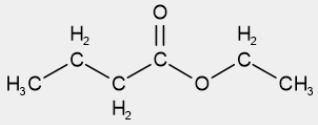
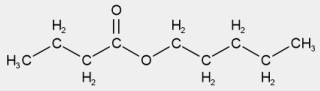
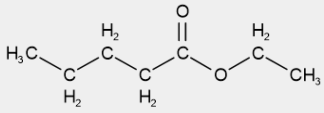
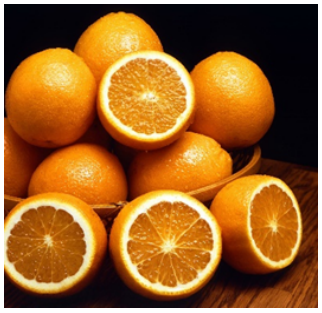


Figure 2.5.4: Electrostatic potential maps of methyl acetate ($\mathrm{CH_3COOCH_3}$),. C's are gray, H's are white, and O's are red in the model. Blue area is $\delta+$, read is $\delta-$, and green is neutral in the map. (Copyright: Public domain)

Easters are pretty common in nature. Small esters are volatile and soluble in water, making them easier to smell and taste. The fragrances of many perfumes and flavors of several fruits are due to esters, as shown in Table 3.

Table 3: Some esters and the flavor/taste of the fruit they are associated with are shown in the next row.

 pentyl acetate	 pentyl methanoate	 ethyl heptanoate	 propyl acetate
 Banana	 Plum	 Grape	 Pear
 octyl acetate	 ethyl butanoate	 pentyl butanoate	 propyl pentanoate
 Orange	 Pineapple	 Apricot	 Apple

Esters in fats

Fats are esters of carboxylic acids that contain a long chain hydrocarbon, an alkane or alkene with cis double bonds, called fatty acids, and propane-1,2,3-triol also called glycerol, as shown in one example in Figure 2.5.5.

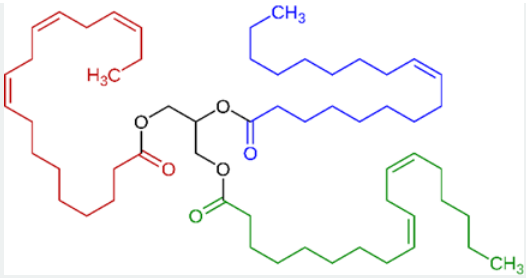
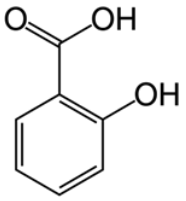
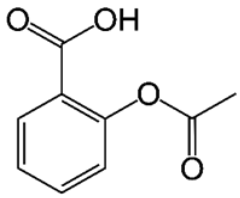
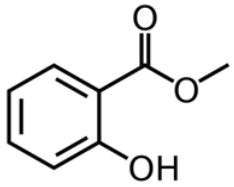


Figure 2.5.5: Representative triglyceride found in linseed oil, a triester (triglyceride) derived from linoleic acid (green), α-linoleic acid (red), and oleic acid (blue). (Copyright;Neutroic, Public domain, via Wikimedia Commons)

Aspirin -an ester in medical use

Aspirin is an ester of salicylic acid found in a willow tree's bark. Salicylic acid reduces pain and fever, but it irritates the stomach lining. Aspirin, an ester of salicylic acid, overcomes this problem and is commonly used to reduce pain and fever and as an anti-inflammatory agent. Methyl salicylate is another ester of salicylic acid found in wintergreen oil and used as skin ointments to soothe sore muscles.

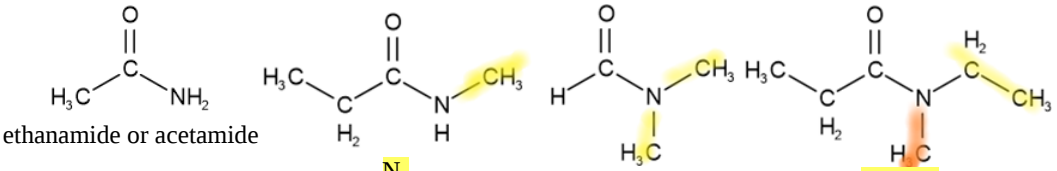
		
Salicylic acid	Aspirin	Methyl salicylate

Amides

Amidess have an acyl group ($(\mathrm{R-\overset{\overset{\mathrm{O}}{\|}}{C}-})$) bonded with a nitrogen group ($-\mathrm{NRR'}$), i.e., an amide ($-\overset{\overset{\mathrm{O}}{\|}}{\mathrm{C}}-\mathrm{NRR'}$) group, where R and $\mathrm{R'}$ may be H or a hydrocarbon group.

Nomenclature of amides

- IUPAC name of an amide is the name of the acid corresponding to the acyl ($(\mathrm{R-\overset{\overset{\mathrm{O}}{\|}}{C}-})$) group with the suffix -oic acid replaced with the word amide.
  - If a hydrocarbon group is bonded to nitrogen, its name appears as a prefix preceded by N-.
  - If two identical hydrocarbon groups are bonded to nitrogen, the group name is preceded by N, N-di.
  - If two different hydrocarbon groups are bonded to nitrogen, the group names, each preceded by N-, are listed alphabetically as prefixes. Some examples are shown below.



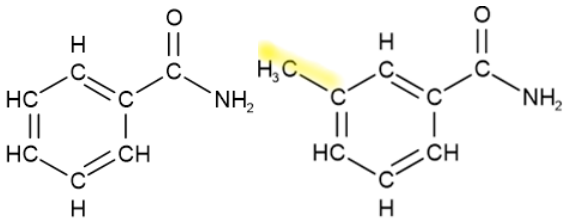
ethanamide or acetamide

N-methyl propanamide

N,N-dimethyl methanamide or N,N-dimethyl formaide

N-ethyl - N-methyl propanamide

- An amide group attached to a benzene ring takes the base name 'benzamide,' as shown in the following examples.
-

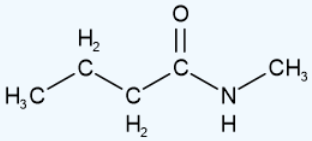


benzamide

3-methyl benzamide

Example 2.5.1

What is the IUPAC name of the compound shown on the right?



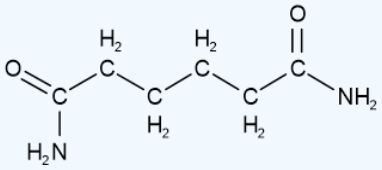
Solution

- There is an amide group on a four C chain, i.e., butanoic acid, which changes to butanamide.
- There is a methyl group on nitrogen that becomes the prefix N-methyl.

Answer: N-methylbutanamide

Example 2.5.2

What is the IUPAC name of the compound shown on the right?



Solution

- There are two amide groups, so it takes the suffix -diamide.
- The corresponding acid is a six C diacid named hexanedioic acid. Replace -oic acid with amide.

Answer: hexanediamid

Physical properties of amides

Amide group has a polar bonds, i.e., $\overset{\delta+}{\mathrm{C}}=\overset{\delta-}{\mathrm{O}}$, $\overset{\delta+}{\mathrm{C}}-\overset{\delta-}{\mathrm{N}}$, and $\overset{\delta-}{\mathrm{N}}-\overset{\delta+}{\mathrm{H}}$ groups, resulting in a polar group: $-\overset{\delta+}{\mathrm{C}}(=\overset{\delta-}{\mathrm{O}})-\overset{\delta-}{\mathrm{N}}-\mathrm{R}'\mathrm{R}''$, as shown in Figure 2.5.5.

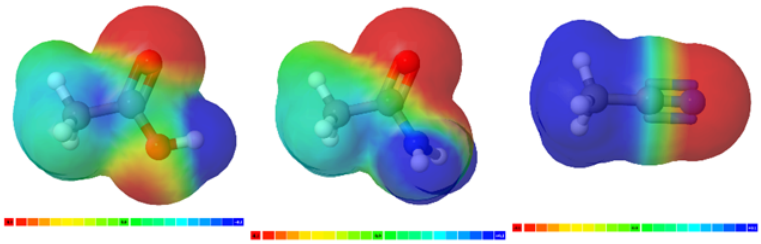


Figure 2.5.5: Electrostatic potential maps of acetic acid (CH_3COOH), acetamide (CH_3CONH_2), and acetonitrile ($CH_3C \equiv N$). C′s are gray, H′s are white, O′s are red, and Cl is green in the model. Blue area is $\delta+$, read is $\delta-$, and green is neutral in the map. (Copyright: Public domain)

Comparison of electrostatic potential maps in Figure 2.5.5 shows that the carbonyl C of amide is less $\delta+$, i.e., less electrophilic than that of the corresponding carboxylic acid. This is because of two reasons i) N is less electronegative and draws electrons away less than O and ii) being less electronegative, N sends its lone pair of electrons more to the carbonyl C neutralizing its $\delta+$ by resonance than O. Both of these factors make the carbonyl C less $\delta+$ and less electrophilic, which makes amides one of the lest reactive and the most stable carboxylic acid derivatives that are found commonly in nature. The resonance effect is illustrated in Figure 2.5.6 below. Due to the resonance, the C−N bond has a significant double bond character.

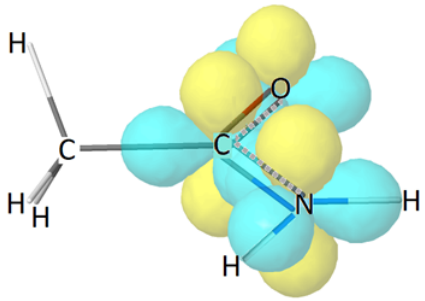


Figure 2.5.6: Illustration of the resonance shown by the dotted lines showing partial double bonds formed by the overlap of three consecutive p-orbitals (yellow lobes) on O, carbonyl C, and N. Recall that the N when it is next to an sp^2 hybridized atom converts from sp^3 to sp^2 hybridization for the resonance to happen. The blue lobs represent the sp^2 orbitals (Copyright; Public domain)

Since the lone pair of electrons of amide are occupied in the resonance, they are less available to protons of acids. Therefore, amides are much less basic than amines. Amides have $\overset{\delta-}{N}-\overset{\delta+}{H}$ bonds that allow them to make hydrogen bonds with water molecules. Therefore, amides containing up to five C′s are soluble in water due the hydrogen bonding. Those with larger alkyl groups, i.e., with more than five C′s are slightly soluble or insoluble due to the hydrophobic character of the alkyl groups dominating over the hydrophilic nature of the amide group.

Amides in nature and medicines

Proteins are made of small repeat units, called amino acids, joined through amide groups, as illustrated in Figure 2.5.7, which makes amide one of the most common groups present in nature.

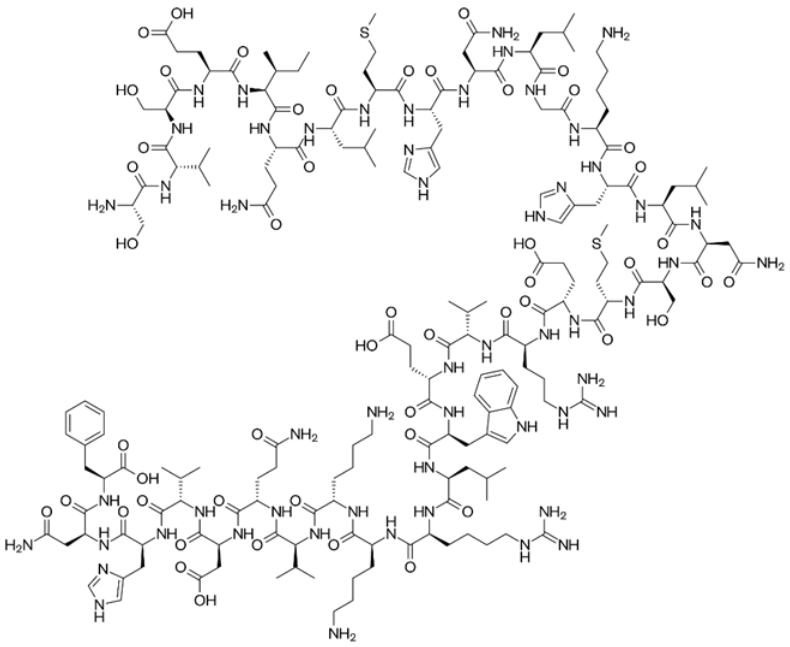
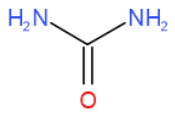
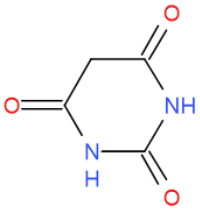
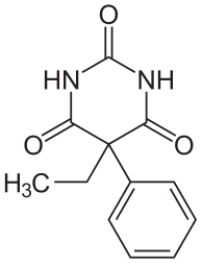
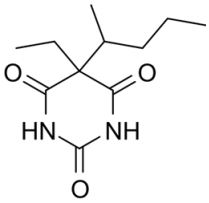
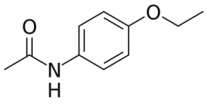
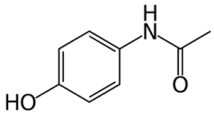


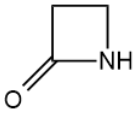
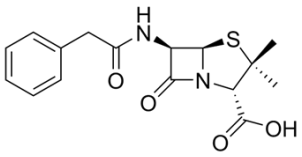
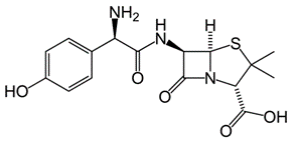
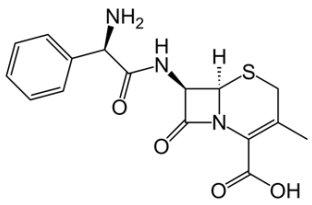
Figure 2.5.7: Skeletal formula of teriparatide—the biologically active region of the 84-amino acid human parathyroid hormone. (Copyright; Vaccinationist, Public domain via Wikimedia commons)

During the digestion of proteins, N′s end up in urea excreted by kidneys. If kidneys malfunction, urea may build to a toxic level, resulting in uremia. Urea is also used as a fertilizer.

Barbiturates derived from barbituric acid have sedative and hypnotic effects and are used in medicines for these effects. Barbiturate drugs include phenobarbital and pentobarbital. Phenacetin and acetaminophen are amides used in Tylenol as alternatives to reduce fever and pain but with little anti-inflammatory effect. The structures of these amides are shown below.

					
urea	barbituric acid	phenobarbital	pentobarbital	phenacetin	acetaminophen

cyclic amides are called lactams. A four-member lactam is a common feature in the structure of Penicillin and related synthetic antibiotics, as shown below.

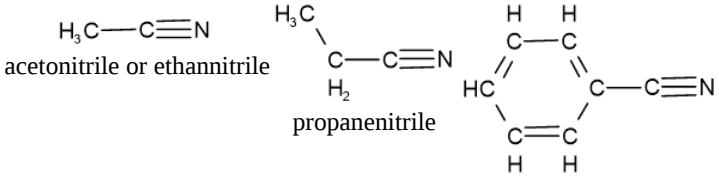
			
A four-membered lactam	Penicillin G	Amoxicillin	Cephalexin

Nitriles

Nitriles have cyano group that is a polar bond $-\overset{\delta+}{\mathrm{C}}\equiv\overset{\delta-}{\mathrm{N}}:$ with a lone pair in one of the sp orbital of N, as illustrated in Figure 2.5.5.

IUPAC name of nitriles is composed of the name of the hydrocarbon skeleton, including the C in the nitrile group with the suffix -nitrile.

Another way of naming them is to take the name of the corresponding carboxylic acid but replace the suffix -oic acid with -onitrile, as shown in the following example.



acetonitrile or ethannitrile

propanenitrile

benzonitrile

Nitriles are not very common in nature but are important as intermediates in synthesizing organic compounds.

Phosphorous groups

Phosphorous (P) is in the same group with N in periodic table. Like N in ammonia ($\mathrm{H-\ddot{N}-H}$, with H), the P can have three bonds and a lone pair, as in phosphine ($\mathrm{H-\ddot{P}-H}$, with H) with eight valence electrons, i.e., octet complete. Unlike N, the P is in fourth row in the periodic table and, like other elements of the fourth and higher row, can have more than eight valence electrons in its compounds, e.g., ten valence electrons in phosphoric acid ($\mathrm{HO-\overset{\overset{O}{\|}}{P}-OH}$, with OH). Phosphorous groups are important in biological systems, e.g., they are part of DNA molecules, phospholipids in cell membranes, and energetic molecules like adenosine triphosphate that are used as energy currency in biochemical reactions.

Phosphoric acid and phosphoric anhydrides

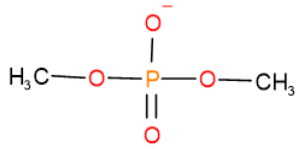
Phosphoric acid or orthphosphosporic acid has three $-\mathrm{OH}$ groups and one $=\mathrm{O}$ bonded to a P atom, i.e., $\mathrm{HO-\overset{\overset{O}{\|}}{P}-OH}$, with OH. Like carboxylic acid anhydride, which is two carboxylic acids joined through a common O, two phosphoric acids joined through a common O is a diphosphoric acid or a pyrophosphoric acid, and three phosphoric acids condensed in this way is a triphosphoric acid. Their corresponding anions formed after ionizing the acidic protons from the $-\mathrm{OH}$ groups are called phosphates, and alkoxy ($-\mathrm{OR}$ replacing one or more $-\mathrm{OH}$ groups are phosphate esters, as shown below.

$\mathrm{HO-\overset{\overset{O}{\|}}{P}-OH}$, with OH Phosphoric acid or orthphosphosporic acid	$\mathrm{HO-\overset{\overset{O}{\|}}{P}-O-\overset{\overset{O}{\|}}{P}-OH}$, with OH diphosphoric acid or a pyrophosphoric acid	$\mathrm{HO-\overset{\overset{O}{\|}}{P}-O-\overset{\overset{O}{\|}}{P}-O-\overset{\overset{O}{\|}}{P}-OH}$, with OH triphosphoric acid
$\mathrm{^{-}O-\overset{\overset{O}{\|}}{P}-O^{-}}$, with O^{-} Phosphate ion or orthphosphate ion	$\mathrm{^{-}O-\overset{\overset{O}{\|}}{P}-O-\overset{\overset{O}{\|}}{P}-O^{-}}$, with O^{-} diphosphate ion or a pyrophosphate ion	$\mathrm{^{-}O-\overset{\overset{O}{\|}}{P}-O-\overset{\overset{O}{\|}}{P}-O-\overset{\overset{O}{\|}}{P}-O^{-}}$, with O^{-} triphosphate ion

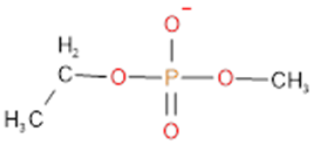
Phosphoric esters

Like carboxylic acid changes to an ester when its $-OH$ group is replaced with an alkoxy ($-OR$) group, phosphoric acid changes to mono phosphoric ester when one of its $-OH$ groups are replaced with alkoxy ($-OR$), to diester when two $-OH$ groups are replaced with alkoxy ($-OR$), and to triester when three $-OH$ groups are replaced with alkoxy ($-OR$).

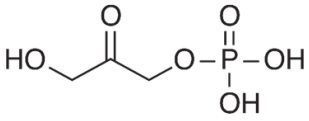
- The phosphoric esters are named by listing the names of the alkyl parts of the alkoxy groups in alphabetic order, followed by the world phosphate.
- In more complex phosphodiesters, it is common practice to name the organic compound followed by the word 'phosphate' or prefixed phospho-, as shown in the following examples.



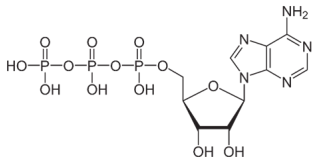
dimethyl phosphate



ethyl methyl phosphate



dihydroxyacetone phosphate



adenosine triphosphate

CHAPTER OVERVIEW

3: Stereochemistry

3.1: Introduction to stereochemistry

Learning Objectives

- Identify isomers and differentiate constitutional isomers from sterioisomes.
- Identify subclasses of constitutional isomers, including skeletal isomers, functional group isomers, and positional isomers; and subclasses of sterioisomers, including conformers, enantiomers, diasteriomers.

What is stereochemistry?

Stereochemistry is the study of the relative arrangement of atoms in molecules and their manipulation. A major area of stereochemistry is the study of isomers which is introduced below.

Isomers

Isomers are species with the same atoms in the same numbers, i.e., the same molecular formula but are arranged differently in space.

Several types of isomers can be classified in different ways. Figure 3.1.1 presents one scheme of classifying isomers.

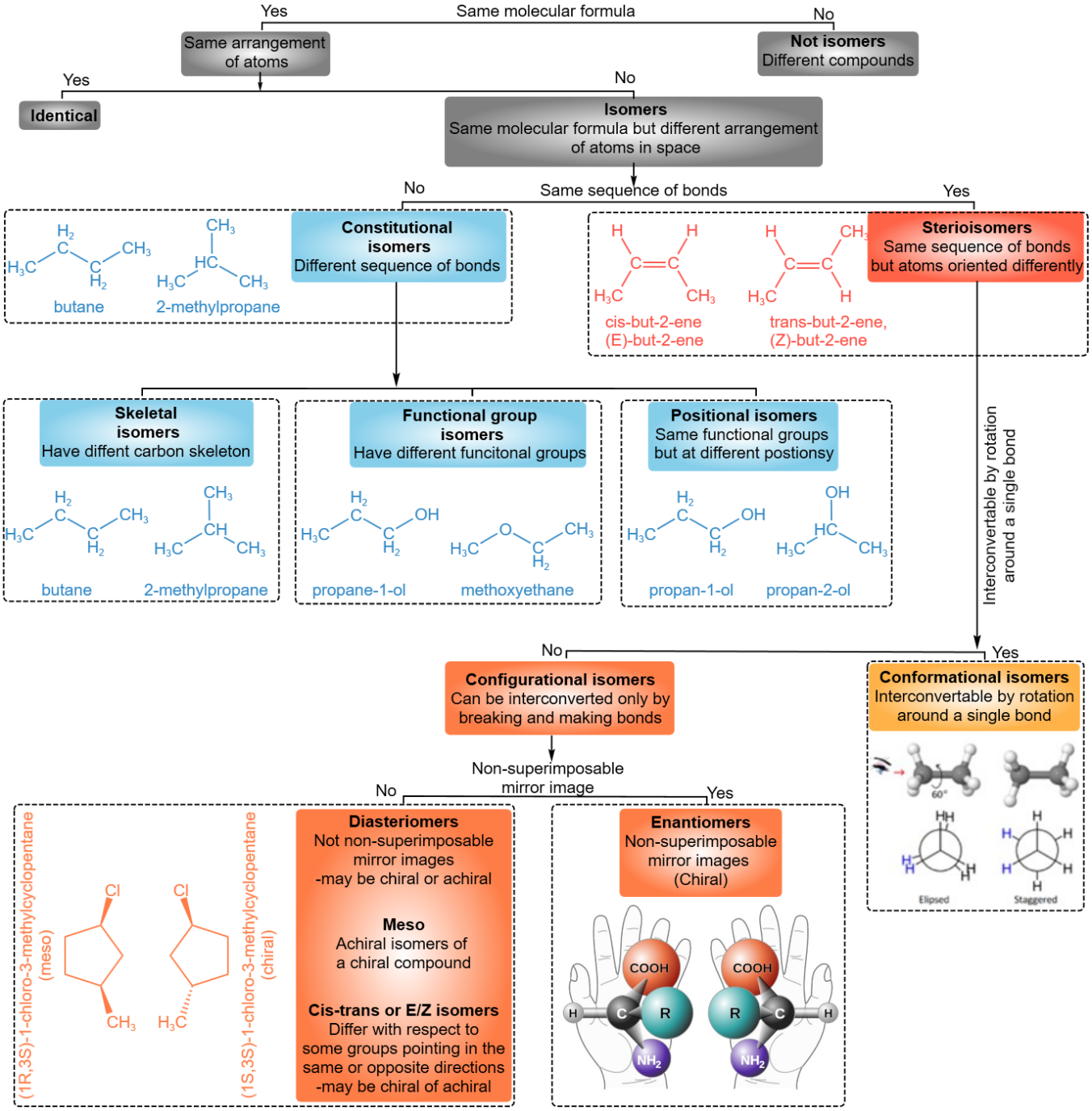


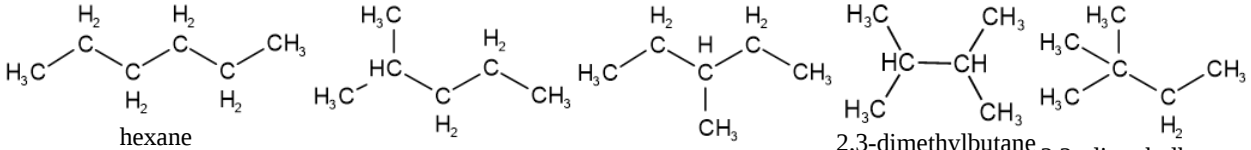
Figure 3.1.1: Classification of isomers. (Copyright; Public domain)

Constitutional isomers

Constitutional isomers have a different sequence of bonds that may result in another skeleton, i.e., **skeletal isomers**, different reactive features, i.e., **functional groups isomers**, or the same reactive features but placed at different locations, i.e., **positional isomers**.

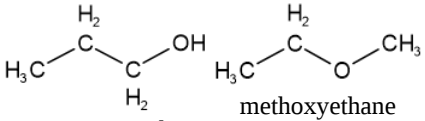
Skeletal isomers

Skeletal isomers have different carbon skeletons. For example, hexane, 2-methylpentane, 3-methylpentane, and 2,2-dimethylbutane are skeletal isomers having the same molecular formula C_6H_{14}.



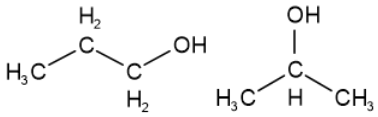
Functional group isomers

Functional group isomers have different reactive features, i.e., different functional groups. For example, propan-1-ol is alcohol, and methoxyethane is ether, but they are isomers having the same molecular formula C_3H_8O.



Positional isomers

Positional isomers have the same reactive features, i.e., the same functional groups, but the groups are attached at different positions. For example, propan-1-ol -a primary alcohol having an alcohol group linked to terminal C and propan-2-ol -a secondary alcohol having an alcohol group attached to a non-terminal C, are isomers having the same molecule formula C_3H_8O.

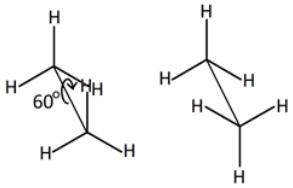


Stereoisomers

- **Stereoisomers** have the same sequence of bonds but the atoms or groups of atoms are oriented differently in space.
  - **Conformers** or **conformational isomers** are the stereoisomers in which the different orientations of atoms are a result of rotation around single bonds. The specific arrangement of atoms in a conformational isomer is also called **conformation**.
  - **Configurational isomers** are stereoisomers that can be interconnected only by breaking and making some bonds. The specific arrangement of atoms in a configuration isomer is also called **configuration**.

Conformers or conformational isomers

Conformers are the isomers that are the result of rotation around single bonds. They are also called different **conformations** of the same molecule. For example, rotation around $C-C$ bond of ethane places a set of three H's on one C at different positions relative to H's on the other C, as illustrated below.



Different arrangements of H's in ethane due to rotation around $C-C$ bond

Illustration of rotation around $C-C$ bond in ethane (Copyright; mailto:ralf@ark.in-berlin.de, CC BY 2.5 via Wikimedia Commons)

Usually, the rotation around single bonds happens rapidly at room temperature. So, the conformers usually exist as a mixture and can not be easily separated.

Configurational isomers

Configurational isomers are stereoisomers that can be interconnected only by breaking and making some bonds. For example, cis-but-2-ene and trans-but-2-ene are configurational isomers in which $-CH_3$ groups connected to two C's of a double bond are oriented differently as shown in Figure 3.1.1. Another example is L-glyceraldehyde and D-glyceraldehyde in which four different groups are connected to the same C but oriented differently as shown in Figure 3.1.1.

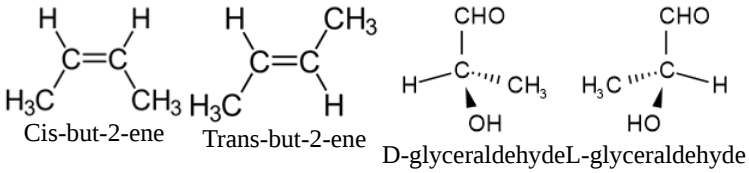


Figure 3.1.1: Examples of configurational isomers: cis- and trans-but-2-ene, and D- and L-glyceraldehyde. (Copyright; Public domain).

The configurational isomers are subdivided into enantiomers and diastereomers.

Enantiomers

A pair of stereoisomers that are related to each other as non-superimposable mirror images are **Enantiomers.**

For example, D-glyceraldehyde and L-glyceraldehyde shown above are enantiomers of each other. Imagine there is a mirror between D- and L-glyceraldehyde shown above. You will notice that they are mirror images of each other. If you try to overlap L-glyceraldehyde onto D-glyceraldehyde, two groups may overlap, but the other two will not overlap, no matter how you may rotate the molecule. D- and L- represent two different configurations of glyceraldehyde shown in Figure 3.1.1. Enantiomers are chiral molecules.

Chirality

An object or molecule that cannot be superimposed on its mirror image by any translation, rotational, or conformational changes is a **chiral** object. This geometric property is called **chirality**.

Achiral is not chiral, i.e., the objects or molecules that are identical to their mirror image are achiral.

For example, an amino acid with four different groups attached to the same carbon and hand are chiral, having a non-superimposable mirror image as illustrated in Figure 3.1.2. Like left and right hands that have a thumb and fingers in the same order, but are mirror images and not the same, chiral molecules have the same things attached in the same order, but are mirror images and not the same.

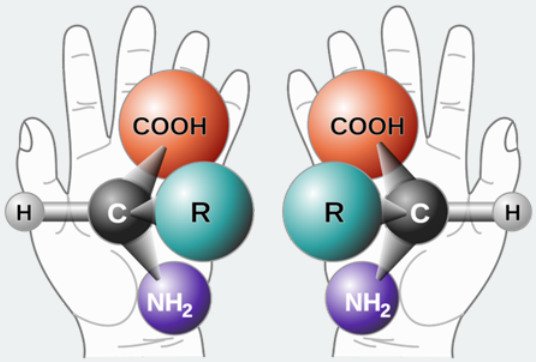


Figure 3.1.2: Illustration of chiral objects and their non-superimposable mirror images -a hand and an amino acid with four different groups attached to the same C. (Copyright; περήλιο, Public domain, via Wikimedia Commons)

Diastereomers

Stereoisomers that are not enantiomers are **diastereomers**.

For example, cis-but-2-ene and trans-but-2-ene shown in Figure 3.1.1 are diastereomers because they have the same formula, and the same atom-connectivity, but methyl groups are oriented in the same direction in cis- and in opposite directions in trans-isomer, they do not mirror each other. The cis- and trans- represent two different configurations of but-2-ene as shown in Figure 3.1.1.

3.2: Confirmations

Learning Objectives

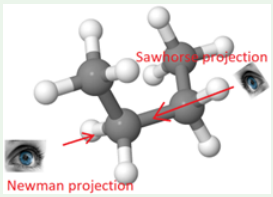
- Draw sawhorse projections and Newman projections of simple organic compounds.
- Name, predict relative stability, and energy barriers to interconversion of conformations of simple organic compounds, including ethane, butane, and cyclohexane.
- Draw chair conformations of cyclohexane, substituted cyclohexanes, and their flipped forms, and predict the stability of the bulky groups at the axial and equatorial positions.

Representing orientation of groups connected to a $C-C$ bond

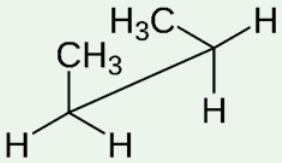
There are multiple ways of visualizing the orientation of atoms along a $C-C$ bond, including sawhorse projection, Newman projection, and Fisher projection.

Sawhorse projection and Newman projection of a molecule

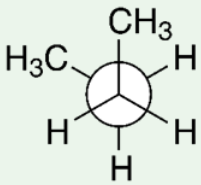
- **Sawhorse projection** looks at the $C-C$ bond from an oblique angle, as shown in Figure $3.2.1$.
- **Newman projection** looks from front to back with the front atom represented as a dot with bonds originating from the dot and the back atom is represented as a wheel with bonds shown as lines originating from the wheel, as illustrated in Figure $3.2.1$.



Model showing viewpoints by arrows



Sawhorse projection



Newman projection

Figure $3.2.1$: Model with the viewpoints indicated by arrows, sawhorse projection, and Newman projection of butane ($CH_3CH_2CH_2CH_3$. (Copyright; Public domain)

The relative orientation of groups connected to the two C's of $C-C$ bond is quantitatively shown by the dihedral angle.

Dihedral angle or torsion angel

Dihedral angle or **torsion angle** is the clockwise angle between half-planes through two sets of three atoms, e.g., the half-planes defined by $R-C-C$ and $R'-C-C$ in Figure $3.2.2$, having two atoms in common $C-C$ in this case).

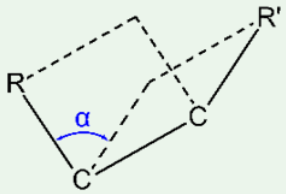


Figure $3.2.2$: Dihedral angle α between the half-planes defined by $R-C-C$ and$R'-C-C$ in a hydrocarbon RCH_2CH_2R' molecule (Copyright; Д.Ильин: vectorization, Public domain via Wikimedia commons)

Conformational isomers

The torsion angle changes due to free rotation around the single bond. However, the rotation is not completely free; there are energy barriers. Therefore, as the torsion angle gradually changes from 0^o to 360^o, the relative potential energy of the molecule goes through local minima to local maxima, often at intervals of 60^o.

Conformations

- **Conformations** are the different arrangements of atoms in a molecule resulting from rotation around a single bond.
  - **Eclipsed conformation** is a conformation in which two substituents on adjacent atoms are at 0^o torsion angle and the
  - **Staggered conformation** is the confirmation in which the two substituents on adjacent atoms are at 60^o torsion angle, as illustrated in Figure 3.2.3.

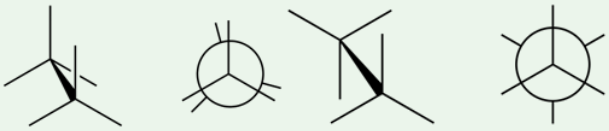
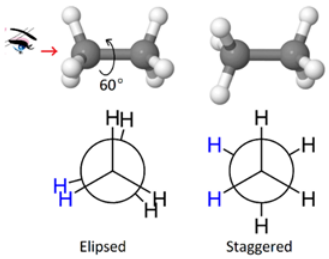


Figure 3.2.3: Eclipsed conformation in sawhorse projection (left) and Newman projection (right) and the corresponding staggered conformation (Copyright; Jovianeye, Public domain via Wikimedia commons).

Conformations of ethane

The eclipsed and staggered conformations of ethane are shown in the figure on the right. The eclipsed conformation of ethane is higher energy (unstable) relative to staggered confirmation by 12.5 kJ/mol. The eclipsed conformation changes to staggered, and vice versa, after every 60^o rotation around the $C-C$ bond as illustrated in Figure 3.2.4.



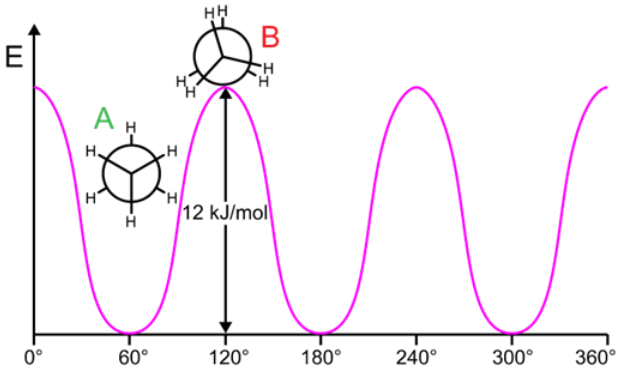


Figure 3.2.4: Newman projections of ethane conformations for rotation around C-C bond & their relative energy differences (not total energies) vs. torsional/dihedral angle curve, where the conformation A is staggered, and B is eclipsed. (Copyright; Keministi, Public domain via Wikimedia commons)

Factors in stabilizing or destabilizing conformations

Factors that contribute to the stabilization or destabilization of the conformation are the following.

1. Destabilization due to electrostatic repulsion between the same charge dipoles in close proximity in eclipsed conformation,
2. destabilization due to steric repulsion, i.e., physical repulsion of the groups due to space requirement in close proximity in eclipsed conformation, and
3. stabilization due to hyperconjugation, i.e., an overlap of a σ-bond on one atom with an antibonding orbital on the neighboring atom in staggered conformation.

These factors are illustrated in Figure 3.2.5 for the case of ethane.

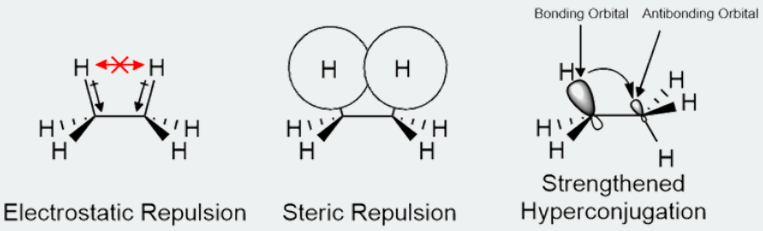
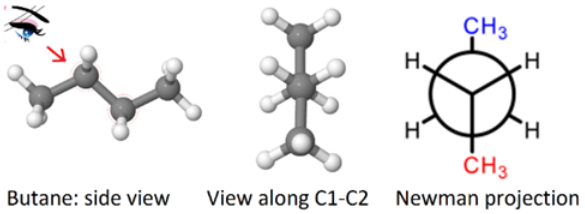


Figure 3.2.5: Illustrates the contributions of steric and electrostatic interactions as well as hyperconjugation to the energy barrier of rotation in ethane (Copyright; UMMedChemStudent, CC BY-SA 3.0, via Wikimedia Commons)

Conformations of butane

The most stable conformation of butane ($CH_3CH_2CH_2CH_3$) for a view of C2-C3 bond is illustrated on the right. This staggered conformation with the bulky ($-CH_3$ groups in this case) fathers apart is given a special name,'anti' and labeled A in the relative energy vs. tortion angle curve shown in Figure 3.2.6.



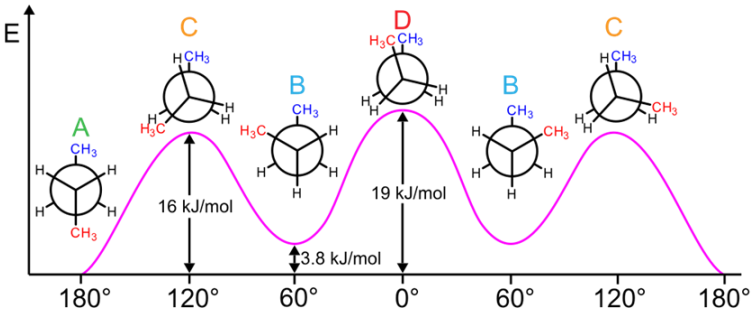


Figure 3.2.6: Newman projections of butane conformations for rotation around C2-C3 bond and their relative energy differences (not total energies) vs. torsional/dihedral angle curve. A: antiperiplanar, anti or trans, B: gauche, C: eclipsed, D: synperiplanar or cis. (Copyright; aKeministi, Public domain via Wikimedia commons)

A 60° rotation converts the anti-conformation to eclipsed conformation, labeled C in the figure, which is 16 kJ/mol higher energy relative to the anti-confirmation. A further 60° rotation converts the eclipsed conformation to another staggered conformation called gauche-conformation, labeled B, at a local energy minimum but 3.8 kJ/mol higher than anti-conformation. The gauche-conformation has the bulky ($-CH_3$ in this case) groups at 60° torsion angle where they are close enough to cause some steric repulsion between each other. Another 60° rotation converts the gauche-conformation to an eclipsed conformation called cis, labeled D, in which the bulky ($-CH_3$ in this case) groups are eclipsing each other. This cis-conformation is 3 kJ/mol higher in energy than the other eclipsed confirmation (labeled B) and it is the highest energy conformation of butane. These conformations keep repeating at an interval of 60° torsion angle as illustrated in Figure 3.2.6.

Conformational isomers

- Conformations corresponding to local minima on the potential energy curve, e.g., anti- and gauche-conformations in the case of butane, are called **conformational isomers**, **conformers**, or **rotamers**.
- The conformations corresponding to the local maxima on the energy curve, e.g., eclipsed conformations in the case of butane, are the **transition states** between the conformational isomers.

Energy cost of eclipsing and gauche interaction of $-H$ and $-CH_3$ groups

It can be calculated form Figure 3.2.4 and Figure 3.2.6 that each

- $-H$ to $-H$ eclipsing costs 4 kJ/mol,
- $-H$ to $-CH_3$ eclipsing costs 6 kJ/mol,

- $-CH_3$ to $-CH_3$ eclipsing costs 11 kJ/mol, and
- $-CH_3$ to $-CH_3$ gauche interaction costs 3.8 kJ/mol.

Conformations and strains in cycloalkanes

If a cycloalkane is in a planar conformation, i.e., all the C′s in the chain be in the same plane, there will be two kinds of strain:

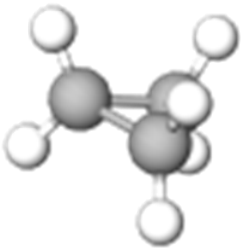
1. **angle strain,** i.e., the strain caused by bending the bond angles from the ideal tetrahedal bond angles of 109.5°, and
2. **eclipsing strains** between a pair of H on each C with pairs of H′s on the two neighboring carbons.

Cyclopropane has a planar ring with bond angles reduced to 60° and three pairs of eclipsing H′s, resulting in an overall ring strain of ~120 kJ/mol.

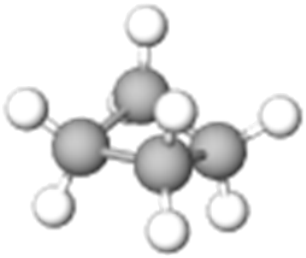
A planar cyclobutane would have bond angles reduced to 90° and four pairs of eclipsing H′s, but it changes to a puckered confirmation that increases angle strain a little but significantly reduces the eclipsing strain, resulting in an overall ring strain ~110 kJ/mol.

A planar cyclopentane would have bond angles of 108°, i.e., close to the regular value of 109.5°, but five pairs of eclipsing H′s. Cyclopentane is puckered, increasing angle strain slightly but significantly reducing the eclipsing strain, resulting in an overall strain of ~25 kJ/mol. Four of the C′s of cyclopentane are almost in one plane and the fifth is raised, giving it an envelope shape, also called envelop conformation. Cyclopentane is one of the most common ring structures found in nature due to its low angle strain. For example, ribose -a sugar that is a component of DNA monomers, and fructose or fruit sugar are five-membered rings.

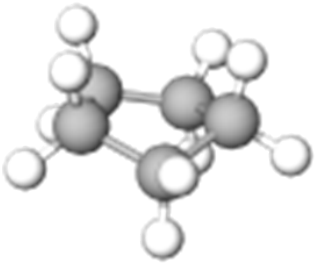
Cyclohexane acquires a bent conformation that looks like a chair -called chair conformation, in which all bond angles are close to the tetrahedral value of 109.5° and all H′s are staggered with no ring strain. Cyclohexane is the most common ring structure due to its zero ring strain. For example, glucose -the most common monosaccharide (sugar) is six-membered rings. Starch, a carbohydrate component of food, and cellulose, a major component of wood, are made of glucose units. The conformations of cyclopropane, cyclobutane, cyclopentane, and cyclohexane are illustrated below.



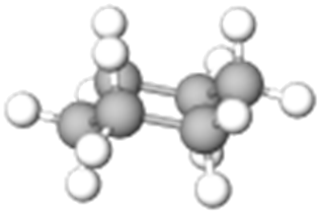
Cyclopropane,
Ring strain ~120 kJ/mol



Cyclobutane
Ring strain ~110 kJ/mol

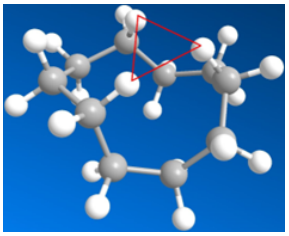


Cyclopentane
Ring strain ~25 kJ/mol



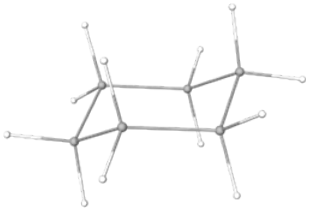
Cyclohexane
No ring strain

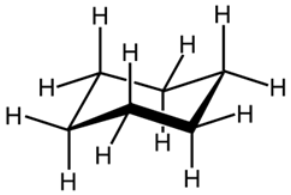
Medium-size rings (7-13 C′s) can pucker/bend to minimize angle strain but some **transannular strain**, i.e., steric repulsion between groups attached to non-adjacent ring atoms, appear between groups on some C′s with other groups at a distance across the ring. For example, the figure on the right highlights the transannular strain between some H′s with a red triangle in the lowest energy conformation of cyclodecane (Copyright; Chem540f09grp4, Public domain, via Wikimedia Commons). In larger rings (14 or more C′s), there is little or no ring strain. The order of strain in cyclic hydrocarbons is the following: cyclopropane > cyclobutane > cyclopentane > cyclohexane < cycloheptane <cyclooctane.



Conformations of cyclohexane

Cyclohexane which is the most common cyclic structure, exists primarily in a chair conformation in which all bonds are 111°, i.e., very close to tetrahedral value 109.5°, and all six pairs of H′s are staggered as shown in Figure 3.2.7.


Model of chair conformation


Edge-on view of chair conformation

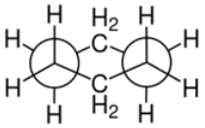

Newman projection

Figure 3.2.7: Cyclohexane: a model of chair conformation, edge-on view, and Newman projection from the left side of the edge-one view showing staggered conformation. (Copyright; model: public domain, edge-one view: Roland.chem, Public domain, via Wikimedia Commons, and Newman projection: Chem Sim 2001, Public domain, via Wikimedia Commons)

Drawing chair conformation of cyclohexane

Follow the following 5 steps to draw the chair form of a cyclohexane ring.

1. Draw two parallel lines slanting down, separated by about half their length, and the bottom line starting from about the middle of the first as shown in Figure 3.2.8.
2. Start a line from the lower end of the bottom line making a wide V shape, and extend it to about parallel to the top of the top line. Draw a line parallel to the first one in this step, starting from the top of the top line and extending to about the bottom of the bottom line. Figure 3.2.8 shows these two lines in blue.
3. Connect the top of the right side line of step 2 to the bottom end of the top line of step 1 and do the same to the other line of step 2. These are shown in red in Figure 3.2.8. The chair skeleton of the cyclohexane ring is complete at this stage.
4. Draw lines starting from the corners of the ring, going upwards from the corners pointing up, and going downwards from the corners pointing down. These bonds are approximately parallel to the axis of the ring called **axial bonds,** shown in green in Figure 3.2.8.
5. Draw lines starting from the corners making a big V-shape to the axial bond and approximately parallel to the bond one bond away in the ring. These are approximately along the equator of the ring and called **equatorial bonds**, shown in black, blue, and red, corresponding to the color of the bonds one bond away to which they are drawn parallel in Figure 3.2.8.

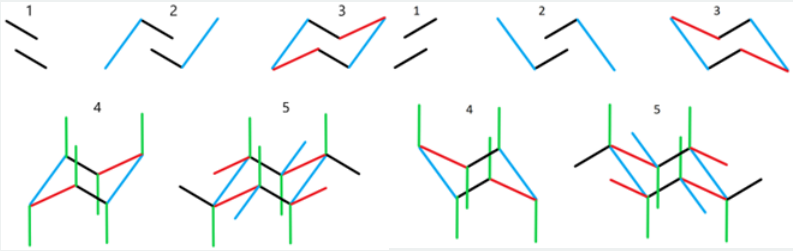


Figure 3.2.8: Illustration of steps to draw chair conformation of cyclohexane (right) and its flipped form (right). (Copyrights; Public domain)

There are two chair conformations. They are flipped form of each other. To draw the flipped form of the first, follow the same five steps but start with two parallel lines slanting upwards, as illustrated in Figure 3.2.8.

Ring flipping in cyclohexane

There is limited rotation around C−C in a cyclohexane chain that allows flipping of the cyclohexane ring that converts all the axial bonds into equatorial and all the equatorial bonds to axial as illustrated by animation and drawing in Figure 3.2.9.

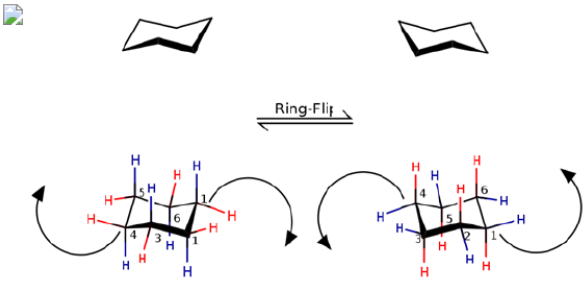


Figure 3.2.9: Animation of cyclohexane ring flipping (left) and sketch showing all axial bonds become equatorial and voice versa upon flipping the ring. (Copyrights; animation: Cleanthis, Public domain, via Wikimedia Commons, and the sketch: Naturwiki, Public domain, via Wikimedia Commons)

The ring flipping goes through twist boat conformation and half-chair and boat transition states with higher energy conformations as shown in Figure 3.2.10. The two chair conformations are of equal energy in the case of cyclohexane and predominant in the mixture due to their low energy. For every 10,000 molecules in chair conformations, there is no more than one molecule in the twist boat conformation.

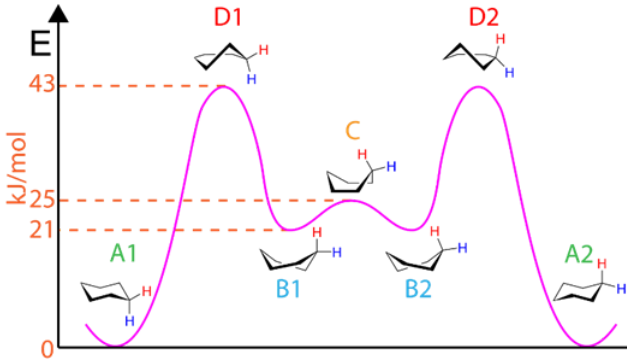


Figure 3.2.10:Cyclohexane chair flip (ring inversion) reaction plotted against their energy differences. Inversion happens quickly & constantly at room temperature. A1 & A2: chair; B1 & B2: twist-boat; C: boat; and D1 & D2: half-chair. (Copyright; Keministi, Public domain via Wikimedia commons)

Conformations of monosubstituted cyclohexanes

Two chair conformations of cyclohexane are of equal energy. When one of the H is replaced with a bulky group like $-CH_3$ group, the bulky group is at an equatorial position in one and at an axial position in the other chair conformation and the two are not the same energy. The bulky group at the axial position has steric repulsion with the other two axial H′s on the same face of the ring, called 1,3-diaxial interaction, as illustrated in Fig. 11. Therefore, the chair conformation with the bulky group at equatorial position is more stable and more predominant at equilibrium than its flipped form with the bulky group at axial position. This effect depends on the size of the group; the larger the group the stronger the effect.

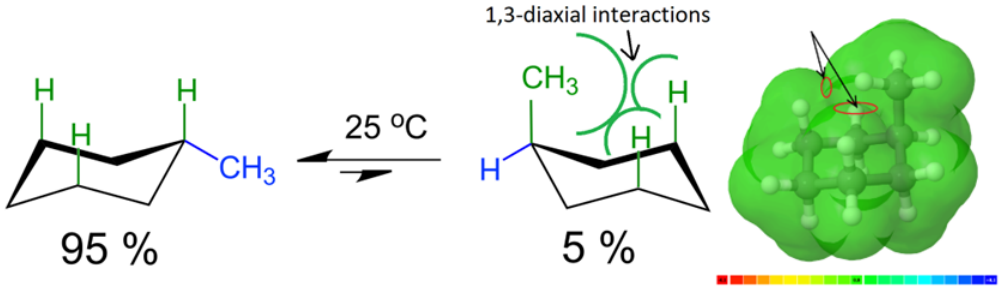


Figure 3.2.11: Illustration of the equilibrium between axial and equatorial methylcyclohexanes with an illustration of the destabilizing steric repulsion in the axial methylcyclohexane, sketch (left) and model (right) with the steric strain area highlighted by circles and indicated by arrows. (Copyright; Public domain)

Preference of equatorial groups over axial groups in nature

Six-membered rings are the most common. Among them, those with all or most of the bulky groups in equatorial positions are significantly more common because of the stability of the equatorial relative to the axial bulky groups. For example, several six-membered ring monosaccharides (sugars) have the most bulky groups in equatorial positions. Glucose with all the bulky groups at the equatorial positions is the most common among them. Glucose units combine through an equatorial group to form a polymer called cellulose, the major component of wood. Starch is also a polymer of glucose but with one connecting bond axial. This changing connecting bond from equatorial to axial changes cellulose, a hard structural material, to starch, a carbohydrate. Wood can survive centuries, but starch would get rotten and decompose in a few days if left out in the open. The structures of glucose, starch, and cellulose are shown in Figure 3.2.12

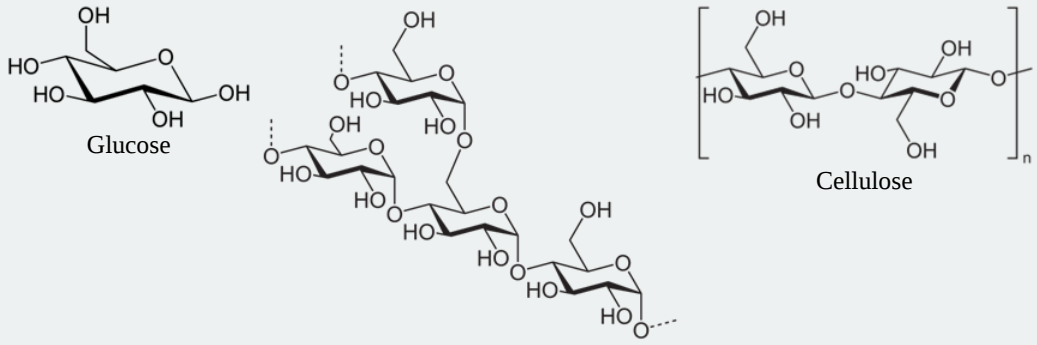


Figure 3.2.12: Structure of glucose, starch, and cellulose. (Copyright; glucose: Yikrazuul, Public domain, via Wikimedia Commons, Starch, and cellulose: NEUROtiker, Public domain, via Wikimedia Commons)

Fused ring systems also exist in nature; most have six- and five-membered rings. For example, the basic skeleton of steroid hormones is three six-membered rings designated A, B, and C, and one five-membered ring designated D, fused, as shown in Fig 13. The C−C bond shared between the two fused rings is the **bridge**. The two C′s at the ends of the bridge are **bridgehead** C′s. Each bridgehead C has two C−C bonds, one in each of the fused rings. **Trans ring fusion** has all four bonds, from the bridgehead C′s into the two fused rings, equatorial and **cis ring fusion** has one of the four bonds axial. Steroid hormones always have trans-fused rings B, C, and D, as shown in Fig. 13. Ring A may be cis-fused, as in example 5β in Figure 3.2.1, but in a majority of cases, it is also trans-fused as in example 5α in Figure 3.2.1. This is because the cis fusion with one axial C−C bond at the ring connection is associated with steric strain.

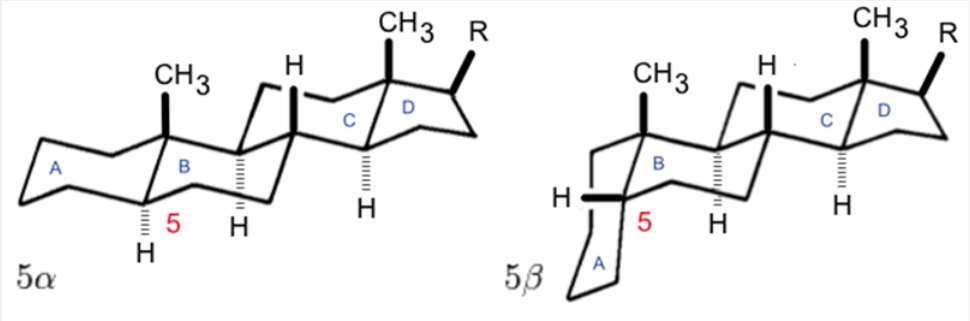


Figure 3.2.1: Steroid 5α and 5β stereoisomers. (Copyright; Leprof 7272, CC BY-SA 3.0, via Wikimedia Commons)

3.3: Configurations

Learning Objectives

- understand the meaning of absolute configuration, chiral centers, and draw perspective drawings and Fisher projections to represent chiral compounds.
- Apply rules to assigning D/L or R/S stereodescriptors to the chiral compounds containing one or two chiral centers.
- Understand optical activity as a physical characteristic of chiral compounds and the factors that affect it.
- Recognize stereochemical relationships between organic compounds, including enantiomers, diastereomers, and meso.

What is a configuration?

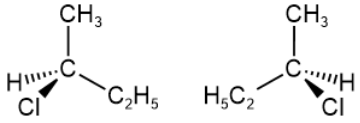
A molecule's permanent geometry that results from its bonds' spatial arrangement is called **configuration**. Stereoisomers have different configurations for the same set of atoms and the same group of bonds. For example, cis-but-2-ene has a different configuration than its stereoisomer trans-2-butene, shown in Figure 3.1.1.

Representing a configurations

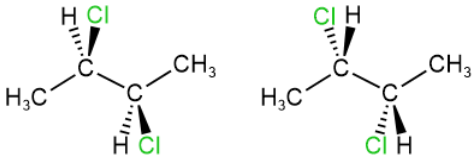
Representing configuration around a single atom will be described first. An sp-hybridized atom has a linear geometry, as in the case of hydrogen cyanide $H-C\equiv N$ or carbon dioxide $O=C=O$ molecules. An sp^2-hybridized atom is a trigonal planer, i.e., there are three σ-bonds at 60^o to each other in the same plane and one π-bond with one of the σ-bond, as in the case of formaldehyde $H_2C=O$ represented as  . An sp^3-hybridized atom is tetrahedral, i.e., four σ-bonds as two V's of 109^o internal angle joined perpendicular to each other at vortexes, e.g., methane CH_4 modeled as  . Tetrahedral geometry is a 3D geometry usually represented by perspective drawings or Fisher projections, described in the next sections.

Perspective drawing

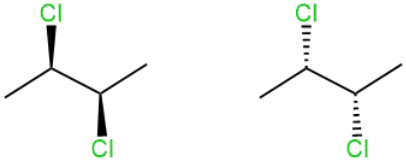
Since the tetrahedral geometry is two V's perpendicular to each other, one V is placed in the plane of the screen (or page) and represented by solid lines. The other V is represented by a solid wedge representing the bonding coming out of the screen towards the viewer and a hashed wedge representing the bond going beyond the screen away from the viewer, as illustrated below for the case of two configurations of 2-chlorobutane. The view angle is from the direction of the top-left corner. The V in the plane of the screen appears larger and the other appears smaller. Usually, the solid wedge is near the viewer relative to the hashed wedge.



If the configuration of two or more connected $C's$ needs to be shown, the V's in the plane of the screen is usually drawn pointing downwards or upwards, making a zig-zag line as illustrated below for the case of two configurations of 2-chlorobutane.



In the case of skeletal formulas, $C's$ and $H's$ are not shown. However, the configuration is still readable correctly from the bonds shown, as illustrated below for the case of two configurations of 2-chlorobutane.



Fisher projections

Fisher projections represent the configuration as a projection of the molecule in 2D with four bonds in the shape of a cross, as explained in Figure 3.3.1.

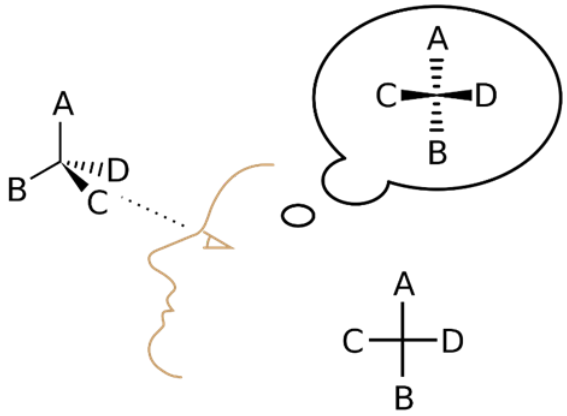
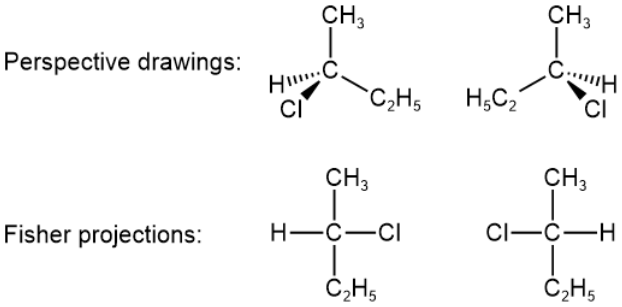


Figure 3.3.1: Draw Fisher projections by: i) looking at the central atom such that the bonds extending towards the viewer are placed horizontally and the bonds extending away from the viewer are drawn vertically, and ii) replacing the sold and hashed wedges with straight lines assuming the horizontal lines are bonds forward and the vertical lines are bonds backward. (Copyright; Slashme, CC BY-SA 3.0, via Wikimedia Commons)

Examples of Fisher projections of the two configurations of 2-chlorobutane are shown below, along with the perspective drawings.



When more than one consecutive C′s needs to be shown by a Fisher projection, the C chain, i.e., the skeleton, is shown vertically. For this purpose, C−C single bond is rotated by 180° resulting in a semi-cyclic conformation. Then the C′s of interest are viewed from the edge-on so that the horizontal bonds are pointing towards the viewer and drawn vertically from top to bottom, as illustrated Figure 3.3.2.

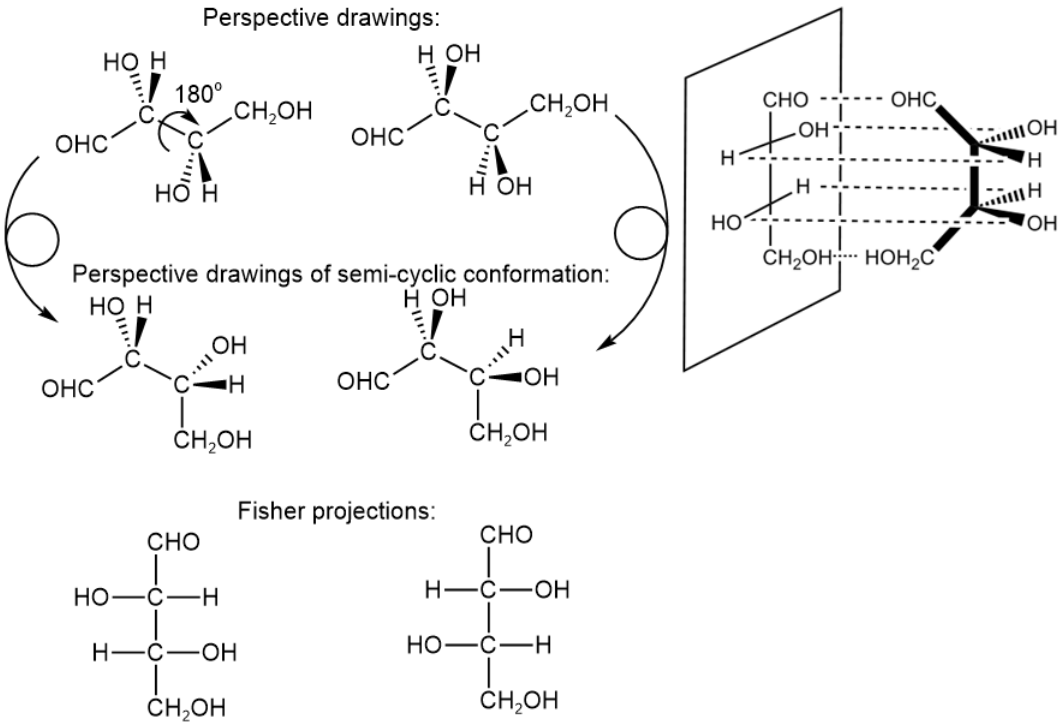


Figure 3.3.2: Illustration of how to modify perspective drawing of more than one chiral center into a semi-circle conformation (left) and draw its Fisher projection (right). (Copyright: Left: Public domain (left) and right: Olion17, Public domain, via Wikimedia Commons

Chiral center

A compound with sp-, sp^2-hybridized central atom, or a sp3-hybridized central atom with two or more same groups have a superimposable mirror image, as illustrated in Figure 3.3.3. Since the compound and its mirror image are identical in these cases, these compounds are achiral, i.e., not chiral.

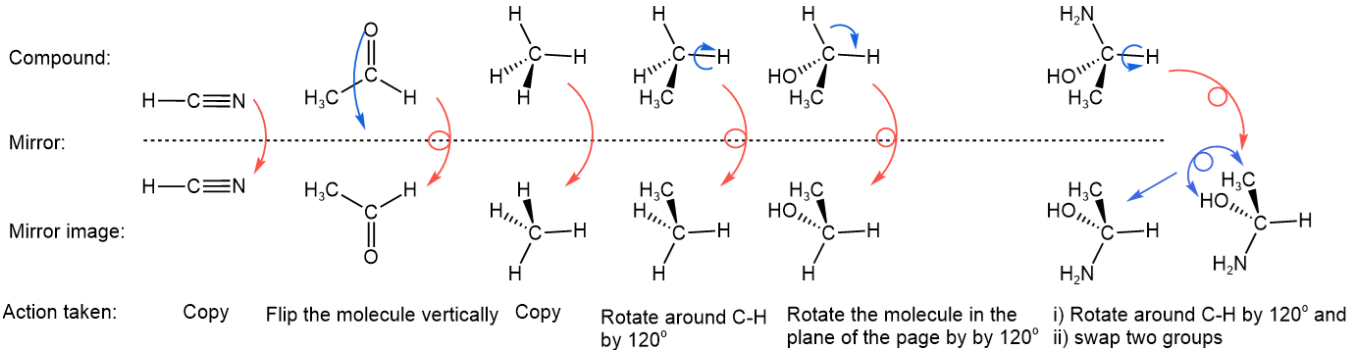


Figure 3.3.3: Illustration of converting a compound into its mirror image: simply copy or 'rotate and copy' for achiral, but making and breaking bonds is needed for chiral compounds. Green arrows indicate flip or rotation. (Copyright; Public domain)

A compound with four different groups attached to an sp^3-hybridized central atom has two distinct configurations, which are related as non-superimposable mirror images of each other, i.e., chiral. Since the two configurations represent different compounds that are stereoisomers, making and breaking of bonds, i.e., a chemical change, is needed to inter-convert them, as illustrated in Figure 3.3.3.

> An atom, often a C with four different groups attached to it, is called a **chiral center**.

A chiral center is often a cause of the chirality of the molecule. A stereochemical descriptor, e.g., D or L and R or S, is used to distinguish the two configurations of a chiral center.

An **absolute configuration** is the spatial arrangement of the atoms of a chiral molecular entity specified by a stereochemical descriptor.

Stereocenter

A **stereocenter** is an atom, axis, or plane with at least three different groups attached to it, so interchanging any two groups creates a new stereoisomer. The chiral center is a sub-class stereocenter, as illustrated in Figure 3.3.3. The C′s of the C=C-bond of cis-but-2-ene and trans-but-2-ene shown in Figure 3.1.1, have three different groups and they are sterocenters. For example, swapping $-CH_3$ and $-H$ on any of the two C′s converts cis-but-2-ene to its stereoisomer trans-but-2-ene and vice versa.

Stereochemical descriptors

Two sets of stereochemical descriptors, i.e., D or L and R or S, are in common use and described next.

D/L Stereochemical descriptors

The D and L stereochemical descriptors were developed for monosaccharides, i.e., sugars. The monosaccharides have an aldehyde or a ketone group placed on the top or near the top in Fisher projections. Every other C usually has an alcohol ($-OH$) group on them, and they are chiral except the terminal C′s, which are achiral.

Looking at the bottom-most chiral C in the Fisher projection of a monosaccharide, if the $-OH$ group is towards the right of the Fisher projection, it is designated D and if it is towards the left, it is designated L, as illustrated in Figure 3.3.4.

`Figure 3.3.4: Illustration of D and L monosaccharides (glyceraldehyde and glucose), and amino acid (alanine). (Copyright; Public domain)

A hyphen separates the name of the monosaccharide from the D or L stereodescrptor. The D or L defines the absolute configuration of the bottom-most chiral C and the name that follows it defines the absolute configuration of all other chiral C′s in a monosaccharide.

Amino acids have a chiral (\ce{C}\) with an amine ($-NH_2$) and a carboxylic acid ($-COOH$) attached to it. The $-COOH$ group is placed on the top and the $-NH_2$ group is either on the left or right side of the Fisher projection.

If the $-NH_2$ group is towards the right of the Fisher projection of an amino acid it is designated D. If it is towards the left, it is designated L, as illustrated in Figure 3.3.4.

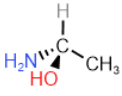
- Natural monosaccharides (sugars) have D-configurations, and natural amino acids have L-configurations with few exceptions.
- Although the common names of monosaccharides and amino acids with a D or L descriptor are not IUPAC names, these are commonly used in biochemistry.

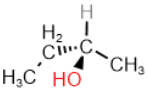
R/S Stereochemical descriptors

The R and S stereochemical descriptors are part of the IUPAC nomenclature. For this purpose, the four groups attached to the chiral center are assigned priority 1, 2, 3, and 4, where 1 is the highest priority and 4 is the lowest. Then R or S is assigned based on the groups' relative orientations, considering the priority order.

Sequence rules

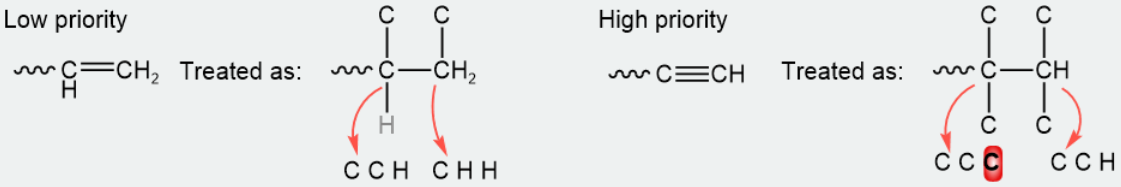
The following is the sequence rules, also called Cahn-Ingold-Prelog (CIP) sequence rules for assigning priority.

1. A group attached to a stereocenter through an atom of a higher atomic number (Z) has higher priority. For example, the priority of the four groups in  is: $-OH > -NH_2 > -CH_3 > -H$.

2. If there is a tie, look at atoms two bonds away from the stereocenter. List them according to their atomic number (Z) and compare them, one by one, starting from the highest atomic number. Stop at the first point of difference and assign higher priority to the higher atomic number at the first point of difference. For example, in  there is a tie between $-CH_3$ and $-CH_2-CH_3$ due to both having C at one bond away from chiral center. Comparison of atoms two bonds away: $\begin{matrix} H,H,H \\ C,H,H \end{matrix}$ assigns higher priority to $-CH_2-CH_3$ over $-CH_3$, i.e., the priority order is: $-OH > -CH_2-CH_3 > -CH_3 > -H$.
3. If the tie does not break at the distance of two bonds, move to atoms at the distance of three bonds from the chiral center and repeat the process until the first point of difference arrives.

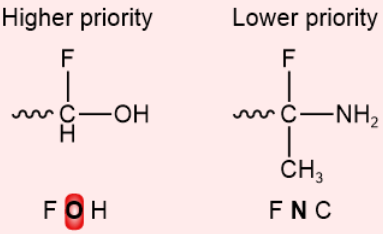
CIP rule for atoms with double or triple bond

Atom bonded by a double bond is treated as connected twice, and by a triple bond is treated as connected three times. For example, $-CH=CH_2$ and $-C\equiv CH$ are the same for the atom directly attached to the chiral center. For comparing atoms two bonds away, the groups are treated as explained in the illustration below, with the first point of difference highlighted in red.



Caution

The rule is to stop at the first point of difference, even if the order may change later, as illustrated below with the first point of difference highlighted in red.



Assigning R or S to the chiral center

After all four groups on the chiral center have been assigned priority numbers 1, 2, 3, and 4, two situations arise:

- **Case 1:** The lowest priority, i.e., #4, is pointing away from the viewer, i.e., it is on a hashed line in the perspective drawing or on a vertical line in Fisher projection. In this case, R (*rectus*, Latin for right) is assigned if an arrow drawn from #1 to #2 to #3 is clockwise or directed to the right, and S (*sinister*, Latin for left) is assigned if it is directed to the left or counterclockwise, as illustrated in Figure 3.3.5.
- **Case 2**: The lowest priority, i.e., #4 is not pointing away from the viewer. In this case,
  - swap #4 with the group pointing away from the viewers,
  - since a swap of any two groups changes the configuration, reverse the assignments, i.e., assign R if an arrow drawn from #1 to #2 to #3 is clockwise or directed to the right and S if it is directed to the left or counterclockwise, as illustrated in Figure 3.3.5.

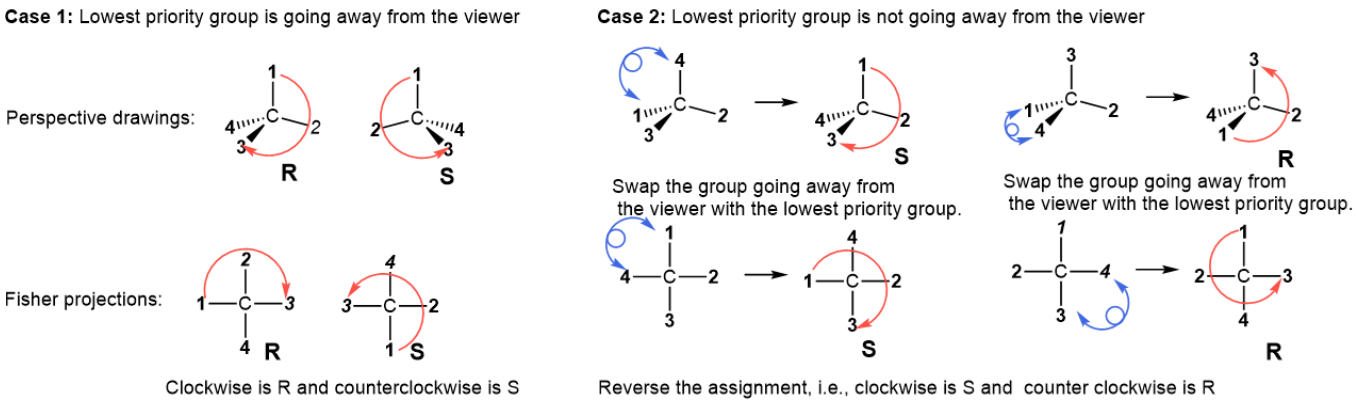
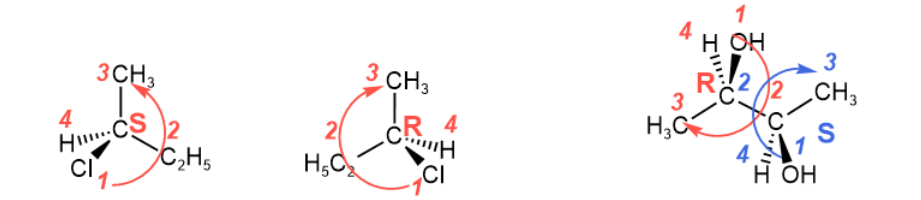


Figure 3.3.5: Illustration of assigning R or S configuration to a chiral center based on priority# 1, 2, 3, and 4, where 1 is the highest priority, and 4 is the lowest. (Copyright; Public domain)

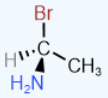
The stereochemical descriptor R or S describes the absolute configuration of the chiral center. They are placed within small brackets at the beginning of the IUPAC name, separated by a hyphen, as shown in the examples below. If there is more than one chiral center in a compound, the locant number precedes each chiral center's R or S descriptor, as explained in the examples below.



(S)-2-chlorobutane (R)-2-chlorobutane (2R,3S)-butane-2,3-diol

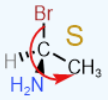
Example 3.3.1

Assign R/S to the compound shown on the right.



Solution

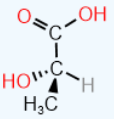
1. Assign priority based on the atomic number of atoms attached to the chiral center: $-\mathrm{Br} > -\mathrm{NH_2} > -\mathrm{CH_3} > -\mathrm{H}$.
2. Swap #4 with the group away from the viewer (hashed wedge). No action needed as $-\mathrm{H}$ is already at the hashed wedge.
3. Draw an arrow from #1 to #2 to #3. The arrow is counterclockwise, i.e., the stereodescriptor is S.



Answer: S

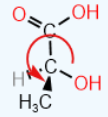
Example 3.3.2

Assign R/S to the compound shown on the right.



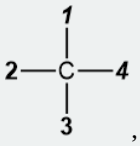
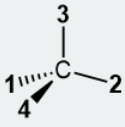
Solution

1. Assign priority based on the atomic number of atoms attached to the chiral center: $-\mathrm{OH}$ is #1, there is a tie between $-\mathrm{CH_3}$ and $-\mathrm{COOH}$, and $-\mathrm{H}$ is #4.
    1. Tiebreaker is O two bonds away $\overset{O,O,O}{H,H,H}$, assigning #2 to $-\mathrm{COOH}$ and #3 to $-\mathrm{CH_3}$.
2. Swap #4 with the group away from the viewer (hashed wedge), i.e., $-\mathrm{OH}$ with $-\mathrm{H}$ in this case.
3. Draw an arrow from #1 to #2 to #3. The arrow is counterclockwise. Since two groups were swapped, reverse the assignment i.e., counterclockwise is R.

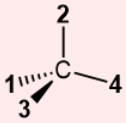


Answer: R

Trick for assigning R and S

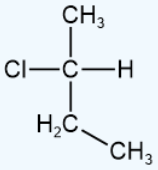
When the lowest priority group #4 is on the vertical line in the Fisher projection, as in  or  , there is no need to swap to bring it on the vertical line. Draw the arrow from #1 to #2 to #3 and reverse the assignments, i.e., assign S if the arrow is clockwise and R if it is counterclockwise. This trick also works when the lowest priority group #4 is on the solid wedge in the perspective drawing, as in  . This trick makes the R/S assignment easy, especially for complex organic compounds where redrawing the structure after the swap is not easy.

Caution

Do not try this trick when the lowest priority group #4 is in the plane of the page in the perspective drawing, as in  or  . These two structures have opposite configurations, but the arrow drawn from #1 to #2 to #3 is in the same direction for both. It will lead to an incorrect assignment in one of the two.

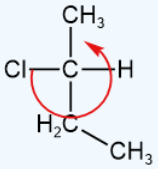
Example 3.3.3

Assign R/S to the compound shown on the right.



Solution

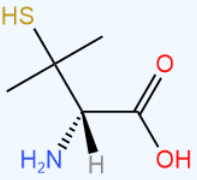
1. Assign priority based on the atomic number of atoms attached to the chiral center: $-\mathrm{Cl}$ is #1, there is a tie between $-\mathrm{CH_3}$ and $-\mathrm{CH_2CH_3}$, and $-\mathrm{H}$ is #4.
    1. Tiebreaker is C two bonds away $\overset{C,H,H}{H,H,H}$, assigning #2 to $-\mathrm{CH_2CH_3}$ and #3 to $-\mathrm{CH_3}$.
2. Swap #4 with the group away from the viewer (hashed wedge). Trick: No need for Fisher projection, as the assignment can be reversed in the last step.
3. Draw an arrow from #1 to #2 to #3. The arrow is clockwise. Since the lowest priority group is on vertical line, reverse the assignment i.e., clockwise is S.



Answer: S

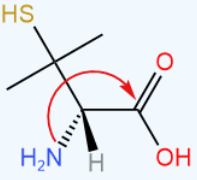
Example 3.3.4

Assign R/S to the compound shown on the right.



Solution

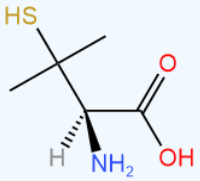
1. Assign priority based on the atomic number of atoms attached to the chiral center: $-\mathrm{NH_2}$ is #1, there is a tie between $-\mathrm{C(CH_3)_2SH}$ and $-\mathrm{COOH}$, and $-\mathrm{H}$ is #4.
    1. Tiebreaker is S two bonds away $\overset{S,C,C}{O,O,O}$, assigning #2 to $-\mathrm{C(CH_3)_2SH}$ and #3 to $-\mathrm{COOH}$.
2. Swap #4 with the group away from the viewer (hashed wedge). Trick: No need for perspective drawing with the lowest priority group on a solid wedge, as the assignment can be reversed in the last step.
3. Draw an arrow from #1 to #2 to #3. The arrow is clockwise. Since the lowest priority group is on the solid wedge, reverse the assignment i.e., clockwise is S.



Answer: S (This compound is (S)-penicillamine)

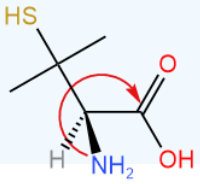
Example 3.3.5

Assign R/S to the compound shown on the right.



Solution

1. Assign priority based on the atomic number of atoms attached to the chiral center: $-\mathrm{NH_2}$ is #1, there is a tie between $-\mathrm{C(CH_3)_2SH}$ and $-\mathrm{COOH}$, and $-\mathrm{H}$ is #4.
    1. Tiebreaker is S two bonds away $\overset{S,C,C}{O,O,O}$, assigning #2 to $-\mathrm{C(CH_3)_2SH}$ and #3 to $-\mathrm{COOH}$.
2. Swap #4 with the group away from the viewer (hashed wedge). Not needed as the lowest priority group is already away from the viewer on hashed wedge.
3. Draw an arrow from #1 to #2 to #3. The arrow is clockwise. Clockwise is R.



Answer: R (This compound is (R)-penicillamine)

Compounds with one chiral center

Compounds containing one chiral center are chiral. They have a stereoisomer called enantiomer, which is their non-superimposable mirror image. Some examples of compounds containing one chiral center, their enantiomers with their absolute configuration assigned by D/L system and R/S system, and their tetetrahederal geometry illustrated are shown in Figure 3.3.6.

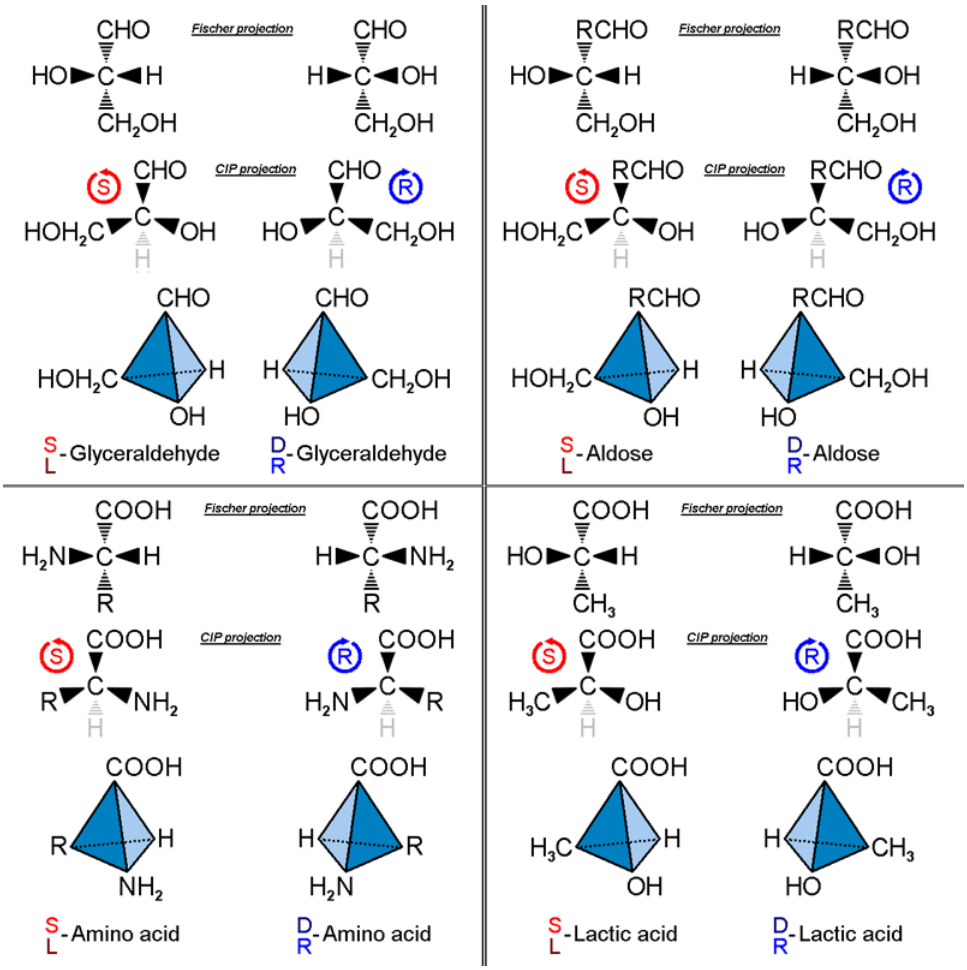


Figure 3.3.6: Absolute configurations assigned according to D/L system and R/S system and tetrahedral geometry illustrated for glyceraldehyde (top-left), the general structure of monosaccharides containing an aldehyde group called aldoses (top-right), general structure of amino acids (bottom-left), and lactic acid (bottom-right). (Copyright; Qniemiec, CC BY-SA 3.0, via Wikimedia Commons)

> Enantiomers have the same physical and chemical properties, except when they react with or interact with other chiral objects or molecules.

For example, hands and gloves are chiral objects. The right-hand glove fits on the right hand, and the left-hand fits on the left. Similarly, enantiomer molecules react equally and when they are produced in a chemical reaction they are produced equally, except when one of the reagents is chiral. For example, enzymes are chiral reagents and usually react with one of the two enantiomers in the reactants and/or produce one of the two enantiomers.

Optical activity

The light has an electric field oscillating perpendicular to the direction of propagation of the light. Ordinary light has an electric field oscillating in all places perpendicular to the direction of propagation. When it passes through a polarizer, it transmits only the waves oscillating in one plane, i.e., plane polarized light, as illustrated in Figure 3.3.7. The plane polarized light is composed of left-handed and right-handed circularly polarized light, which are non-superimposable mirror images of each other, i.e., chiral, as shown in Figure 3.3.7.

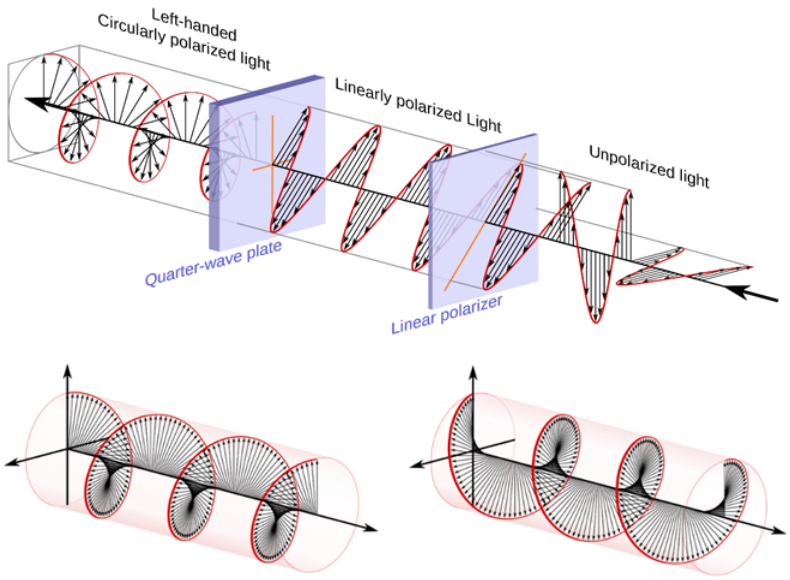


Figure 3.3.7 : Top: Illustration of unpolarized light with an electrical component oscillating in all directions perpendicular to the axis of propagation, linear polarized light in which the oscillation is restricted in one direction, and a left-handed component of the linear polarized light. Bottom: left-handed and right-handed components of plane-polarized light that are non-superimposable mirror images of each other. (Copyright: Dave3457, Public domain, via Wikimedia Commons)

When the plane-polarized light passes through a solution of chiral molecules, one enantiomer absorbs more right-handed and the other more left-handed component. Therefore, if the solution is of a pure enantiomer, the plane of the plane-polarized light rotates clockwise or counterclockwise. The enantiomer that rotates the plane polarized clockwise is called dextrorotatory (abbreviated as d or (+)), and the other that rotates it counterclockwise is called levorotatory (abbreviated as l or (-)).

- The ability of chiral compounds to rotate the plane of plane-polarized light is called **optical activity**.
  - Chiral compounds are **optically active,** and achiral compounds are **optically inactive**.
- An instrument used to measure the ability of a compound to rotate the plane of a polarized light is called a **polarimeter**, illustrated in Figure 3.3.8.

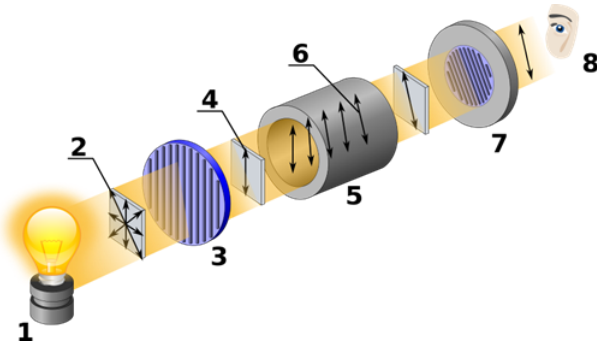


Figure 3.3.8: Illustration of the operating principle of a polarimeter: 1. light source, 2. unpolarized light, 3. Polarizer, 4. Plane polarized light, 5. the tube containing sample solution, 6. rotation of the plane of the polarized light, 7. Rotatable analyzer, 8. Detector. (Copyright; Kaidor, CC BY-SA 3.0, via Wikimedia Commons)

The rotation of the plane of plane-polarized light by an enantiomer is called **observed rotation**, measured in degrees. The observed rotation depends on the concentration of the sample, the path length of the light through the sample, the wavelength of the light used, and temperature.

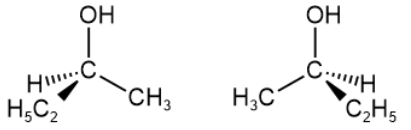
Specific rotation

When the concentration is fixed to 1 g/mL (or density of the pure sample in g/mL) and path length to 1 dm, the observed rotation is **specific rotation** $[\alpha]$.

Usually, the wavelength (λ) is 589 nm, i.e., sodium D light and temperature (T) is room temperature and shown as subscript and superscrpt: $[\alpha]^T_\lambda$, as in the following formula.

$$[\alpha]^T_\lambda = \frac{\text{Observed rotation (in degree)}}{\text{Path length (dm)} \times \text{Concentration (g/mL)}}$$

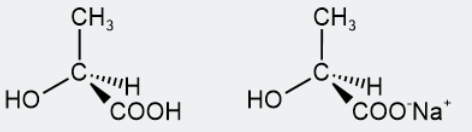
The specific rotation is a characteristic physical property of chiral compounds. If one enantiomer is dextrorotatory, the other is leavorotatory to the same degree, as shown in the following examples.



(S)-(+)-2-butanol $[\alpha]^{25}_D$ + 13.52

(R)-(-)-2-butanol $[\alpha]^{25}_D$ - 13.52

S and R define the absolute configuration of the chiral center, and dextrorotatory (+) or levorotatory (-) is an experimental property of the compound. A compound with S configuration may be dextrorotatory (+) or levorotatory (-), the same for R configuration. For example, (S)-lactic acid is dextrorotatory (+), and its salt (S)-sodium lactate is levorotatory (-), as shown below.



(S)-(+)lactic acid (S)-(-)-sodium lactate

Racemic mixture

Mixing a chiral compound with its enantiomer diminishes its optical activity. A 50:50 mixture of enantiomers is optically inactive called a **racemic mixture**.

Compounds with more than one chiral centers

The maximum number of stereoisomers possible for chiral compounds is given by the formula 2^n, where n is the number of chiral centers. For example, glyceraldehyde shown in Figure 3.1.1 has one chiral center and (2^1= 2) two stereoisomers: D-glyceraldehyde and L-glyceraldehyde. Erythrose, i.e., another monosaccharide, has two chiral centers and (2^2= 4) four stereoisomers, as shown in Figure 3.3.9.

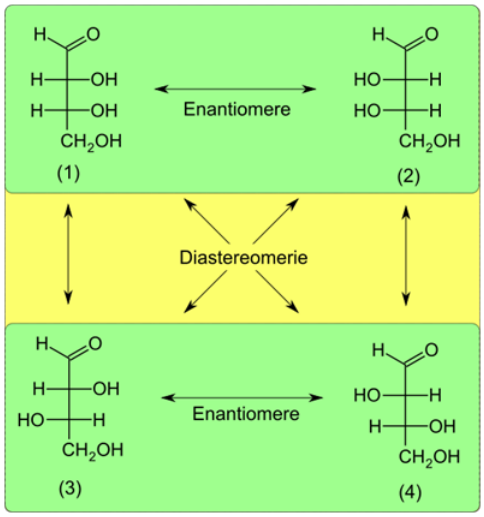


Figure 3.3.9: Illustration of enantiomer and diastereomer relationship among stereoisomers: **1)** D-erythrose, **(2)** L-erythrose, **(3)** L-threose, and **(4)** D-threose. (Copyright; Roland.chem, CC0, via Wikimedia Commons)

D-erythrose and L-erythrose are enantiomers, and D-threose and L-threose are also enantiomers, but erythrose is not an enantiomer of threose.

Diastereomers

Stereoisomers that are not enantiomers are **diastereomers** of each other. For example, D-erythrose is a diastereomer of D-threose and L-threose, as illustrated in Figure 3.3.9.

The formula 2^n tells the maximum possible number of stereoisomers of compounds having n number of chiral centers. The actual number may be less than the maximum. For example, tartaric acid has two chiral centers and three stereoisomers: (+) tartaric acid and (-) tartaric acid, which are enantiomers and achiral stereoisomers that is diastereomers of the first two, as illustrated in Figure 3.3.10. The third isomer is achiral because its mirror image is identical, simply rotated by a 180°.

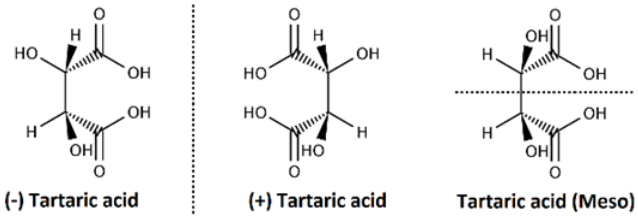


Figure 3.3.10: Illustration of enantiomers of tartaric acid and their achiral (meso) stereoisomer. The dotted line represents a mirror plane between enantiomers and the mirror plane, i.e., the symmetry plane in the meso isomer. (Copyright; Public domain)

Meso compound

A meso compound is an achiral compound with two or more chiral centers. A meso compound has a plane of symmetry that divides into two mirror-image halves, as illustrated in Figure 3.3.10.

Enantiomers have the same physical and chemical properties, except when interacting with chiral objects or reagents. Diastereomers have different physical and chemical properties, as can be observed in Table 1.

Table 1: Some of the physical and chemical properties of enantiomers ((+)-tartaric acid and (-)-tartaric acid) and their diastereomer (meso tartaric acid)

Stereoisomer	(+)-Tartaric acid	(-)-Tartaric acid	Meso tartaric acid
Specific rotation	+12.7	-12.7	0
Melting point (°C)	171-174	171-174	146-148
Density at 20°C (g/mL)	1.7598	1.7598	1.660
Solubility in water at 20°C (g/100 mL)	139	139	125
pK_1 at 25°C	2.98	2.98	3.23
pK_2 at 25°C	4.34	4.34	4.82

Chirality in life

A large number of compounds in living things are chiral. For example, amino acids and carbohydrates are chiral. Macro-molecules like proteins and nucleic acids are also chiral. Chirality is also observed on a macro-scale, e.g., our hands, foot, and ears are chiral. Similarly, horns and spots in axis deer are chiral, as shown in Figure 3.3.11.

Figure 3.3.11: Chiral spotted male axis deer (*Axis axis*) at Wild Florida Safari Park. (Copyright; Nosferattus, CC0, via Wikimedia Commons)

Stereospecific reactions in living things

Most chemical reactions in living things are stereospecific, i.e., one stereoisomer selectively reacts, and one is selectively produced. For example, the human digestion enzyme chymotrypsin has 268 chiral centers and 2^{268} possible stereoisomers, but only one isomer is produced. This is because amino acids that constitute chymotrypsin and other enzymes have L-configuration. The enzyme-catalyzed reactions are often stereospecific because their binding sites are chiral, and only one enantiomer reactant fits in them nicely for the reaction, as illustrated in Figure 3.3.12 Similarly, one specific surface of enzyme-bound substrates is exposed for the reaction that results in the selective production of one enantiomer.

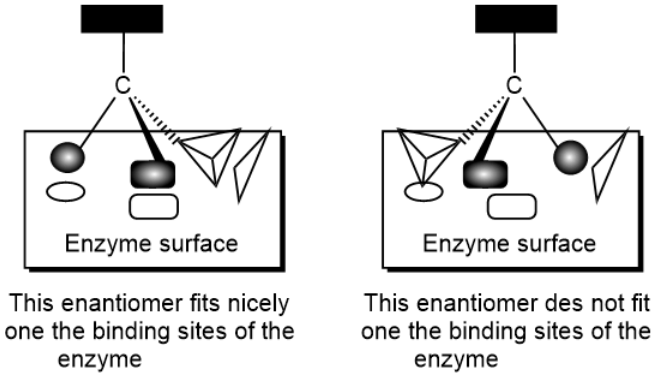
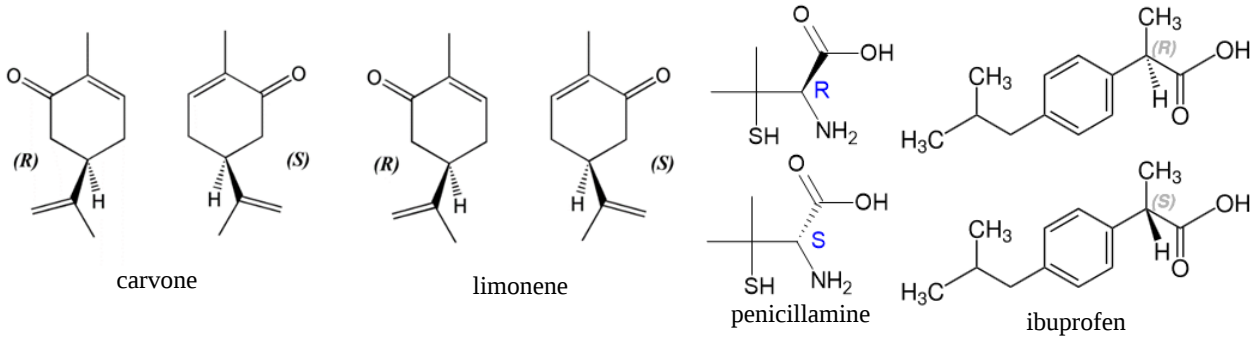


Figure 3.3.12: Illustration of stereoselective enzyme binding sites, taste, order, and drug receptors. (Copyright; Public domain)

Chirality manifests in taste, flavor, odor, and drug action as their receptors have chiral binding sites, as illustrated in Figure 3.3.12 Examples include:

- L-aspartame tastes sweet and is used as an artificial sweetener but D-aspartame is tasteless,
- R-(-)-carvone smells like spearmint, and S-(+)-carvone smells like caraway,
- R-(+)-limonene smells like orange and lemon, and S-(-)-limonene smells like a spruce tree,
- (S)-penicillamine has antiarthritic activity, and (R)-penicillamine is toxic, and
- S-ibuprofen has anti-inflammatory action, and R-ibuprofen is inactive.



CHAPTER OVERVIEW

4: Organic reactions

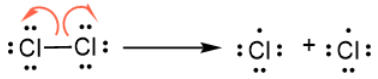
4.1: What is a reaction mechanism

What is a chemical reaction?

Chemical reactions involve rearranging atoms in a substance or substances called reactants resulting in new substances called products. In other words, chemical reactions involve breaking and/or making bonds. There are two major ways of breaking and making bonds, i.e., hemolytic and heterolytic bond breaking and making.

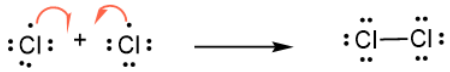
Hemolytic bond breaking and making

A covalent bond is a shared pair of electrons. In hemolytic bond breaking or hemolytic cleavage, the two elections are divided equally between the products as represented by the following reaction of a chlorine molecule (Cl_2) splitting into two chlorine atoms.



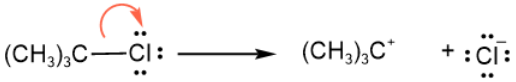
Recall that the half-headed curvy arrow represents the movement of one electron, pair of dots represents lone pair of electrons, a line represents a pair of bonding electrons, and a single dot represents an unpaired valance electron. A specie with an unpaired valence electron is called a free radical, e.g., $:\dot{\underset{\cdot\cdot}{Cl}}:$ in the above reaction is a free radical. Free radicals are usually reactive species.

The Reverse of the above reaction is a hemolytic bond making where two reactants contribute one electron each to make a covalent bond, as shown in the following reaction between two chlorine atoms.

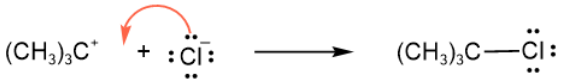


Heterolytic bond breaking and making

Heterolytic bond breaking happens so that a shared pair of electrons in the covalent bond are retained by one of the bonded atoms decreasing the charge by one, and the other bonded atom loses the shared electron increasing the charge by one, as shown in the following reaction.



Note that a regular curvy arrow represents the movement of two electrons, in this case, a bonding pair of electrons ending up as the fourth lone pair on the chlorine atom. Usually, the more electronegative atom receives the bonding pair of electrons and the less electronegative atom loses it in the heterolytic bond breaking. The reverse of the above reaction is heterolytic bond making, where the bonding pair of electrons are donated by one of the reactants, as shown in the following reaction.



Reaction mechanism

Chemical reactions often involve more than one bond-making and/or bond-breaking event. Further, reactions often happen in a sequence of steps that add up to yield the overall reaction equation. Individual reaction steps in the sequence are called **elementary reactions**.

Description of the step-by-step sequence of elementary reaction by which the overall chemical change occurs is called a **reaction mechanism**. There are two major reaction mechanisms, i) free radical reaction mechanisms that involve hemolytic bond breaking and making and ii) polar reaction mechanisms that involve heterolytic bond breaking and making. Some of the important reaction mechanisms are described in the following sections.

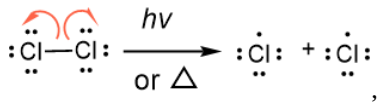
4.2: Free Radical Reactions

Learning Objectives

- Understand free radical reaction mechanisms, including elementary reactions of its three main steps: initiation, propagation, and termination.
- Learn exams of free radical reactions in industry and daily life: converting alkanes to haloalkanes, cracking process, combustion, polymerization, and the aging process.

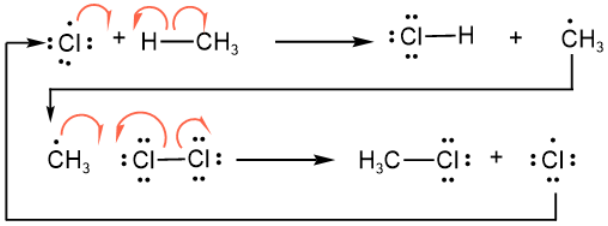
Free radical reaction mechanism

Hemolytic bond breaking requires energy equal to the bond dissociation energy, and hemolytic bond making releases energy equal to the bond dissociation energy. When a molecule absorbs energy in the form of heat or a photon of UV light, some of the weak bonds, like $\mathrm{Cl-Cl}$ bond (239 kJ/mol) or $\mathrm{O-O}$ bond (146 kJ/mol) breaks initiating free radicals, e.g.:



where *hv* represents a photon of UV light and Δ represent heat energy. The reaction in which free radicals are created from neutral species, as in the above example, is called the **free radical initiation** step in a free radical reaction mechanism.

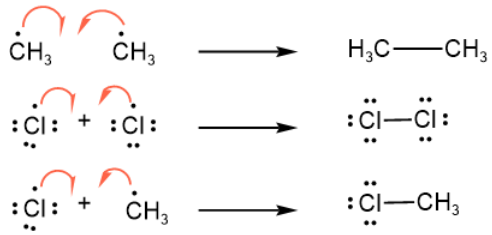
Free radicals are very reactive species. They usually abstract an atom, e.g., an H atom, from a hydrocarbon molecule, e.g.:



This reaction happens easily because the energy needed to break an $\mathrm{H-C}$ bond (413 kJ/mol) is compensated by the energy released by making $\mathrm{H-Cl}$ bond (427 kJ/mol). The $\dot{\mathrm{C}}\mathrm{H}_3$ repeats similar process when it collides with a Cl_2 molecule as shown in the second reaction above. Again, the energy released by making $\mathrm{C-Cl}$ bond (339 kJ/mol) compensates for the energy needed in breaking $\mathrm{Cl-Cl}$ bond (239 kJ/mol). Reactions in which one radical converts into a neutral specie and create another free radical that repeats the process, as in the above two reactions, are called **propagation reactions**. The propagation reactions often happen easily as the energy released in the bond-making compensates fully or partially for the energy needed in bond-making. The propagation reactions shown above happen in a cycle as the free radical produced in one is the reactant in the other and vice versa. The two propagation steps add up to the following overall reaction:

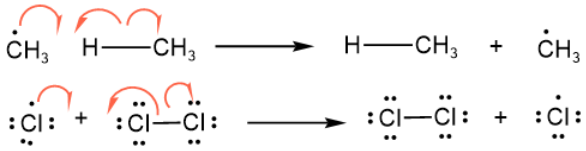
$$\text{Overall reaction: } \mathrm{CH_4 + Cl_2 \longrightarrow CH_3Cl + HCl}$$

This cycle of propagation reaction may repeat hundreds of times until one of the reactant exhaust or a free radical may collide with another free radical and terminate each other, as in the following reactions.



The homolytic bond-making between two free radicals that terminate the two radicals is called the **termination reaction**. The initiation, propagation, and termination are typical elementary reactions in a free radical reaction mechanism.

A $\dot{C}H_3$ may collide with CH_4 molecule and abstract a H atom and a $:\dot{\underset{\cdot\cdot}{\underset{\cdot\cdot}{Cl}}}:$ may collide with a Cl_2 molecule and abstract a Cl atom, but there is no net chemical change in these elementary steps as shown below.



Examples of free radical reactions

Conversion of alkanes to haloalkanes

Exposing a mixture of alkane and halogen (chlorine or bromine) to UV light or heat at ~100ºC converts alkanes to haloalkanes, as described in the free radical reaction mechanism for the case of CH_4 to CH_3-Cl conversion.

$$Cl_2 + CH_4 \xrightarrow{\text{UV or heat}} CH_3Cl + HCl$$

This reaction is useful in industrial organic synthesis for converting alkanes found in petroleum to alkyl halides that serve as intermediates in synthesizing organic compounds. The reaction is not very useful for laboratory organic synthesis as the product of the reaction competes with the initial alkane resulting in a mixture of products, as shown in the following reactions.

$$Cl_2 + CH_3Cl \xrightarrow{\text{UV or heat}} CH_2Cl_2 + HCl$$

$$Cl_2 + CH_2Cl_2 \xrightarrow{\text{UV or heat}} CHCl_3 + HCl$$

$$Cl_2 + CHCl_3 \xrightarrow{\text{UV or heat}} CCl_4 + HCl$$

The haloalkane product can not compete effectively when the initial alkane concentration is high. So, the side products are minimized in the industrial process by employing a higher concentration of alkane. This reaction converts relatively less reactive alkanes to more reactive haloalkanes having polar $\overset{\delta+}{C}-\overset{\delta-}{Cl}$ or $\overset{\delta+}{C}-\overset{\delta-}{Br}$ bond. The haloalkanes are used in the polar reactions described in later sections.

Combustion

Hydrocarbons found in petroleum, coal, and natural gas are primarily consumed in combustion processes for producing heat. In a combustion process, ground state O_2 is first converted to an excited state O_2^*, that strips a H off of an organic molecule in the free radical initiation step.

$$R-CH_3 + O_2^* \longrightarrow R-\dot{C}H_2 + H\dot{O}O$$

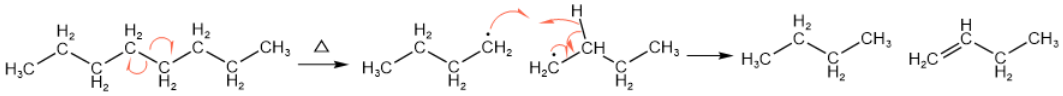
The free radicals, i.e., $R-\dot{C}H_2$ and $H\dot{O}O$ react further producing more radicals and, ultimately, convert the organic compounds to CO_2, H_2O, and heat. The combustion reaction of octane and ethanol are shown below as examples.

$$2\,CH_3CH_2CH_2CH_2CH_2CH_2CH_2CH_3 + 25\,O_2 \longrightarrow 16\,CO_2 + 18\,H_2O + \text{Heat}$$

$$CH_3CH_2OH + 3\,O_2 \longrightarrow 2\,CO_2 + 3\,H_2O + \text{Heat}$$

Cracking

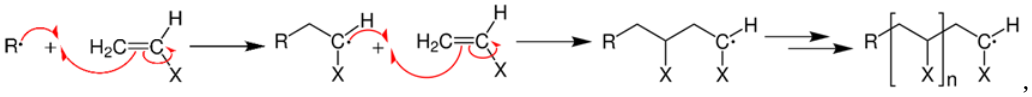
Although $C-C$ (347 kJ/mol) and $C-H$ (413 kJ/mol) bonds are stable at room temperature, at 450 ºC to 900 ºC there is enough thermal energy to break them homolytically. The resulting radical species react further and usually end up in smaller chain alkanes and alkenes, as illustrated in the figure on the right (Pengeldi, CC0, via Wikimedia Commons). This process is called thermal cracking. For example, one of the many reactions involved in the thermal cracking of octane is the following.



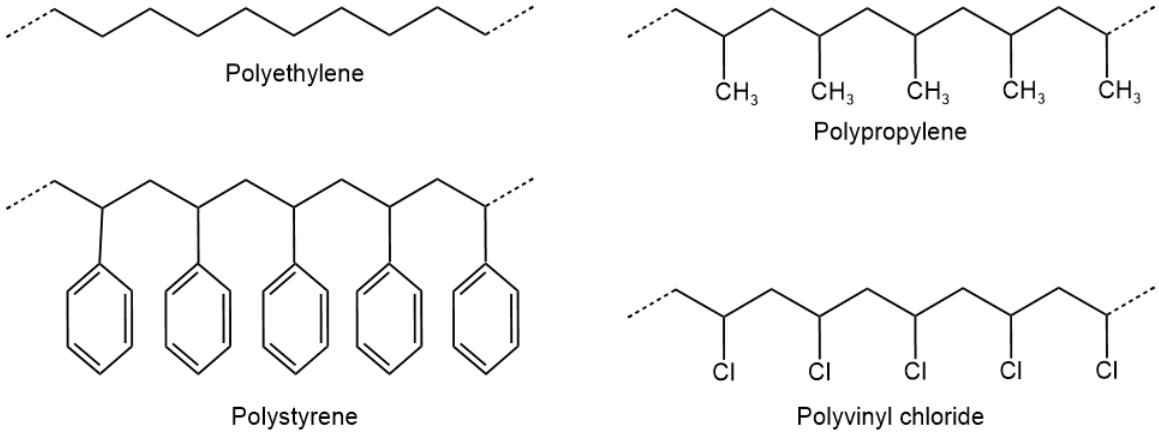
Catalysts and steam may be added to accelerate the sections and/or modify the nature of products. Cracking convert long-chain alkanes to more valuable small-chain alkanes and alkenes.

Polymerization

Polymers are long molecules composed of hundreds or thousands of repeat units called monomers. One polymer synthesis mechanism is free radical polymerization, involving initiation, propagation, and termination steps. The process initiates by homolytic cleavage of a weak bond, like $O-O$ bond in benzoyl peroxide ($C_6H_5COO-OOCC_6H_5$, by heat or UV-photon. The radical produced by the homolytic cleavage adds to a π bond of an alkene producing a new radical. The new radical adds to another alkene repeating the process in propagation steps, as illustrated below.



where X is H in polyethylene, CH_3 in polypropylene, benzene ring C_6H_5 in polystyrene, Cl in polyvinylchloride, etc., as illustrated below.



Ultimately, the radicals react with each other by hemolytic bond =-making that terminates the radicals.

Polyethylene is used in plastic bags, bottles, toys, piping, etc. Polypropylene is used in ropes, carpets, pipes, furniture, food containers, etc. Polystyrene is used in foam packing, building insulation foam, plastic cutlery, foam containers for food and drinks, etc. Polyvinylchloride is used in pipes, window and door frames, bank cards, cable insulation, etc. In polytetrafluorethylene or Teflon, all $H's$ and X in the monomer are $F's$. Teflon is used in non-stick coatings, lubricants, electrical insulation for wires, etc. Various objects of common use made of these polymers are illustrated in Figure $4.2.1$.

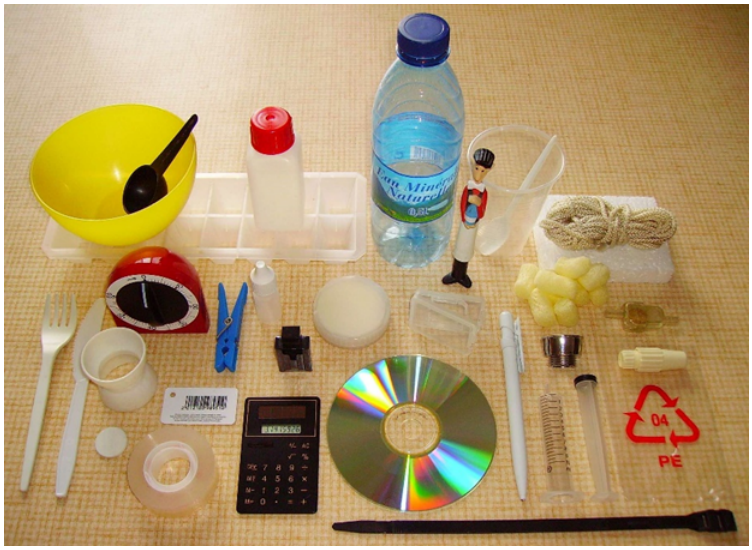
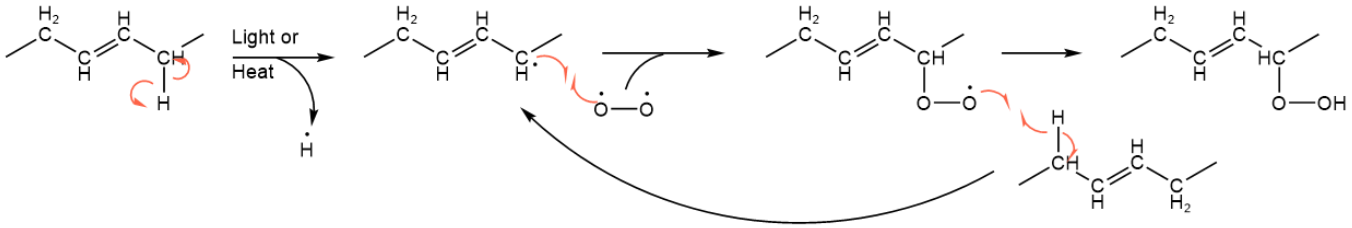


Figure 4.2.1: Common objects made of polymers like polyethylene, polypropylene, polystyrene, polyvinylchloride, etc. (Copyright; Cjp24, Public domain, via Wikimedia Commons)

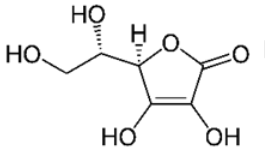
Free radical reactions in biological processes

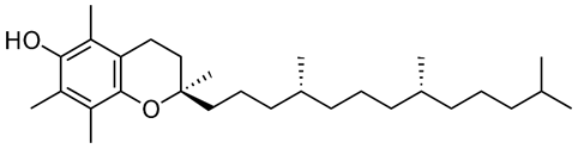
Free radical reactions are involved in the aging of food products. For example, fatty acids have long alkane chains, many of which have $C{=}C$ bonds. The allylic hydrogens in fatty acid chains are susceptible to hemolytic bond cleavage, as shown below.



Oxygen is a diradical that adds to the free radical in a hemolytic bond-making step, producing a new radical. The new radical abstracts allylic hydrogen from another fatty acid producing a radical that repeats the cycle, i.e., propagation steps. The propagation reactions install hydroperoxyl groups ($-O-OH$) on the fatty acids. The hydroperoxides are unstable and decompose to short-chain aldehydes and carboxylic acids responsible for the aged fats' unpleasant "rancid" smell. A similar process happens with the low-density lipoproteins deposited in arteries that lead to cardiovascular diseases. The aging process in organisms is also related to similar free radical reactions.

Living things have a mechanism of getting rid of unwanted free radicals by reacting them with radical scavengers. For example, vitamin C is a radical scavenger in the blood, and vitamin E is a radical scavenger in fats. The radical-scavengers neutralize the toxic radicals by donating H atoms from their $-OH$ groups that break the propagation chain reaction. The new radicals generated in the scavenging reactions are less reactive and easily excreted before they can do more damage.


Vitamin C


Vitamin E

4.3: Acid-base reactions

Learning Objectives

- Recognize acids and bases, understand the strength of acids and based, and predict the strength based on the structural features of the molecules.
- Understand simple acid-base reactions that are elementary steps in organic reaction mechanisms.
- Predict the direction of acid-base equilibrium based on the acid/base strength.
- Predict the ionization of acid/base functional groups in biochemicals under physiological conditions.

What is an acid-base reaction?

An acid-base reaction is a reaction in which a proton (H^+) is exchanged between reactants. For example, when acetic acid (CH_3COOH) is mixed with water, a proton is transferred from acetic acid to water, as shown in the reaction equation below.

$$CH_3COOH + H_2O \rightleftharpoons CH_3COO^- + H_3O^+$$

An acid donates a proton, and a base accepts a proton. CH_3COOH is an acid and H_2O is a base in the forward reaction. CH_3COO^- is a base and H_3O^+ is an acid in the reverse reaction. Two species that are related by the addition or removal of a proton are **conjugate acid-base pairs**. For example, CH_3COOH and CH_3COO^- are conjugate acid-base pair. Similarly, H_3O^+ and H_2O are conjugate acid-base pair. Since a proton has a +1 charge, its removal from an acid decreases the charge by one, and its addition to a base increases the charge by one. So, a general formula of an acid-conjugate acid-base pair could be represented as HB and B^-, e.g., CH_3COOH and CH_3COO^- or as HB^+ and B, e.g., H_3O^+ and H_2O. Acid-base reactions are generally fast and reversible, establishing equilibrium. That is why equilibrium arrows are shown separating reactants and products.

Strength of an acid and a base

Strength of an acid

The strength of an acid is a measure of its ability to donate a proton to a base. Often the solvent is water, and the reference base is H_2O, as in the following general reaction.

$$HA + H_2O \rightleftharpoons A^- + H_3O^+$$

The relationship of molar concentrations of reactants and products in an equilibrium reaction is expressed by the following formula:

$$K_{eq} = \frac{[A^-][H_3O^+]}{[HA][H_2O]}$$

,where square brackets indicate the molar concentration of the specie within the brackets and K_{eq} is a constant called the equilibrium constant. Since water is a solvent, its concentration $[H_2O]$ is almost a constant and merged with the constant to define a new constant, as shown below.

$$K_a = K_{eq}[H_2O] = \frac{[A^-][H_3O^+]}{[HA]}$$

, where K_a is a constant called acid dissociation constant that expresses the strength of the acid, the higher the value of K_a; the stronger the acid. Since K_a is often a large negative number, it is expressed on a log scale with its signed reversed, as shown in the following formula.

$$pK_a = -Log_{10}(K_a)$$

The pK_a is a quantitative measure of the acid strength, i.e., the smaller the pK_a means the large the K_a and stronger the acid.

Strength of a base

The strength of a base is a measure of its ability to accept a proton from an acid. Often the solvent is water, and the reference base is

H_2O, as in the following general reaction.

$$A^- + H_2O \rightleftharpoons HA^+HO^-$$

The relationship between molar concentrations of reactants and products in an equilibrium reaction is expressed by the following formula:

$$K_{eq} = \frac{[HA][HO^-]}{[A^-][H_2O]}$$

Again $[H_2O]$ is assumed to be a constant and merged with the constant to define a new constant, as shown below.

$$K_b = K_{eq}[H_2O] = \frac{[HA][HO^-]}{[A^-]}$$

, where K_b is a constant called base dissociation constant that expresses the strength of the base. the higher the value of K_b the stronger the base. again take a log and reverse the sign to arrive at a constant called pK_b as shown in the following formula.

$$pK_b = -Log_{10}(K_b)$$

The pK_b is a quantitative measure of the base strength, i.e., the smaller the pK_b means the large the K_b and stronger the base.

Relationship between pK_a and pK_b of conjugate acid-base pair

The K_a of an acid HA and K_b of a conjugate acid-base A^- pair are reciprocal of each other as proven below.

$$K_a \times K_b = \frac{[A^-][H_3O^+]}{[HA]} \times \frac{[HA][HO^-]}{[A^-]} = [HO^-][H_3O^+] = K_w = 10^{-14}$$

That rearranges to,

$$K_b = \frac{10^{-14}}{K_a}$$

It shows that if acid is a strong acid (large K_a), its conjugate base is a weak base ((small K_b), and vice versa. Taking the log and changing the sign on both sides leads to a relationship between pK_a and pK_b of conjugate acid-base pair, as shown below.

$$-Log_{10}K_b = -Log_{10}\left(\frac{10^{-14}}{K_a}\right)$$

$$pK_b = 14 - pK_a$$

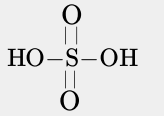
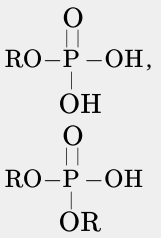
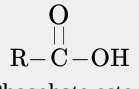
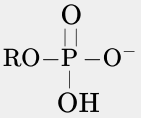
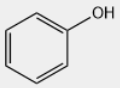
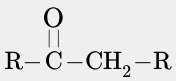
Therefore, pK_a values are tabulated in the reference books, and pK_b values, if needed, are calculated from the pK_a values of their conjugate acids using the above formula.

Factors affecting the strength of acids

The pK_a is a quantitative measure of acid strength, i.e., the smaller the pK_a, the stronger the acid. For example, acetic acid ($CH_3COOH, pK_a = 4 \cdot 76$) is a stronger acid than ethanol ($CH_3CH_2OH, pK_a = 15 \cdot 9$). Approximate pK_a of acids commonly encountered in organic chemistry are listed in Table 1.

Table 1: Approximate pKa values of acids commonly encountered in organic chemistry (Acidic protons are shown in the formula, and R– is an alkyl group or hydrogen)

pKa:	<0	2	5	10	15

Compounds:	Strong acids like sulfuric acid 	Phosphate esters 	Carboxylic acids  Phosphate esters 	Phenols  Ammonium ions $R-NH_3^+$ Thiols $R-SH$	Water H_2O Alcohols $R-OH$
pKa:	**20**	**25**	**35**	**45**	**>50**
Compounds:	Aldehydes and ketones 	Terminal alkynes $R-C\equiv CH$	Hydrogen $H-H$ Amines $R-NHR$	Alkenes $R_2C=CHR$	Alkanes $R-CH_2-R$

The major factors affecting the acid strength are the following.

Electronegativity

Within the same row of the periodic table, the more electronegative the atom to which the proton is bonded, the more acidic the proton is. Electronegativity (EN) increases from left to right in the same row, e.g., C, EN 2.6 < N, EN 3.0 < O, EN 3.4 < F, EN 4.0.

> The acid strength increases as the electronegativity of the atom carrying acidic proton increases within the same row of the periodic table, e.g., CH_4, pK_a ~60 < NH_3, pK_a ~38 < H_2O, pK_a 14.0 < HF, pK_a 3.2.

This effect of electronegativity explains why alcohols ($R-OH$, pK_a~16) are more acidic than amines ($R-NHR$, pK_a~40) and amines are more acidic than alkanes ($R-CH_3$, pK_a~60). This trend related to electronegativity does not hold when comparing elements from different rows.

Size of atom

The electronegativity decreases from top to bottom of a periodic table, e.g., for halogens: F, EN 3.98 > Cl, EN 3.16 > Br, EN 2.69 > I, 2.66. . The size of atoms has an opposite trend, i.e., the size increases from top to bottom of the periodic table, e.g., the size of halogens varies in this order: F < Cl < Br < I.

> The acid strength increases as the size of the atom carrying the acidic proton increases in the same period, e.g., HF, pK_a 3.2 < HCl, pK_a -7 < HBr, pK_a -8 < HI, pK_a -9.9.

This is because as the size of the atom increases, its bond with a proton becomes weaker and easier to break. Further, the negative charge a proton leaves behind is more easily stabilized when spread over a larger atom than a smaller atom. The effect of the size of an atom on acid strength is opposite to that of electronegativity within the same period.

> The effect of the size of the atom bearing acidic proton on acidity is dominant over the effect of electronegativity.

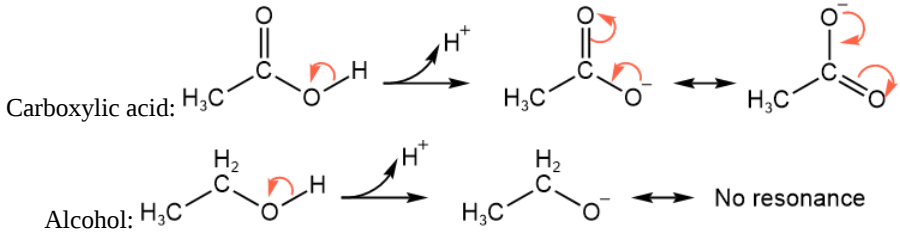
The effect of size explains why thiols ($R-SH$, pK_a~10) are stronger acids than alcohols ($R-OH$, pK_a~16).

Resonance

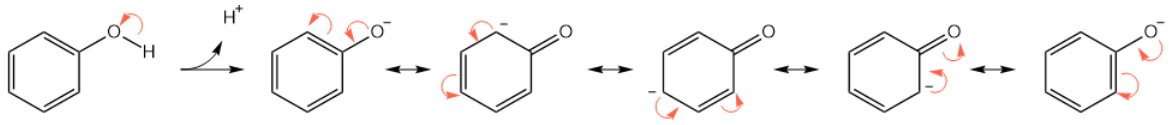
When a proton leaves an acid HA, it usually leaves a negative charge on the conjugate base A^-. Resonance can stabilize the negative charge by distributing it on more than one atom.

> An acid with conjugated base resonance stability can let go of its acidic proton more easily. It is stronger than an equivalent acid with no resonance stabilization of its conjugate base.

For example, both carboxylic acids and alcohols have acidic proton on an $-\mathrm{OH}$ group, but carboxylic acids are stronger acids ($\mathrm{pK_a}$, ~5) than alcohols ($\mathrm{pK_a}$, ~16) due to resonance stabilization of the negative charge on their conjugate bases, as illustrated below.

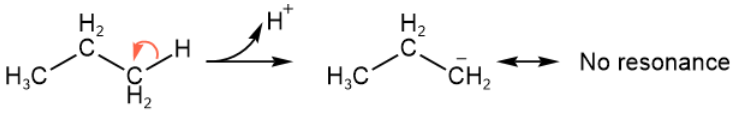


Phenols are also stronger acids ($\mathrm{pK_a}$, ~10) than alcohols for the same reason, as shown below.



However, phenols are weaker acids than carboxylic acids because the negative charge is shared by more electronegative O in carboxylic acids than by less electronegative C in phenols.

Protons on $\alpha\mathrm{C}$ to a $\mathrm{C{=}O}$ group, as in aldehydes, ketones, and carboxylic acid derivatives, are more acidic ($\mathrm{pK_a}$, ~20) than protons on alkanes ($\mathrm{pK_a}$, ~60) due to the resonance effect, as shown below.



Hybridization

Electrons in an s-orbital are nearer to and more attracted by the nucleus than in a p-orbital of the same shell. This is because of the spherical shape of the s-orbital placing electrons nearer to the nucleus versus the dumbbell shape of the p-orbital in the same shell. Therefore, the valence electrons in an sp-orbital having 50% s-character are attracted more to the nucleus than an sp^2 orbital with 33% s-character and an sp^3 orbital with 25% s-character. Recall that electronegativity is a measure of the ability of a nucleus to attract valence electrons.

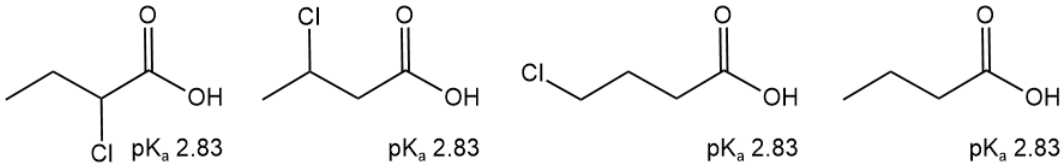
> The electronegativity of an atom and, consequently, the acidity of a proton attached to it increase with a change in hybridization in this order:$sp^3 < sp^2 < sp$.

For example, the acidity of protons of sp^3-hybridized ethane, sp^2-hybridized ethene, and sp-hybridized ethyne increases in this order: $\mathrm{CH_3CH_3}$ pK_a 51 < $\mathrm{CH_2{=}CH_2}$ pK_a 44 < $\mathrm{CH \equiv CH}$ pK_a 25.

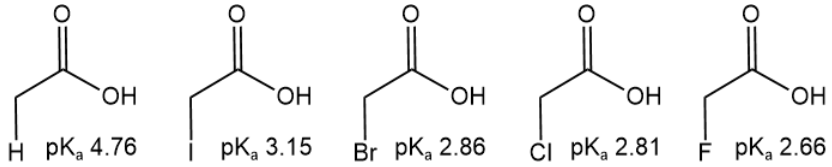
Inductive effect

> The electron-withdrawing effect of an electronegative atom, i.e., the inductive effect increases the acidity of the proton adjacent to it.

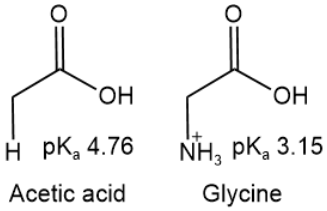
- Near the electronegative atom stronger, the effect, as illustrated in the example below where Cl is the electronegative atom affecting the acidity of $-\mathrm{COOH}$ proton.



- The more electronegative the atom stronger the effect, as illustrated in the example below, where the electronegativity increases in this order: $\overrightarrow{\mathrm{H} < \mathrm{I} < \mathrm{Br} < \mathrm{Cl} < \mathrm{F}}$.



Amino acids are more acidic than carboxylic acids due to the inductive effect of ammonium ion, e.g., glycine is more acidic than acetic acid as shown below.



Charge

> A proton H^+ on a positively charged specie is easier to remove and, consequently, more acidic than on the same species in a neutral state.

The acidity of the following species reflects this: $\overrightarrow{\mathrm{H_2O, pKa\ 15 \cdot 7 < H_3O^+, pKa\ -1 \cdot 74}}$ and $\overrightarrow{\mathrm{NH_3, pKa\ 38 < NH_4^+, pKa\ 9 \cdot 24}}$. The effect of the charge is reduced but still exists when the charge is on an atom other than the atom bearing the acidic proton. Negatively charges species, conversely, are less acidic than the same specie in a neutral or positively charged state, as illustrated below for the case of phosphoric acid species.

$$\mathrm{HO{-}\overset{\overset{\mathrm{O}}{\|}}{\underset{\underset{\mathrm{OH}}{|}}{\mathrm{P}}}{-}OH,\ pKa\ 2 \cdot 14 > {}^{-}O{-}\overset{\overset{\mathrm{O}}{\|}}{\underset{\underset{\mathrm{OH}}{|}}{\mathrm{P}}}{-}OH,\ pKa\ 7 \cdot 20 > {}^{-}O{-}\overset{\overset{\mathrm{O}}{\|}}{\underset{\underset{\mathrm{O^-}}{|}}{\mathrm{P}}}{-}OH,\ pKa\ 12 \cdot 37}$$

Examples of organic acid-base reactions

Reactions of organic acids

Carboxylic acids are organic acids. The following example shows that when a carboxylic acid is dissolved in water, it dissociates by donating its proton to water.

$$\mathrm{CH_3COOH + H_2O \rightleftharpoons CH_3COO^- + H_3O^+}$$

The reaction is written as an equilibrium because it goes in both directions, forward and reverse. acetic acid $\mathrm{CH_3COOH}$ is the acid in reactants and hydronium ion $\mathrm{H_3O^+}$ is the acid in products.

> The general rule is that a stronger acid goes to a weaker one in an acid-base reaction. It is equally correct to say a stronger base goes to a weaker base.

For example, it is more accurate to say that the above reaction is reverse directed, or there is more concentration of reactants than the products at equilibrium, because H_3O^+ is a stronger acid than CH_3COOH. This fact is represented by a longer equilibrium arrow in the direction of weaker acid in the equilibrium, as shown below.

$$\underset{\text{pKa 4·76}}{CH_3COOH} + H_2O \rightleftharpoons CH_3COO^- + \underset{\text{pKa -1·74}}{H_3O^+}$$

This reaction is made strongly forward directed by using a strong base like ^-OH that produces a weaker acid H_2O, making the reaction almost irreversible, as shown below.

$$\underset{\text{pKa 4·76}}{CH_3COOH} + {}^-OH \longrightarrow CH_3COO^- + \underset{\text{pKa 14·0}}{H_2O}$$

This is an example of an acid-base neutralization reaction. The ^-OH ion is obtained by dissolving sodium hydroxide $NaOH$ in water, but Na^+ is a spectator ion, i.e., it does not take part in the reaction and is usually not shown in the equation. As shown below, other organic acids, like phenols and thiols, do similar acid-base neutralization reactions.

$$\underset{\text{pKa 9·95}}{C_6H_5OH} + {}^-OH \longrightarrow C_6H_5O^- + \underset{\text{pKa 14·0}}{H_2O}$$

$$\underset{\text{pKa 10·4}}{CH_3SH} + {}^-OH \longrightarrow CH_3S^- + \underset{\text{pKa 15·7}}{H_2O}$$

Terminal alkyne has acidic protons but they are weak acids ($pKa \sim 25$). A base like ^-OH is not a strong enough base to remove all of the alkyne protons, as shown by the following equilibrium reaction.

$$\underset{\text{pKa 25}}{R-C\equiv CH} + {}^-OH \rightleftharpoons R-C\equiv C^- + \underset{\text{pKa 14·0}}{H_2O}$$

An amide anion ($^-NH_2$ in the form of $NaNH_2$) is a sufficiently strong base to neutralize alkyne, as shown below.

$$\underset{\text{pKa 25}}{R-C\equiv CH} + {}^-NH_2 \longrightarrow R-C\equiv C^- + \underset{\text{pKa 38}}{NH_3}$$

Alcohols ($R-OH$, $pKa \sim 25$) are also weak acids and not fully neutralized with bases like ^-OH, as shown below.

$$\underset{\text{pKa 16}}{R-OH} + {}^-OH \rightleftharpoons R-O^- + \underset{\text{pKa 14}}{H_2O}$$

A hydride ion (H^- in the form of NaH) is a sufficiently strong base to neutralize alcohols, as shown below.

$$\underset{\text{pKa 16}}{R-OH} + H^- \longrightarrow R-O^- + \underset{\text{pKa 35}}{H_2}$$

Reactions of organic bases

Amines ($R-NH_2$) are organic bases that produce ^-OH ions in water, as shown below.

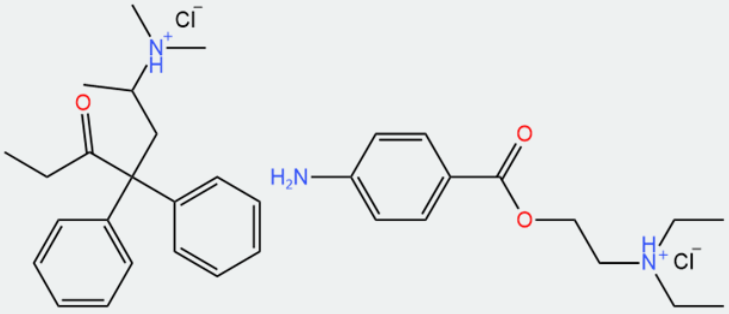
$$\underset{\text{pKa 15·7}}{H_2O} + R-NH_2 \rightleftharpoons {}^-OH + \underset{\text{pKa 10}}{R-NH_3^+}$$

Amines neutralize strange acids like (HCl) and produce ammonium salts, as shown below.

$$\underset{\text{pKa -7}}{HCl} + R-NH_2 \longrightarrow R-NH_3^+\,Cl^-$$

Ammine.HCl drugs -increasing shelf-life and solubility in body fluids

Many drugs contain amine functional groups. Amines are usually unstable and insoluble in water. Neutralizing the amine ($R-NH_2$) with HCl converts them to ammonium chloride salts ($R-NH_3^+\,C^-$), also called amine.HCl salts ($R-NH_2\cdot HCl$), which are water-soluble and more stable. Therefore, drugs containing amine functional groups are usually sold in ammonium chloride form, making them soluble in body fluids like blood plasma and increasing their shelf-life. For example, methadone, a narcotic analgesic, and procaine, a local anesthetic, are marketed as hydrochloride salts, as shown below.



Methadone.
HCl

Procaine.
HCl
(also called novocain)

Alcohols are amphoteric substances like water, i.e., they can donate a proton as acid and accept a proton as bases. Like alcohols, the oxygen of carbonyl (C=O group also has two lone pairs and can accept a proton. A strong acid like HCl or H_2SO_4can protonate an alcohol or a carbonyl group, as shown below.

$$\underset{\text{pKa -7}}{HCl} + R-OH \rightleftharpoons Cl^- + \underset{\text{pKa ~-3}}{R-OH_2^+}$$

$$\underset{\text{pKa -9}}{H_2SO_4} + R-\overset{\overset{O}{\|}}{C}-R \rightleftharpoons Cl^- + \underset{\text{pKa ~-7}}{R-\overset{\overset{+OH}{\|}}{C}-R}$$

The acid-base reactions that produce cations by accepting protons, like $R-OH_2^+$ or $R-\overset{\overset{+OH}{\|}}{C}-R$ and anions by donating protons, like $R-O^-$, $R-C\equiv C^-$, $R-S^-$, etc. are important intermediates in organic reactions described in the later sections.

Ionization of acidic and basic functional groups at physiological pH

The Henderson-Hasselbach equation for an acid (HA) is:

$$\text{Henderson-Hasselbach equation: } pH = pK_a + Log_{10}\left(\frac{[A^-]}{[HA]}\right)$$

, where $pH = -Log_{10}[H_3O^+]$, and $pK_a = -Log_{10}K_a$. For a base (B) is:

$$pH = pK_a + Log_{10}\left(\frac{[HB^+]}{[B]}\right)$$

, where pKa is that of the conjugate acid HB^+. The Henderson-Hasselbach equation helps estimate the ratio of acid (HA) to its conjugate base ((A^-) and of base (HB) to its conjugate acid ((B^+) at the pH of the medium. According to Henderson-Hasselbach equaiton,

- for an acid (HA), if pH of medium is two units lower than the pKa of the acid, it will be present substantially in its de-protonated form A^-, and
- for a base (B), if pH of medium is two units higher than the pKa of its conjugated acid HB^+, it will be present substantially in its protonated form HBA^+.

Following conclusions are drawn by applying the above rules to the functional groups commonly found in cells.

- The pH of cells is generally between 7 and 8.5. Carboxylic acids ($R-\overset{\overset{O}{\|}}{C}-OH$) have pKa 4 to 5. So, carboxylic acids exist in an ionized form $R-\overset{\overset{O}{\|}}{C}-O^-$ in a physiological medium.

- The conjugate acid forms of amines $R-NH_3^+$ have pK_a 9 to 11. So, amines exist in the ionized form $R-NH_3^+$ in a physiological medium.
- Amino acids $R-\underset{\underset{NH_2}{|}}{CH}-\overset{\overset{O}{||}}{C}-OH$ have amine and carboxylic acid groups. Amino acids exist as $R-\underset{\underset{NH_3^+}{|}}{CH}-\overset{\overset{O}{||}}{C}-O^-$. A specie with separate positive and negative charge groups on it is called a zwitterion. Amino acids exist as a zwitterions.
- Phosphodiesters $RO-\overset{\overset{O}{||}}{\underset{\underset{OR}{|}}{P}}-OH$ present in DNA and RNA molecules have pK_a 1 to 3. So, phosphodiester also exist in an ionized form $RO-\overset{\overset{O}{||}}{\underset{\underset{OR}{|}}{P}}-O^-$ in a physiological medium.
- The conjugate acid forms of amines, i.e., ammonium ion $R-NH_3^+$ have pK_a 9 to 11. So, amines exist as ammonium ions $R-NH_3^+$ in a physiological medium.

Derevation of Henderson-Hasselbalch equation

The equation is derived by performing the following steps: i) re-arranging K_a expression:

$$K_a=\frac{[A^-][H_3O^+]}{[HA]}$$

$$[H_3O^+]=K_a\times\frac{[HA]}{[A^-]}$$

Taking the log of both sides and reversing the sign:

$$-Log_{10}[H_3O^+]=-Log_{10}(K_a\times\frac{[HA]}{[A^-]})$$

$$-Log_{10}[H_3O^+]=-Log_{10}K_a-Log_{10}(\frac{[HA]}{[A^-]})$$

making a substitution for pK_a, and pH:

$$\boxed{\text{Henderson-Hasselbach equation: } pH=pK_a+Log_{10}(\frac{[A^-]}{[HA]})}$$

, where $pH=-Log_{10}[H_3O^+]$, and $pK_a=-Log_{10}K_a$.

A similar derivation of a general base B and its conjugate acid HB^+ leads to this equation:

$$\boxed{pH=pK_a+Log_{10}(\frac{[HB^+]}{[B]})}$$

, where pKa is that of the conjugate acid HB^+.

4.4: Nucleophilic substitution and elimination reactions

Learning Objectives

- Identify nucleophiles and electrophiles, their similarities, and differences from acids and bases.
- Understand S_N2, E2, S_N1, or E1 reaction mechanisms, factors that affect them, and conditions that dictate which one will take place.
- Learn some examples of S_N2, E2, S_N1, and E1 reactions of alcohols, ethers, amines, and thiols.

Nucleophile and electrophile

Nucleophile is a neutral or anionic specie that can donate a lone pair or π bonding electrons to make a covalent bond.

Examples of nucleophile are negative or partial negative ($\delta-$) atoms with lone pair or π bond in the following $H_3N:$, $H_2\ddot{O}:$, $H\ddot{O}:^-$, and $H_2C=CH_2$.

Electrophile is an electron-deficient atom of a neutral or cationic species that can receive an electron pair to make a covalent bond.

Examples of electrophile are positive charge or partial positive ($\delta+$) charge C′s or H′s in the following species: $(CH_3)_3\overset{+}{C}$, $\overset{\delta+}{C}H_3-\overset{\delta-}{Cl}$, $(CH_3)_2\overset{\delta+}{C}=\overset{\delta-}{O}$, and $\overset{\delta+}{H}-\overset{\delta-}{Cl}$.

Differences between nucleophile-electrophile and acid-base

The base is a substance that donates a pair of electrons to a proton to make a covalent bond. So, a base is a sub-class of nucleophiles and an acidic proton is a subclass of electrophiles. Acid-base reactions are fast reactions. In acid-base reactions, emphasis is on thermodynamic, specifically on the equilibrium constant K quantified in terms of pKa, i.e., smaller the pKa means larger K and stronger acid. Nucleophiles donate electrons to an electron-deficient carbon or some atom usually other than a proton. Emphasis in nucleophilic-electrophilic reaction is on the kinetics, a good nucleophile reacts faster and a poor nucleophile reacts slower.

Similarities and differences in nucleophilicity and basicity

The similarity is that the stronger bases are stronger nucleophiles when compared within the same row of the periodic table. For example, basicity follows the order $\xrightarrow{\qquad\qquad} HO^- < NH_2^- < F^-$ and their nucleophilicity follows the same order. Similarly, anionic species, e.g., NH_2^-, HO^- and RO^- are stronger bases and good nucleophiles compared to the corresponding neutral species of the same element, i.e., NH_3, H_2O and ROH.

The difference is that the basicity and nucleophilicity are affected differently when comparing atoms of different rows. For example, hydride ion (H^-) from the 1st row is a stronger base but a poor nucleophile. the opposite is true for the 3rd-row and higher elements, e.g., SH^-, Cl^-, Br^-. and I^- are week bases but usually good nucleophiles. Unlike basicity, nucleophilicity is affected by steric factors and solvents. Most of these differences are related to the fact that in acid-base reactions, the electron recipient is the 1s orbital of a proton, and in nucleophilic-electophilic reactions, the electron recipient is 2s, 2p, or larger orbital of carbon or some other element. The details of these factors are out of the scope of this book. Base strength is described as stronger or weak, and nucleophilic strength is described as good or poor.

Nucleophilic substitution mechanisms

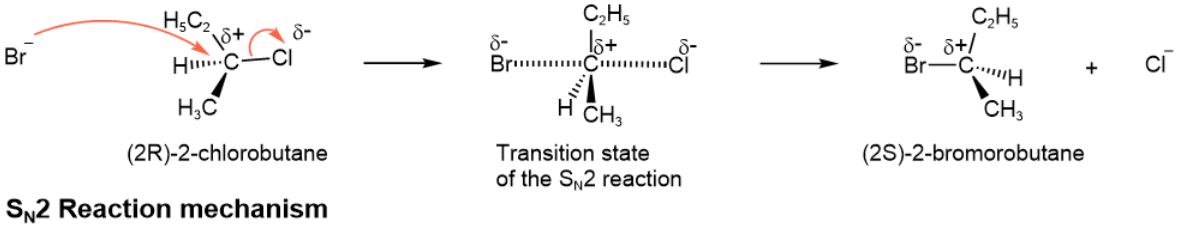
In a nucleophilic substitution reaction, a nucleophile ($Nu:^-$) attacks and makes a covalent bond with $\delta+$ atom of the target molecule, called **substrate** ($\overset{\delta+}{R}-\overset{\delta-}{X}$). The polar bond of the target molecule breaks heterolytically, leaving the bonding electrons with the more electronegative end called **leaving group** ($X:^-$), as shown in the following generalized reaction.

$$\ddot{Nu}^{-} + \overset{\delta+}{R}-\overset{\delta-}{X} \longrightarrow \overset{\delta+}{R}-\overset{\delta-}{Nu} + \ddot{X}^{-}$$

Since one nucleophile, i.e., $\ddot{Nu}^{-}$, replaces another nucleophile, i.e., $\ddot{X}^{-}$ on the substrate, it is called a **nucleophilic substitution reaction**. The nucleophilic substitution reactions are based on the fact the incoming nucleophile has a stronger tendency to make a covalent bond with the electrophilic center than the leaving group. In other words, the incoming nucleophile is stronger and the leaving group is a weaker nucleophile. Mechanisms of the nucleophilic substitution reactions are described below.

Nucleophilic substitution bimolecular (S_N2)

One mechanism for nucleophilic substitution reaction is concerted bond-making and breaking in a single step, as shown below.



S_N2 Reaction mechanism

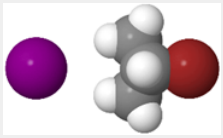
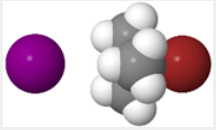
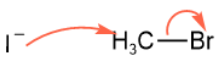
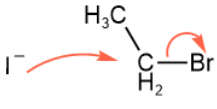
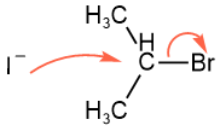
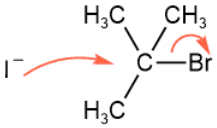
The incoming nucleophile approaches the electrophilic $\overset{\delta+}{C}$ from the side opposite to the leaving group. The incoming nucleophile starts making the bond and the leaving group starts breaking the bond simultaneously. The other three groups pointing away from the leaving group, start moving to the other side, away from the incoming nucleophile. In the transition state, the three groups acquire a trigonal planar arrangement perpendicular to the bond being broken and formed. This reaction mechanism is called S_N2, where S is for substitution, N is for nucleophilic, and 2 is for bimolecular.

- If the electrophilic $\overset{\delta+}{C}$ is a chiral center, its configuration is inverted in the product of S_N2 reaction.
- This elementary step of S_N2 reaction involves two reactants in its rate-determining step, the incoming nucleophile, and the substrate, and it is a bimolecular reaction.

The rate of an S_N2 reaction is directly proportional to the substrate concentration and the nucleophile concentration. The rate depends on the substrate's structure, the nature of the leaving group, the nature of the nucleophile, and the solvent, as described next.

Effect of the structure of the substrate

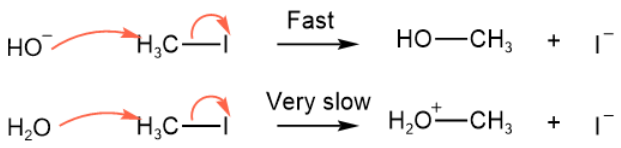
The rate follows the following order with respect to the nature of the electrophilic $\overset{\delta+}{C}$: $\xleftarrow{\text{methyl} > \text{primary} > \text{secondary} > \text{tertiary}}$. The steric hindrance posted by the substrate to the nucleophile explains it, which follows the same order, as illustrated below.

Type of electrophile carbon	Methyl	Primary	Secondary	Tertiary
Model				
Structures with mechanism arrows	 bromomethane	 bromoethane (primary)	 2-bromopropane (secondary)	 2-bromo-2-methyl-propane (tertiary)

The relative S_N2 reaction rate	145	1	0.008	no reaction

Effect of the nucleophile

Good nucleophiles react faster. Particularly, anionic nucleophiles, like HO^- or RO^- are employed for S_N2 as they react much faster than their corresponding neutral counterparts, i.e., H_2O or ROH, as shown in the following example.



Effect of the leaving group

Leaving group propensity has a trend opposite to the basicity of the leaving group, i.e., good leaving groups are weaker bases. For example, the basicity of halide ions follows this order: $\overleftarrow{F^- > Cl^- > Br^- > I^-}$ but the rate of reaction of alkyl halides, under the same conditions, follows the opposite trend, i.e., $\overrightarrow{R-F < R-Cl < R-Br < R-I}$. The effect of the leaving group on the rate of the reaction is shown in the following table.

Reaction	Relative rate
$HO^- + RCH_2-F \longrightarrow RCH_2-OH + F^-$	1
$HO^- + RCH_2-Cl \longrightarrow RCH_2-OH + Cl^-$	200
$HO^- + RCH_2-Br \longrightarrow RCH_2-OH + Br^-$	10,000
$HO^- + RCH_2-I \longrightarrow RCH_2-OH + I^-$	30,000

Effect of the solvent

The solvent is needed to dissolve both the substrate and the nucleophile. The substrate is a polar compound like $\overset{\delta+}{C}H_3-\overset{\delta-}{Br}$ and the nucleophile is usually in the form of an ionic solid like Na^+OH^- or $CH_3O^-Na^+$. Nonpolar solvents do not work because polar and ionic substances do not dissolve in nonpolar solvents. Polar solvents can dissolve polar and ionic compounds. Polar solvents fall into two categories:

- **polar protic** solvent that have an acidic proton like water (H_2O) and methanol (CH_3-OH), and
- **polar aprotic** like acetone ($(CH_3)_2\overset{\delta+}{C}=\overset{\delta-}{O}$) and dimethylsulfoxide ($(CH_3)_2\overset{\delta+}{S}=\overset{\delta-}{O}$ }\)) which do not have an acidic proton. Their $\delta+$ end is at the center of the molecule, surrounded by bulky groups.

Polar protic solvent dissolves ionic compounds by ion-dipole interactions forming a layer of solvent around cation and anion. They are not suitable for S_N2 reactionas the solvent layer prevents nucleophiles from approaching electrophilic $\overset{\delta+}{C}$.

Aprotic solvents dissolve the ionic compound by the ion-dipole interaction but leave the anion, i.e., nucleophile, almost free. This is because the anion is prevented from approaching the $\delta+$ pole of the solvent due to steric hindrance, as illustrated in Figure 4.4.1. Recall that electrostatic force is inversely proportional to the square of the distance between the +ve charge and the -ve charge, i.e., the longer the distance, the weaker the force.

Polar aprotic solvents are suitable for S_N2 reaction as the nucleophiles are relatively free to approach electrophilic $\overset{\delta+}{C}$ of the substrate.

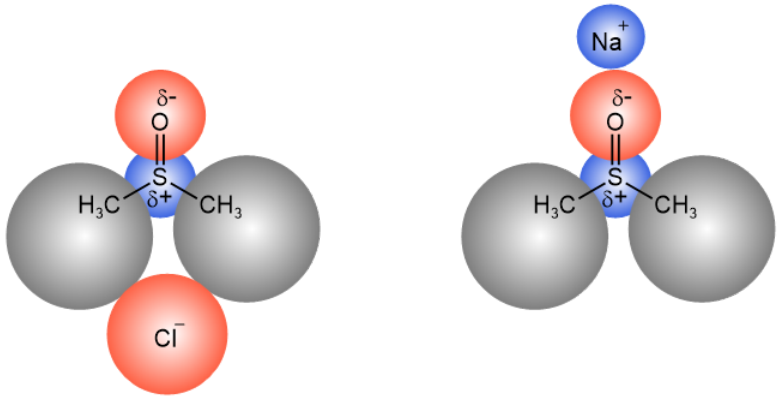


Figure 4.4.1: Illustration of dissolution of Na^+Cl^- in dimethylsulfoxide ($(CH_3)_2\overset{\delta+}{S}=\overset{\delta-}{O}$ }\)). The aprotic solvent dissolves ionic compounds by making strong ion-dipole interaction with the cation but leaves the anion relatively free as it is prevented by steric hindrance from approaching the \(\delta{+} pole of the solvent molecule.

Examples of S_N2 reactions

Alkyl halides are derived from alkanes by free radical reactions. The alkyl halides are then converted to a number of other classes of organic compounds by S_N2 reaction, e.g., ethyl bromide can be converted to:

$$\text{alcohol: } CH_3CH_2Br + OH^- \longrightarrow CH_3CH_2OH + Br^- \ ,$$

$$\text{a thiol: } CH_3CH_2Br + HS^- \longrightarrow CH_3CH_2SH + Br^- \ ,$$

$$\text{an ether: } CH_3CH_2Br + CH_3CH_2O^- \longrightarrow CH_3CH_2OCH_2CH_3 + Br^- \ ,$$

$$\text{a thioether: } CH_3CH_2Br + CH_3CH_2S^- \longrightarrow CH_3CH_2SCH_2CH_3 + Br^- \ ,$$

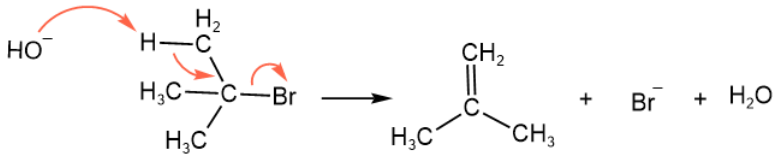
$$\text{an amine: } CH_3CH_2Br + {}^-NH_2 \longrightarrow CH_3CH_2NH_2 + Br^- \ ,$$

$$\text{a nitrile: } CH_3CH_2Br + {}^-C\equiv N \longrightarrow CH_3CH_2C\equiv N + Br^- \ , \text{ and}$$

$$\text{and an alkyne: } CH_3CH_2Br + {}^-C\equiv CCH_3 \longrightarrow CH_3CH_2C\equiv CCH_3 + Br^- \ .$$

Elimination bimolecular (E2)

Recall that the C connected to a functional group is αC, and the one adjacent to it is βC. Also, recall that $\overset{\delta+}{H}-\overset{\delta-}{Br}$ is a strong acid because the acidic polar leaves electron on elecronegative Br. Similarly, protons on βC are acidic because they can send the bonding electrons to Br through the following mechanism.

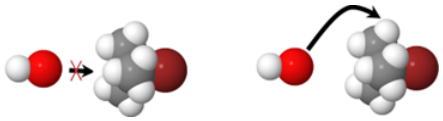


Nucleophiles are bases at the same time. The base HO^- attacks βH, the H−C bonding electrons establish a π-bond between αC and βC, and the αC let go the leaving group Br^-. Since a βH is eliminated along with the leaving group, This reaction mechanism is called β-elimination. It is a bimolecular reaction because of two species involved in this elementary step: the base and the substrate. This specific β-elimination is called **elimination bimolecular (E2)**, where E stands for elimination and 2 for bimolecular.

S_N2 vs. E2

S_N2 and E2 reactions compete because the same reagent is nucleophile in S_N2 and base in the E2 reaction. However, the effect of the structure of the substrate is opposite in S_N2 and E2. The structure of substrate affects S_N2 in this order: $\xleftarrow{\qquad\qquad\qquad\qquad}$ methyl > primary > secondary > tertiary . Tertiary substrates do not react by S_N2 mechanics because the nucleophile is

prevented from approaching αC by steric hindrance. However, βH is still exposed and easily approachable by the base, as shown below.



Further, E2 is facilitated by lowering steric crowding around the αC. Therefore, tertiary substrates yield E2 product under the S_N2 conditions. The steric crowding around αC of secondary substrates is less than tertiary. So, E2 and S_N2 compete, and elimination and substation products are formed from the secondary substrates. There is no steric crowding in the case of primary or methyl substrates. So, S_N2 dominated over E2, and substitution product is formed almost exclusively in the case of primary and methyl substrates. The effect of the structure of the substrate on E2 reaction is in this order: $\xrightarrow{\text{methyl} < \text{primary} < \text{secondary} < \text{tertiary}}$, which is the opposite of S_N2. E2 reactions synthesize alkenes from alkyl halides, as in the above example, and form similar functional groups that will be described later.

Substitution nucleophilic unimolecular S_N1

Ionic compounds like Na^+Cl^- dissociate into ions in polar protic solvents like H_2O better than in polar aprotic solvents like. This is because unlike polar aprotic solvents solvating only cations, polar protic solvents solvent both cations and anions. Polar covalent bonded compounds like $\overset{\delta+}{H}-\overset{\delta-}{Br}$ also dissociate in polar portic solvent by the same mechanism. Polar organic compounds like $H_3\overset{\delta+}{C}-\overset{\delta-}{Br}$ have less tendency to dissociate. However, if the substrate is tertiary like $(H_3C)_3\overset{\delta+}{C}-\overset{\delta-}{Br}$, the effect of steric crowding on αC and the dissociation ability of the polar protic solvent together are strong enough to dissociate polar organic compounds, as in 1st step of the mechanism shown in Figure 4.4.2.

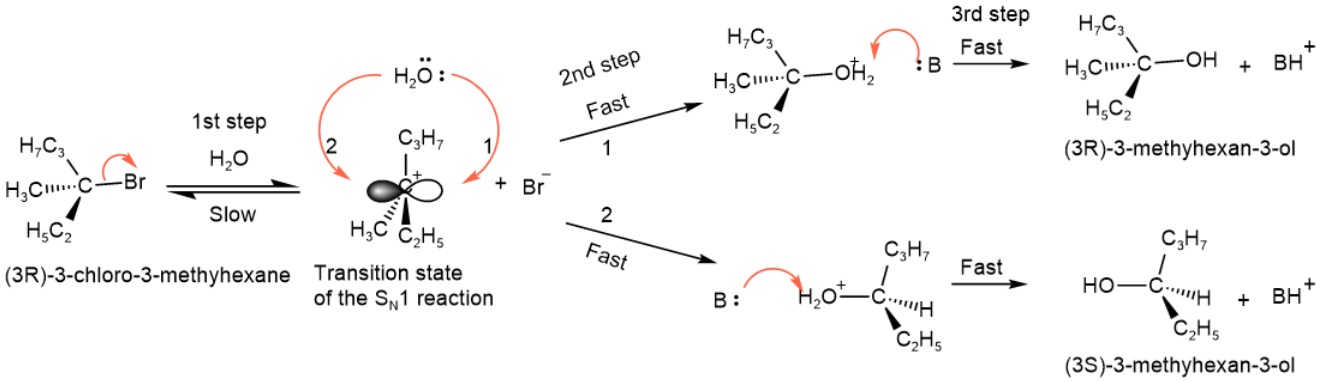


Figure 4.4.2: S_N1 reaction mechanism illustrated. (Copyright; Public domain)

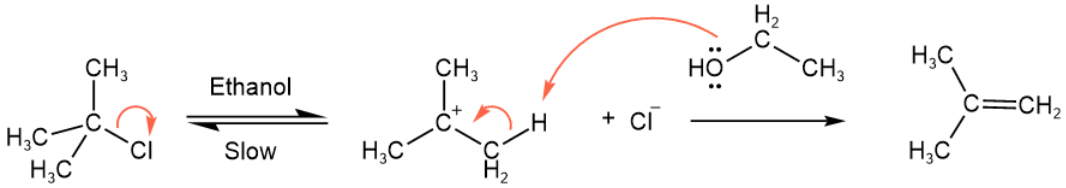
The carbonation produced in the 1st step, e.g., $(H_3C)_3C^+$ in this case, is a strong electrophile and reacts with any nucleophile around, including solvent molecules, e.g., (H_2O in this case. Since the 2nd step in non-selective, the solvent wins over any other nucleophile in the system because of its higher concentration. The neutral nucleophile becomes a cationic group in the product of the 2nd step, e.g., $R-OH_2^+$ in this case. The strong acid like $R-OH_2^+$ donates its proton to any base ($:B$) in the medium, including the solvent, e.g., H_2O in this case.

The overall reaction is substitution nucleophilic. The overall reaction is unimolecular proportional to the substrate concentration in the 1st step. So, this mechanism is called substitution nucleophilic unimolecular S_N1, where S stands for substitution, N or nucleophilic, and 1 for unimolecular.

- S_N1 is unimolecular.
- The carbonation intermediate is trigonal planar and the nucleophile can attack it from either side to make a new covalent bond. Therefore, if S_N1 reaction happens on a chiral carbon, around 50% of the product retains the configuration of the reactant, and the other 50% has the inverted configuration, i.e., **the product is a racemic mixture**.
- S_N1 Reaction takes place on tertiary alkyl halides. Secondary and primary alkyl halides do not undergo S_N1 reactions.

Elimination unimolecular (E1)

The carbocation formed in S_N1 reaction has acidic protons on βC because the electrons left on the conjugate base are stabilized as a π-bond, as in the mechanism shown below.



This mechanism eliminates a leaving group and a proton β to the leaving group. The 1st step, which is the rate-determining step, is unimolecular. This mechanism is called elimination unimolecular (E1), where E stands for elimination and 1 for unimolecular. E1 always competes with S_N1 because they have a common first step.

What decides the reaction will happen by S_N2, E2, S_N1, or E1 mechanism?

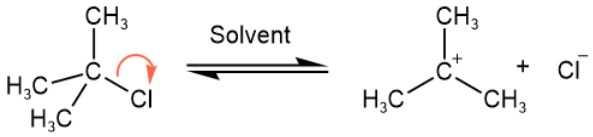
All of the S_N2, E2, S_N1, or E1 involve two nucleophiles, an incoming nucleophile and a leaving group which is a nucleophile attached to the substrate. Incoming nucleophile substitutes the leaving group in substitution reactions and eliminates βH and the leaving group in elimination reactions. The question is what decides which mechanism the reaction will follow? This section addresses the answers to this question. The solvent and the nucleophile determine S_N2 or S_N1 condition.

> Strong nucleophiles, usually anionic like HO^-, RO^-, and polar aprotic solvent like acetone ($(CH_3)_2C{=}O$) or dimethylsulfoxide ($(CH_3)_2S{=}O$) define **S_N2 and E2 condition**.

The substrate dictates S_N2, E2, or both will happen: S_N2 happens on the primary or methyl substrate, E2 happens on the tertiary substrate, and both happen simultaneously on a secondary substrate. S_N2 and E2 reactions do not take place in polar protic solvents.

> Polar protic solvents like water (H_2O) or methanol ($CH_3{-}OH$) and neutral nucleophiles like H_2O, ROH define **S_N1 and E1 condition**.

Anionic nucleophiles can not exist in protic sovlents due to acid-base neutralization reactions between them. Neutral species are poor nucleophiles but are OK in S_N1 or E1 reactions because they are not involved in the rate-determining step. The following data on the dissociation rate of tertiary butyl chloride shows the effect of polar protic solvent on the rate-determining step of S_N1 and E1 reactions.



Solvent	Polarity (Dielectric constant)	Protic or aprotic	Relative rate
Water (H_2O)	78	Polar protic	40
Ethanol ($CH_3CH_2{-}OH$)	24	Polar protic	1
Acetone ($(CH_3)_2C{=}O$	21	Polar aprotic	0.005

S_N1 and E1 usually compete as both have the same rate-determining step. Tertiary substrate easily react by S_N1 and E1 mechanism. Secondary substrate may or may not take place by S_N1 and E1. For example, secondary alkyl halides do not react, but secondary alcohols and ethers react by S_N1 and E1. Primary substrates usually do not react by S_N1 and E1 mechanisms.

Effect of leaving group on S_N2, E2, S_N1, and E1 mechanisms

Recall that the stronger the acid, the weaker the conjugate base, and voice versa. For example, HI is a strong acid and I^- is a weak base, while HF is a weak acid and F^- is a strong base. In other words, strong bases do not tend to protons, and weak bases easily leave protons in acid-base reactions. The same applies to leaving groups in S_N2, E2, S_N1, and E1 reactions.

> Strong bases are poor leaving groups, and weak bases are good leaving groups in S_N2, E2, S_N1, and E1 reactions.

The basicity and hence the leaving propensity of leaving groups follows these trends:

1. Basicity decreases from top to bottom in a group of periodic table. For example, the base strength of halides follows this order: $\xleftarrow{\qquad\qquad\qquad\qquad}$ $F^- > Cl^- > Br^- > I^-$. Leaving propensity follows the opposite trend i.e., $-I$ is the best-leaving group, while $-F$ is the worst leaving group. $-F$ does not act as a leaving group.
2. Basicity decreases from right to left in a row of the periodic table. For example, the base strength of 2nd row elements follows this order: $\xleftarrow{\qquad\qquad\qquad\qquad\qquad\qquad}$ $F^- < OH^- < NH_2^- < CH_3^-$. Leaving propensity follows the opposite trend, i.e., $-OH$ $-NH_2$. $-CH_3$ are the worst leaving group than $-F$ and do not act as leaving group.
3. Basicity decreases upon protonation in the cases of amphoteric species. For example, H_2O is a weaker base than OH^-. That is why, $-OH$ is not a leaving group, but $-\overset{+}{O}H_2$ is a good leaving group that leaves as a weak base H_2O.

Nucleophilic substitution and elimination reactions of alcohols, ethers, amines, and sulfur compounds

Reactions of alcohols

Substitution reactions

The O of an alcohol ($R-OH$) is protonated by a strong acid to convert it from not leaving group (($-OH$) to a good leaving group (($-\overset{+}{O}H_2$). The protonated alcohols under go S_N1 or E1 reactions, except when the substrate is primary. that undergoes S_N2 or E2 reactions. For example, primary, secondary, and tertiary alcohols undergo nucleophilic substitution with HCl, HBr, or HI as shown below.

$$(CH_3)_3C-OH + HBr \rightleftharpoons (CH_3)_3C-\overset{+}{O}H_2 + Br^- \xrightarrow{\Delta} (CH_3)_3C-Br + H_2O \quad \text{by } S_N1 \text{ mechanism}$$

$$(CH_3)_2CH-OH + HBr \rightleftharpoons (CH_3)_2CH-\overset{+}{O}H_2 + Br^- \xrightarrow{\Delta} (CH_3)_2CH-Br + H_2O \quad \text{by } S_N1 \text{ mechanism}$$

$$CH_3CH_2-OH + HBr \rightleftharpoons CH_3CH_2-\overset{+}{O}H_2 + Br^- \xrightarrow{\Delta} CH_3CH_2-Br + H_2O \quad \text{by } S_N2 \text{ mechanism}$$

Elimination reactions may compete with substitution reactions described above, but the alkene products of elimination add HCl, HBr, or HI and end in the same final product as the substitution products. These reactions of alkens will be described in a later section.

Elimination reactions

Inorder to perform elimination reactions, the O of alcohols ($R-OH$) is protonated by sulfuric acid (H_2SO_4) in water solution. The conjugate base of sulfuric, i.e., $[HSO_4]^-$ is a weak nucleophile and does not cause the substitution reaction. Water is a nucleophile, but its creation does not change the intermediate. However, water acts as a base picking up βH, causing elimination reactions, as shown below.

$$(CH_3)_3C-OH \overset{H2SO4+H2O}{\rightleftharpoons} (CH_3)_3C-\overset{+}{O}H_2 \xrightarrow{\Delta+HSO4^-+H2O} (CH_3)_2C=CH_2 + H_2O \quad \text{by E1 mechanism}$$

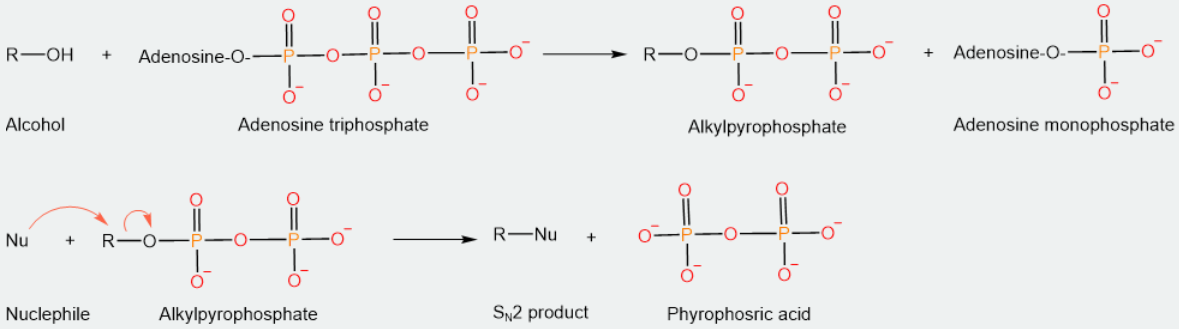
$$(CH_3)_2CH-OH \overset{H2SO4+H2O}{\rightleftharpoons} (CH_3)_2CH-\overset{+}{O}H_2 \xrightarrow{\Delta+HSO4^-+H2O} CH_3CH=CH_2 + H_2O \quad \text{by E1 mechanism}$$

$$CH_3CH_2-OH \overset{H2SO4+H2O}{\rightleftharpoons} CH_3CH_2-\overset{+}{O}H_2 \xrightarrow{\Delta+HSO4^-+H2O} CH_2=CH_2 + H_2O \quad \text{by E2 mechanism}$$

These reactions are called dehydration of alcohols that convert alcohols to alkenes.

Activating alcohols in biochemical systems

Alcohols are common intermediates in biochemical reactions. However, unlike in a chemical laboratory, strong acids like \(\ce{HBr}) or \(\ce(H2SO4) needed to activate alcohol groups do not survive under physiological conditions. Phosphoric acid and its anhydrides, i.e., pyrophosphoric acid and triphosphoric acids are weak acids and their conjugate bases, i.e., phosphate, pyrophosphate, and triphosphate, are weak bases and good leaving groups. Therefore, biological systems convert alcohols into phosphate or pyrophaste esters by reacting with adenosine triphosphate (ATP). It allows them to participate in the nucleophilic substitution reactions under physiological conditions as shown below.



Reactions of ethers

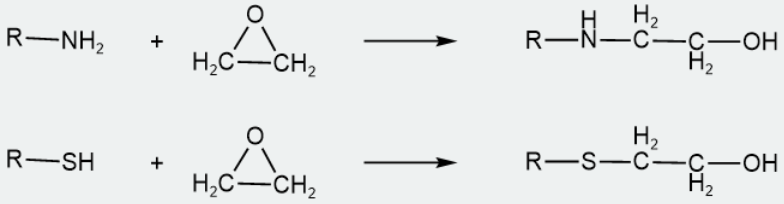
Like alcohols, the O of ethers ($\mathrm{R-O-R'}$) is protonated by a strong acid to convert it from not leaving group (($-\mathrm{OR'}$) to a good leaving group (($-\overset{+}{\mathrm{O}}\mathrm{HR}$). For example, ethers undergo nucleophilic substitution reactions with HCl, HBr, or HI producing alcohols and an alkyl halides, as shown below.

$$(\mathrm{CH_3})_3\mathrm{C-O-C(CH_3)_3} + \mathrm{HBr} \rightleftharpoons (\mathrm{CH_3})_3\mathrm{C-}\overset{+}{\mathrm{O}}\mathrm{HC(CH_3)_3} + \mathrm{Br^-} \xrightarrow{\Delta} (\mathrm{CH_3})_3\mathrm{C-Br} + (\mathrm{CH_3})_3\mathrm{C-OH} \quad \text{by } S_N1 \text{ mechanism}$$

$$\mathrm{CH_3CH_2-O-CH_3} + \mathrm{HBr} \rightleftharpoons \mathrm{CH_3CH_2-}\overset{+}{\mathrm{O}}\mathrm{HCH_3} + \mathrm{Br^-} \xrightarrow{\Delta} \mathrm{CH_3CH_2-OH} + \mathrm{CH_3-Br} \quad \text{by } S_N2 \text{ mechanism}$$

Ethylene oxide: A sterilant based on S_N2 reactions

Ethylene oxide is a three-membered cyclic ether. It is a colorless gas with a boiling point of 11 ^{o}C. The $\mathrm{C-O}$ bonds in ethylene oxide are unstable due to angle strain. Therefore, the ether groups in ethylene oxide act as excellent leaving groups because angle strain is released. Ethylene oxide reacts fast by S_N2 mechanism with amino ($-\mathrm{NH_2}$) and sulfhydryl ($-\mathrm{SH}$) groups that are commonly present in biochemicals, as shown below.



These reactions modify the biochemicals, leading to the death of microorganisms. Ethylene oxide is used as a fumigant in foods and textiles and to sterilize surgical instruments in hospitals.

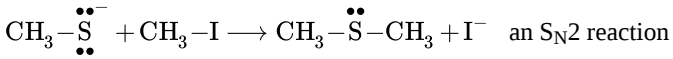
Reactions of amines

Amine $-\mathrm{NH_2}$ and also its protonated form $-\mathrm{NH_3^+}$ are poor as leaving groups and do not act as leaving groups. However, amines are good nucleophiles and act as incoming nucleophiles in various reactions, e.g., in S_N2 reactions with alkyl halides, as shown below.

$$\mathrm{CH_3CH_2-Br} + \mathrm{CH_3-NH_2} \longrightarrow \mathrm{CH_3CH_2-}\overset{+}{\mathrm{N}}\mathrm{H_2-CH_2CH_3} + \mathrm{Br^-} \quad \text{an } S_N2 \text{ reaction}$$

Reactions of thiols

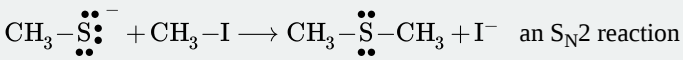
Thiolate ion is a good nucleophile for S_N2 reactions as shown in the following example.



$$CH_3-\ddot{S}^{-} + CH_3-I \longrightarrow CH_3-\ddot{S}-CH_3 + I^- \quad \text{an } S_N2 \text{ reaction}$$

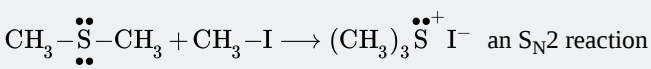
Methylating agent in lab and in biochemical systems

Methyl ($\overset{\delta+}{C}H_3-\overset{\delta-}{X}$ is the best electrophilic site, and iodide ($-I$) is the best-leaving group that makes methyliodide ($\overset{\delta+}{C}H_3-\overset{\delta-}{I}$) the best substrate for methylating any nucleophile in an S_N2 reactions, as shown in the following examples.



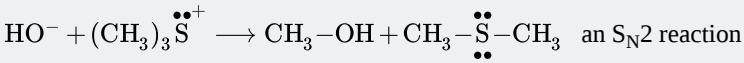
$$CH_3-\ddot{S}^{-} + CH_3-I \longrightarrow CH_3-\ddot{S}-CH_3 + I^- \quad \text{an } S_N2 \text{ reaction}$$

Sulfur is a larger atom that can accommodate three alkyl groups. Therefore, dimethyl sulfide produced in the above reaction can be methylated one more time, producing a solfonium salt, as shown below.



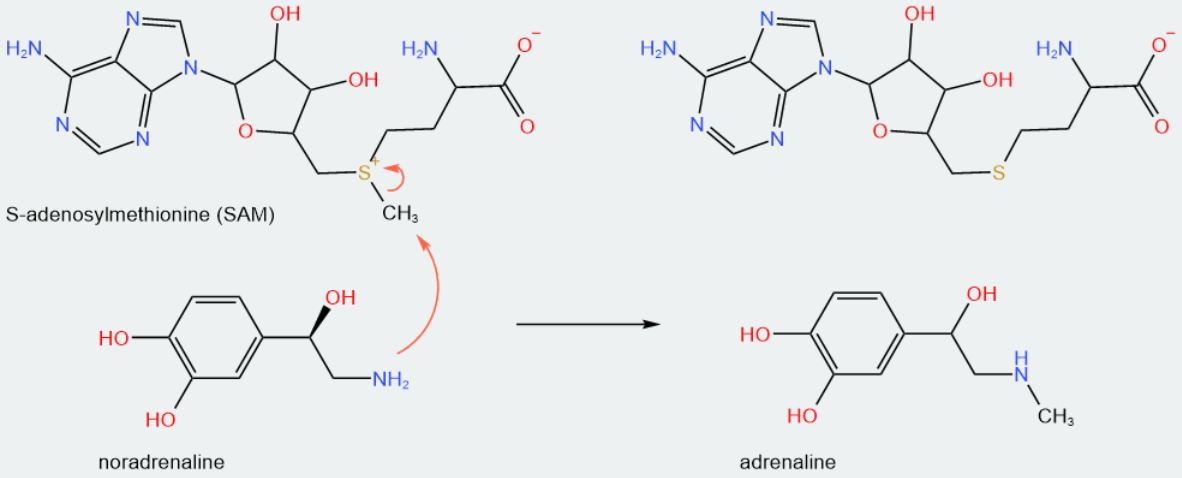
$$CH_3-\ddot{S}-CH_3 + CH_3-I \longrightarrow (CH_3)_3\ddot{S}^{+}\ I^- \quad \text{an } S_N2 \text{ reaction}$$

Dimethyl sulfide ($(CH_3)_2\ddot{S}^{+}-$) is an excellent leaving group that makes trimethylsulfonium ion ($(CH_3)_3\ddot{S}^{+}$) an excellent methylating agent, as shown in an example below.



$$HO^- + (CH_3)_3\ddot{S}^{+} \longrightarrow CH_3-OH + CH_3-\ddot{S}-CH_3 \quad \text{an } S_N2 \text{ reaction}$$

methyliodide (H_3C-I) is a common methylating agent in laboratory,- but it is not available in biological systems. S-adenosylmethionine (SAM), which is similar to trimethylsulfonium ion ($(CH_3)_3\ddot{S}^{+}$) is a common methylating agent in biological systems, as shown in an example below.



In the above example of a biochemical reaction, noradrenaline hormone is converted into a more potent adrenaline hormone by methylation reaction using SAM as a methylating agent.

4.5: Nucleophilic acyl substitution reactions

Learning Objectives

- Understand nucleophilic acyl substitution mechanisms, and factors that affect them.
- Apply nucleophilic acyl substitution to reactions of acid halides, anhydrides, carboxylic acids, esters, and amides.
- Wright the reactions with reagents to convert carboxylic acids into acid halides and phosphate esters in living things.
- Apply nucleophilic acyl substitution reactions to synthesize condensation polymers.

What is a nucleophilic acyl substitution reaction?

A carbon double bonded with oxygen, i.e., $-\overset{\overset{\mathrm{O}}{\|}}{\mathrm{C}}-$ is a **carbonyl group**. The C of the carbonyl group is bonded to two other groups. If one of the group bonded to the carbonyl C is an alkyl ($\mathrm{R}-$), or hydrogen ($\mathrm{H}-$), it become an **acyl group**, i.e., $\mathrm{R}-\overset{\overset{\mathrm{O}}{\|}}{\mathrm{C}}-$. The acyl group has a polar double bond, i.e., $\mathrm{R}-\overset{\delta-}{\overset{\overset{\mathrm{O}}{\|}}{\underset{\delta+}{\mathrm{C}}}}-$ where the carbonyl C is partial positive ($\delta+$), i.e., it is an electrophile. If the acyl group is attached to a nucleophile that can act as a leaving group ($-\mathrm{Lv}$), a stronger nucleophile (Nu^-) can substitute it from the acyl group in reactions called **nucleophilic acyl substitution reactions**, as shown below in a generalized reaction.

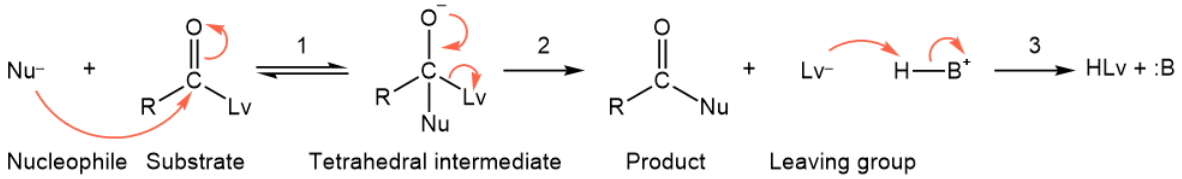
$$\mathrm{Nu}^- + \mathrm{R}-\overset{\overset{\mathrm{O}}{\|}}{\mathrm{C}}-\mathrm{Lv} \longrightarrow \mathrm{R}-\overset{\overset{\mathrm{O}}{\|}}{\mathrm{C}}-\mathrm{Nu} + \mathrm{Lv}^-$$

Mechanisms of nucleophilic acyl substitution reactions

In nucleophilic acyl substitution reactions, there are two reactants, the nucleophile and the acyl group containing substrate. Nucleophiles can exist in anionic base form Nu^- in a basic medium and neutral acid form HNu in a neutral or acidic medium. The anionic form is a better nucleophile than its neutral acid form. Similarly, the substrate can exist as neutral $\mathrm{R}-\overset{\overset{\mathrm{O}}{\|}}{\mathrm{C}}-\mathrm{Lv}$ in neutral or basic medium or as protonated $\mathrm{R}-\overset{\overset{\overset{+}{\mathrm{O}}\mathrm{H}}{\|}}{\underset{\delta+}{\mathrm{C}}}-\underset{\delta-}{\mathrm{Lv}}$ form in an acidic medium. The protonated form has more $\delta+$ charge and is a better electrophile. The anionic nucleophile cannot coexist with protonated substrate, because of their acid-base reaction. The other three combinations, i.e., basic nucleophile + neutral substrate, neutral nucleophile + protonated substrate, and neutral nucleophile + neutral substrate, can coexist as described below.

Base-promoted mechanism

The nucleophile in its more reactive basic form Nu^- and neutral substrate $\mathrm{R}-\overset{\delta-}{\overset{\overset{\mathrm{O}}{\|}}{\underset{\delta+}{\mathrm{C}}}}-$ can coexist in a basic medium. The nucleophile Nu^- attacks the electrophilic carbonyl $\overset{\delta+}{\mathrm{C}}$ and, simultaneously, the $\overset{\delta+}{\mathrm{C}}$ breaks the weakest π-bond, as shown in step#1 of the mechanism below. The electrophilic C changes hybridization from sp^2 to sp^3, i.e., a **tetrahedral intermediate**.

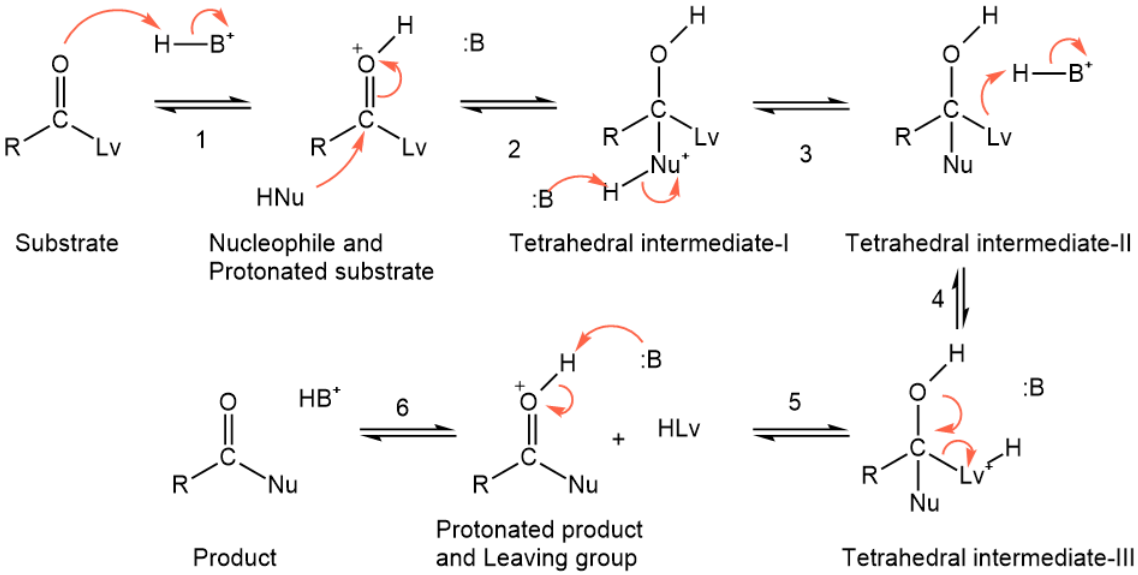


The nucleophilic O^- created in the first step attacks the $\overset{\delta+}{C}$ to re-establish the π-bond, and either $\overset{\delta+}{C}-\overset{\delta-}{Nu}$ or $\overset{\delta+}{C}-\overset{\delta-}{Lv}$-bond breaks. This step is also called the **collapse of the tetrahedral intermediate**. Breakage of $\overset{\delta+}{C}-\overset{\delta-}{Nu}$-bond reverses the first step and breakage of $\overset{\delta+}{C}-\overset{\delta-}{Lv}$-bond leads to the products of step#2. Later will likely happen when the $-Lv$ is a good leaving group. The second step is usually not reversible because Lv^- is usually a poor nucleophile. The leaving group ultimately picks up a proton from any acid molecule (HB^+) in the medium, as shown in step#3.

The rate of reaction is increased by converting the neutral or acid form of the nucleophile (HNu) to its more reactive conjugate base form (Nu^-) in the basic medium in this mechanism. Therefore, it is called **base promoted nucleophilic acyl substitution mechanism**.

Acid-catalyzed mechanism

In this mechanism, the nucleophile is in less reactive neutral HNu form, but the substrate is in a more reactive protonated $R-\overset{\overset{+}{O}H}{\overset{\delta+}{C}}-\overset{\delta-}{Lv}$ form, in an acidic medium. Acid in the medium protonate the carbonyl O in step#1, as shown in the mechanics below.



Nucleophile HNu attacks the protonated electrophilic carbonyl $\overset{\delta+}{C}$ and, simultaneously, the π-bond breaks in step#2, leading to a tetrahedral intermediate-I. The neutral nucleophile becomes +ve charged after donating its lone pair of electrons. It carries an acid proton. Although the medium is acidic in this case, it is usually amphoteric, with some basic groups still present. The incoming nucleophile donates its acidic proton to any basic molecule (: B in the medium in step#3, leading to a tetrahedral intermediate-II. Then the acidic medium protonates either $-Nu$ or $-Lv$-groups. Protonation of $-Nu$ reverses the reaction but protonation of $-Lv$ makes it a better-leaving group in tetrahedral intermediate-III of step#4. The lone pair on O re-establishes the π-bond with the electrophilic C, and at the same time, the leaving group leaves, leading to a protonated acyl product in the step#5. The protonated acyl donates its proton to any base : B in the medium and becomes the product in step#6.

Acid is a catalyst in this mechanism as it is consumed in step#1 but re-generated in step#6, and accelerates two slow steps in this mechanism, i.e., step#2 and step#5. That is why it is called the **acid-catalyzed nucleophilic substitution mechanism**. Specifically, the protonation of the substrates $R-\overset{\overset{\delta-}{O}}{\overset{\delta+}{C}}-$ converts it to a more reactive $R-\overset{\overset{+}{O}H}{\overset{\delta+}{C}}-\overset{\delta-}{Lv}$ form that accelerates step#2, and protonation of $-Lv$ in tetrahedra intermediate-II converts it to a good leaving $-Lv^+H$ group in tetrahedral intermediate-III that accelerates step#5. The other steps in the mechanism are acid-base reactions which are inherently fast.

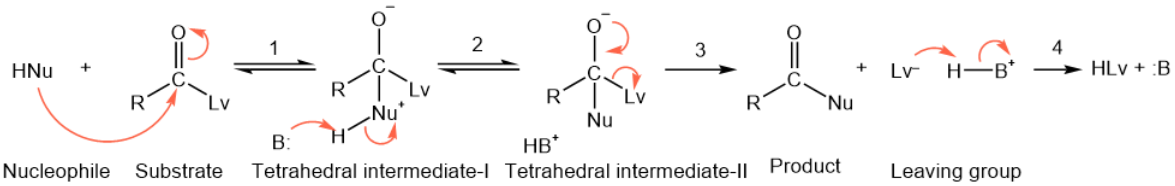
All the steps in this mechanism are reversible. It means there is an equilibrium between reactants and products. Three situations can arise, which are the following.

1. If the incoming nucleophile (HNu) is a poor nucleophile compared to the leaving group (HLv), the reactants dominate,
2. If the incoming nucleophile (HNu) is a good nucleophile compared to the leaving group (HLv), the products dominate, and
3. If the nucleophilicity of the incoming nucleophile (HNu) is comparable to that of the leaving group (HLv), about an equal-equal mixture of reactants and products exists.

In situation#3, the equilibrium can be manipulated to favor products or reactants based on Le Chatelier's principle. Employing one of the reactants in excess or removing one of the products drives the reaction forward. Similarly, adding one of the products in excess or removing one of the reactants drives the reaction in the reverse direction.

Neutral nucleophile and neutral substrate reaction mechanism

In this mechanism, both the nucleophile HNu and the substrate $\mathrm{R-\overset{\delta+}{\underset{\delta-}{\underset{\|}{\overset{\|}{C}}}}O-}$ are in their less reactive neutral forms in a neutral medium. Nucleophile HNu attacks the electrophilic carbonyl $\overset{\delta+}{C}$ and, simultaneously, the π-bond breaks in step#1, leading to tetrahedral intermediate-I. The acidic proton on the incoming nucleophile in intermediate-I is removed by any base molecule (: B in the medium in step#2, leading to tetrahedral intermediate-II, as shown below.



The nucleophilic O^- in tetrahedral intermediate-II re-establishes the π-bond, and, simultaneously, either Nu^- or Lv^- leaves. Departure of Nu^- reverses step#2 and departure of Lv^- leads to the product in step#3. Lv^- is neutralized by any acid molecule present in the medium in step#4.

Since both the nucleophile and the substrate are in their less reactive neutral forms, it works only in situations where a good leaving group ($-Lv$) is attached to the acyl substrate. Step#3 is usually irreversible because the $-Lv$ groups chosen for these reactions are usually good leaving groups and poor nucleophiles.

Effect of leaving group on nucleophilic acyl substitution reactions

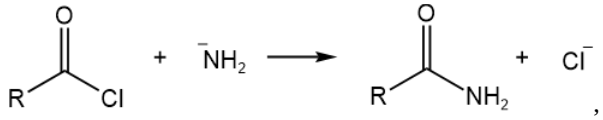
Good leaving groups (Lv^-) are usually weak bases and poor nucleophiles at the same time. Weak bases do not fully share their electrons with acidic protons, and poor nucleophiles do not fully share their electrons with electrophilic C′s. Poor nucleophiles increase the rates of slow steps in the following ways:

1. their bond with C is strongly polar, making the acyl $\overset{\delta+}{C}$ a more reactive electrophile, and
2. they leave easily increasing the rate of collapse of tetrahedral intermediates.

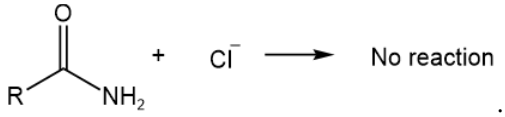
Leaving groups commonly encountered in nucleophilic acyl substitution reactions are halogens like chlorine ($-Cl$) or bromine ($-Br$), carboxylate ($\mathrm{-O-\overset{\|}{\underset{\|}{\overset{O}{C}}}-R}$), hydroxyl ($-OH$), alkoxy ($-OR$), and amine ($-NH_2$) groups. Their basicity and nucleophilicity increases in this order: $\mathrm{Cl^- < R-COO^- < HO^- \approx RO^- < {}^-NH_2}$. Therefore, their ability to leave increases in the opposite order $\mathrm{{}^-NH_2 < HO^- \approx RO^- < R-COO^- < Cl^-}$, i.e., halogens like chlorine ($-Cl$) are the best-leaving groups and amines ($-NH_2$) are the worst-leaving group.

These leaving groups are found in the following subclasses of acyl substrates: amides ($\mathrm{R-\overset{\|}{\underset{\|}{\overset{O}{C}}}-NH_2}$), carboxylic acids ($\mathrm{R-\overset{\|}{\underset{\|}{\overset{O}{C}}}-OH}$), easters ($\mathrm{R-\overset{\|}{\underset{\|}{\overset{O}{C}}}-OR}$), acid anhydrides ($\mathrm{R-\overset{\|}{\underset{\|}{\overset{O}{C}}}-O-\overset{\|}{\underset{\|}{\overset{O}{C}}}-R}$), and acid halides ($\mathrm{R-\overset{\|}{\underset{\|}{\overset{O}{C}}}-Cl}$). These substrates are

carboxylic acids and their derivatives. Their reactivity as acyl substrates in nucleophilic acyl substitution reactions increases in this order: $\mathrm{R{-}\overset{\overset{\mathrm{O}}{\|}}{\mathrm{C}}{-}NH_2} < \mathrm{R{-}\overset{\overset{\mathrm{O}}{\|}}{\mathrm{C}}{-}OH} \approx \mathrm{R{-}\overset{\overset{\mathrm{O}}{\|}}{\mathrm{C}}{-}OR} < \mathrm{R{-}\overset{\overset{\mathrm{O}}{\|}}{\mathrm{C}}{-}O{-}\overset{\overset{\mathrm{O}}{\|}}{\mathrm{C}}{-}R} < \mathrm{R{-}\overset{\overset{\mathrm{O}}{\|}}{\mathrm{C}}{-}Cl}$. It means a group that is higher in the reactivity order can be easily converted to a group that is lower in the order, but the reverse does not happen normally. For example, the following reaction is feasible:



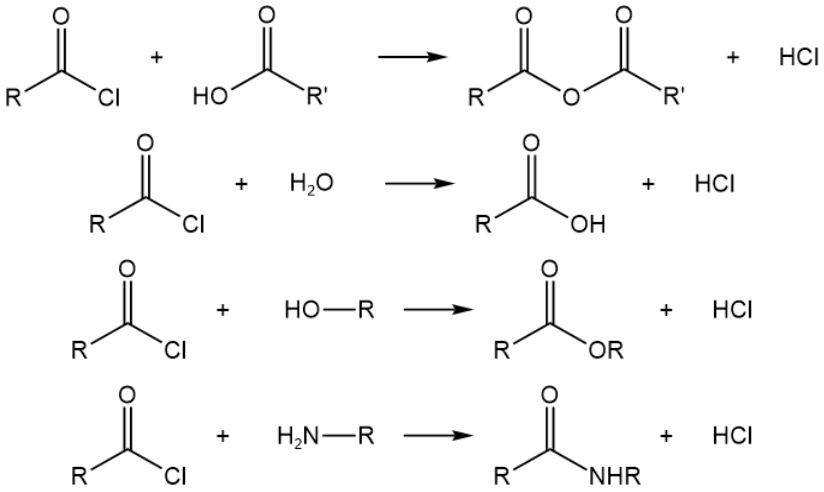
but the following reaction, i.e., the reverse of it, does not happen:



Examples of nucleophilic acyl substitution reactions

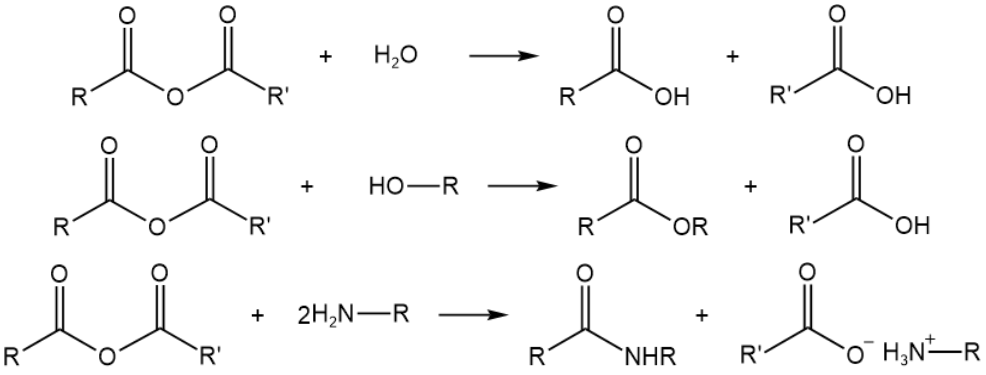
Reactions of acid halides

The acid halides, such as acid chlorides or bromides are the most reactive class of acyl substrates that can react with neutral nucleophiles and neutral substrates in a neutral medium. For example, acid chlorides can react with carboxylic acid, water, alcohol, or amines to form acid anhydrides, carboxylic acids, esters, and amines, respectively, as shown below.

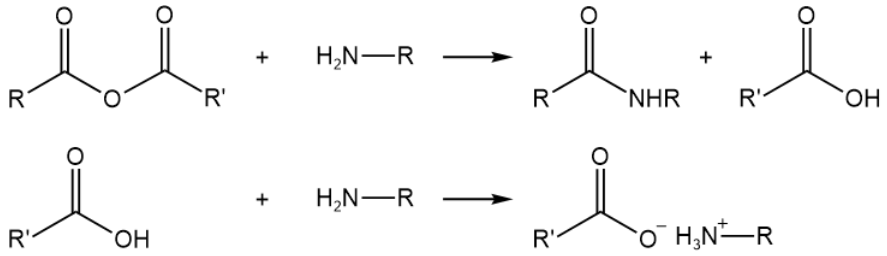


Reactions of acid anhydrides

The acid anhydrides are reactive acyl substrates, second only to acid halides. Anhydride can react with water to produce two equivalents of carboxylic acid; with alcohols to produce one equivalent of ester and one equivalent of carboxylic acid; and with two amines to produce one equivalent of amide and one equivalent of an amine salt of carboxylic acid, as shown below.

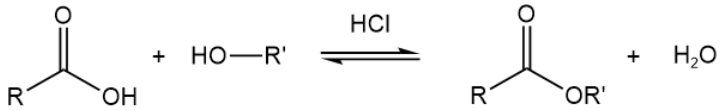


Two moles of amine are needed in the last reaction because initially formed carboxylic acid reacts with the amine base producing an amine salt, as shown below.



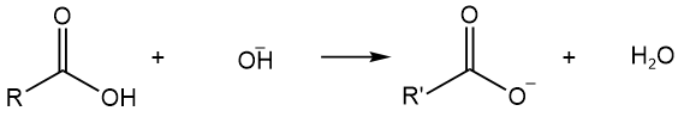
Reactions of carboxylic acids

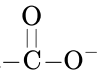
Carboxylic acids react with alcohols to produce ester but slowly. These reactions are reversible because the reactivity of the incoming nucleophiles is about the same as that of leaving nucleophoiles. Acid catalysis is applied to accelerate these reactions, as shown in the example below.

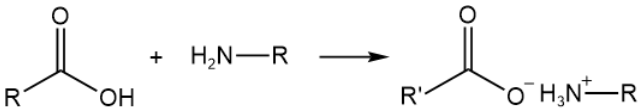


Since there is an equilibrium, excess alcohol is used, or water is removed to drive the reaction forward. These relations are called **alcoholysis reactions** because the reactant alcohol is also solvent. If a reverse reaction is desired, water is employed in excess as a solvent and the reaction is called **hydrolysis** of esters.

Carboxylic acids do not react with halides and another carboxylic acid because the incoming nucleophiles are poor relative to the leaving group. Base can not be added to accelerate the reactions of carboxylic acids because the base neutralizes the substrate, as shown below.



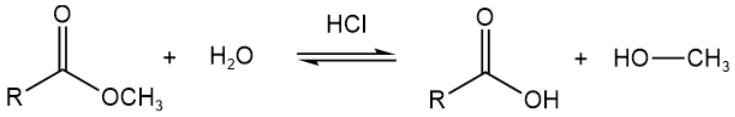
The carboxylate anion ( is the least reactive acyl substrate that is below amides in the reactivity order. Carboxylic acids do not react with amines for the same reason, i.e., amine base neutralizes the carboxylic acid producing ammonium salts, as shown below.



Protonated amines in the ammonium salts are not nucleophiles as they do not have lone pair of electrons on the N.

Reactions of esters

Esters are comparable in reactivity with carboxylic acids but without the acidic protons. Therefore, both acid-catalysis and base-promoted reactions can be carried out with esters. For example, water hydrolyzes esters under acidic conditions, as shown below.



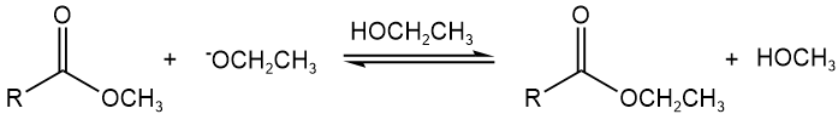
It is a reversible reaction that can be driven forward by employing excess water or by removing the alcohol product. These reactions are called acid-catalyzed **hydrolysis** of esters.

The alkoxy group of esters can be substituted by an alkoxy group of another alcohol, as shown below.

It is a reversible reaction that can be driven forward by employing excess reactant alcohol or by removing the product alcohol. These reactions are called acid-catalyzed **transesterification** reactions.

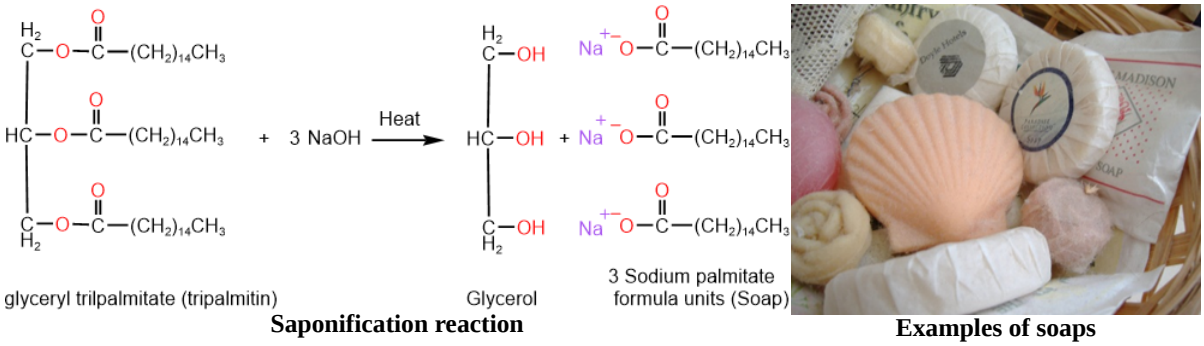
Base-promoted reactions of esters

Transesterification of esters can be performed by employing a conjugated base of the alcohos, as shown in the following example.



Similarly, the example below shows that esters can be hydrolyzed using alkali, i.e., hydroxide ions.

The base-promoted hydrolysis of esters is called **saponification** which is used to hydrolyze fats, as shown in Figure 4.5.1. Fats are tri-esters of glycerol and fatty acids. Saponification of fasts produces sodium salts of fatty acids and glycerol. Sodium salts of fatty acids are soaps.

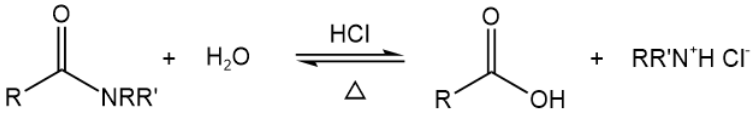


Saponification reaction **Examples of soaps**

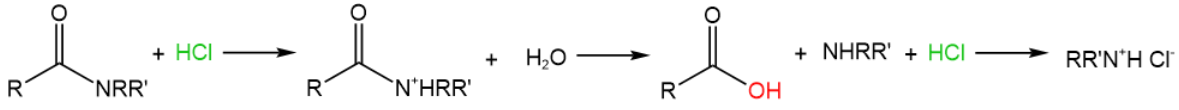
Figure 4.5.1: An example of saponification of fat (tripalmitin) converted into glycerol and soap (sodium palmitate). (Copyright; Public domian)

Reactions of amides

Amides are the least reactive among the acyl substrates. Amides do not react with halides, water, or alcohols under neutral conditions because the incoming nucleophiles are poor than the leaving group. However, water hydrolyzes amides under acid catalysis and heating, as shown below.

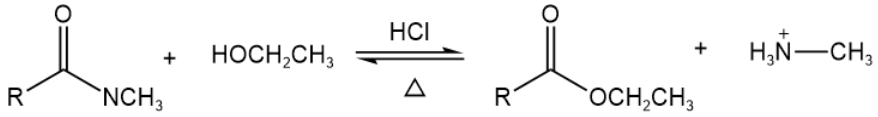


Acid converts poor leaving $-\mathrm{NRR'}$ group to a good leaving $-\overset{+}{\mathrm{N}}\mathrm{HRR'}$ group, and also removes the product amine $\mathrm{NRR'}$ by acid-base reaction, as illustrated below.



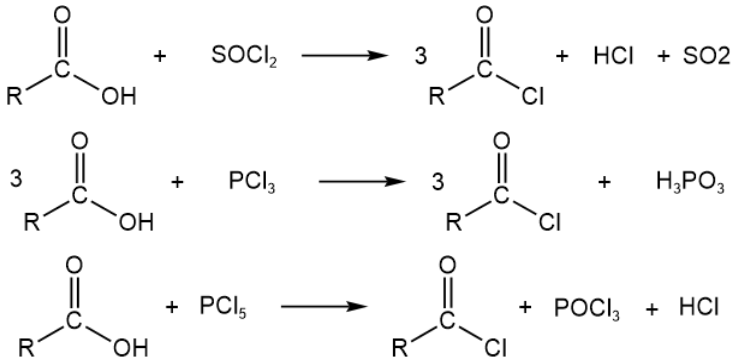
The ammonium ion $(\mathrm{RR'N^+HCl^-}$ does not have a lone pair on N and is not a nucleophile, that makes the reaction irreversible.

As shown below, alcohols react with amides under acid catalysis and heating.

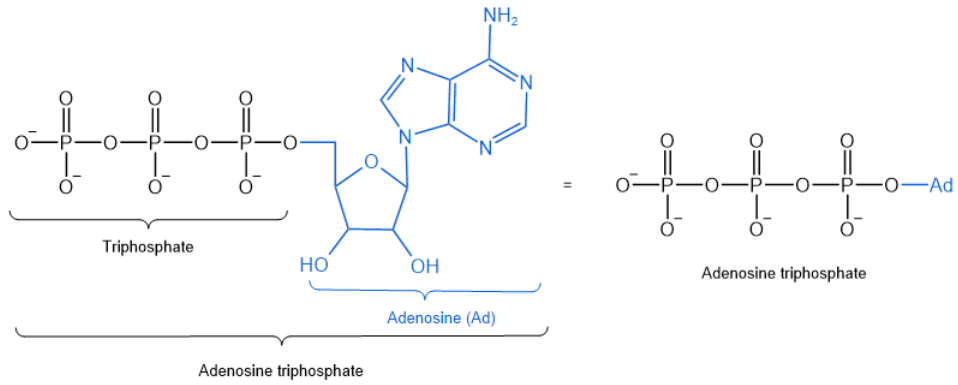


Activating carboxylic acids in the laboratory and in biochemical systems

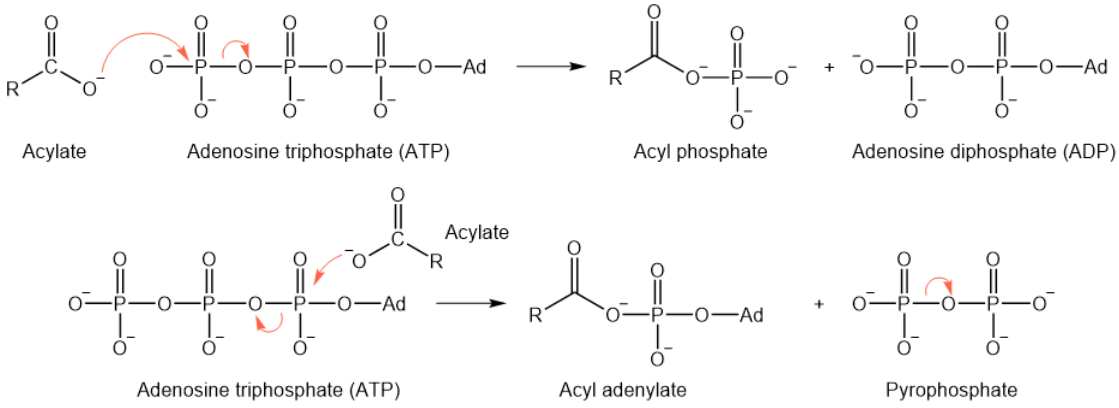
Carboxylic acids are common raw materials but must be converted into reactive derivatives, like acid halide, before conversion to other carboxylic acid derivates. Reagents like thionyl chloride ($SOCl_2$), phosphorous trichloride (PCl_3), or phosphorous pentachloride (PCl_5) are used to convert carboxylic acids into acid chlorides for this purpose, as shown below.



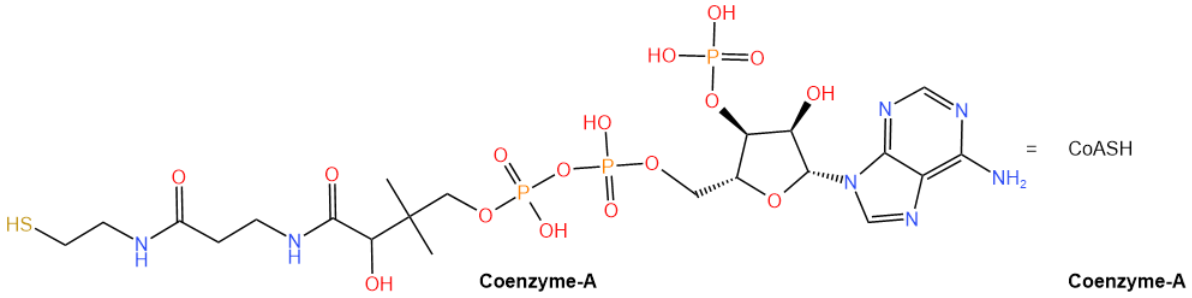
The above reagents are not available in biochemical systems. Adenosine triphosphate, shown below is a common reagent available in biochemical systems.



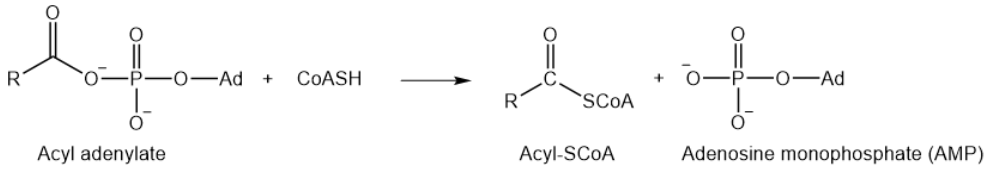
As illustrated below, adenosine triphosphate is used in biochemical systems to convert acylates into acyl phosphates or acyl adenylates.



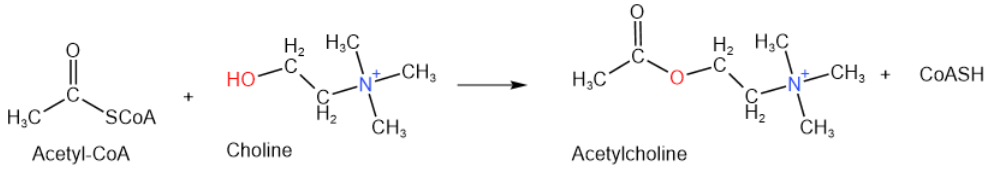
The phosphate and adenylate are good leaving groups. For example, coenzyme-A ($CoASH$) containing nucleophilic thiol ($-SH$) group is common in biochemical systems.



CoASH displaces adenylate group from acyls through nucleophilic acyl substitution mechanism, as illustrated below.

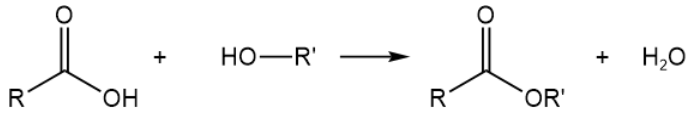


Thioesters, like $-SCoA$, are also good leaving groups. For example, the neurotransmitter acetylcholine is synthesized by reacting acetyl$-SCoA$ with choline through a nucleophilic acyl substitution mechanism, as shown below.



Condensation polymerization reactions

When two molecules combine to form a single molecule, usually with a loss of a small molecule like water or ammonia, the reaction is called a condensation reaction. For example, nucleophilic acyl substitution reactions between carboxylic acids and alcohols are condensation reactions, as shown below.



Polymers called polyesters are produced when molecules with two carboxylic acids and others with two alcohols condense to form esters. For example, benzene-1,4-dicarboxylic acid, commonly known as terephthalic acid, and ethane-1,2-diol, known as ethylene glycol, condense to produce a polyester called Polyethylene terephthalate (PET), as shown in Figure 4.5.2. Polyethylene terephthalate is used to make drink bottles and polyester fabrics.

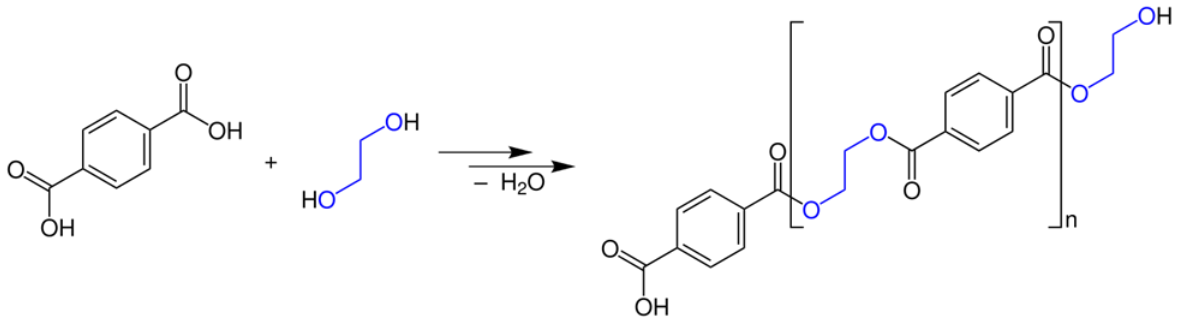


Figure 4.5.2: Polyesterificaiton reaction between terephthalic acid and ethylene glycol producing a polyester Polyethylene terephthalate (PET), where n is an integer. (Copyright; MaChe, Public domain, via Wikimedia Commons)

Nucleophilic acyl substitution reactions between carboxylic acids and amines are another example of condensation reactions that can produce polyamides. For example, condensation of hexanedioic acid and hexane-1,6-diamine, known as hexamethylene diamine, produces a polyamide called nylon 66, as illustrated in Figure 4.5.3.

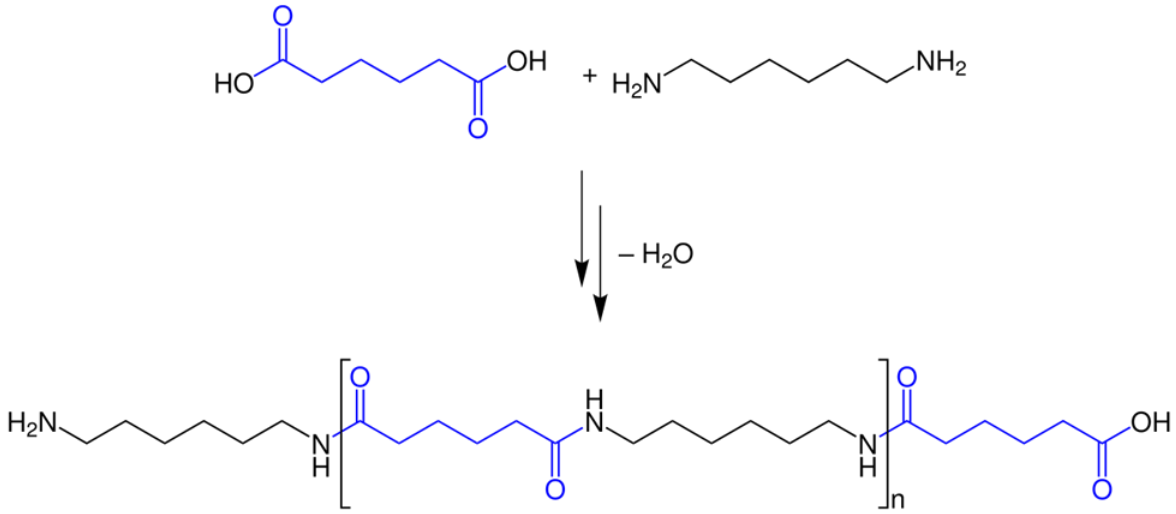


Figure 4.5.3: Polycondensiiton reaction between hexanedioic acid and hexamethylene diamine producing a polyamide called nylon 66, where n is an integer that can be in hundreds. (Copyright; MaChe, Public domain, via Wikimedia Commons)

The first digit of 66 in the name nylon 66, tells number of carbons in one monomer (hexanedioic acid) and the second digit for the other monomer (hexane-1,6-diamine). Nylone 66 is commonly used in clothing, fishing lines, and guitar strings. Kevlar used to make bulletproof jackets is another example of polyamide, shown in Figure 4.5.4

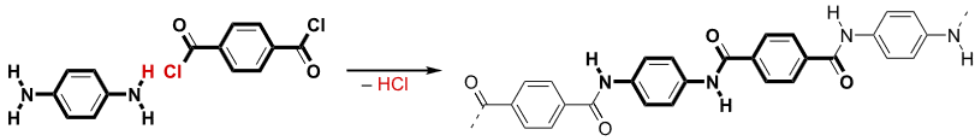


Figure 4.5.4: Reaction of benzene-1,4-diamine (*p* -phenylenediamine) and terephthaloyl chloride to form aramid, also called kevlar. (Copyright; Hbf878, CC0, via Wikimedia Commons,)

Condensation reaction between diisocyanates ($O{=}C{=}N{-}R{-}N{=}C{=}O$) and alcohols ($HO{-}R{-}OH$) produces polyurethane where urethane groups ($-\overset{\overset{H}{|}}{N}-\overset{\overset{O}{||}}{C}-O-$) link the monomer units in the polymer molecule, as illustrated in Figure 4.5.5.

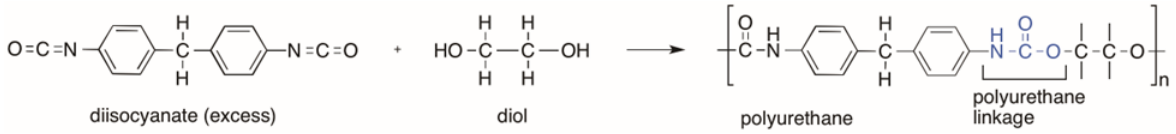
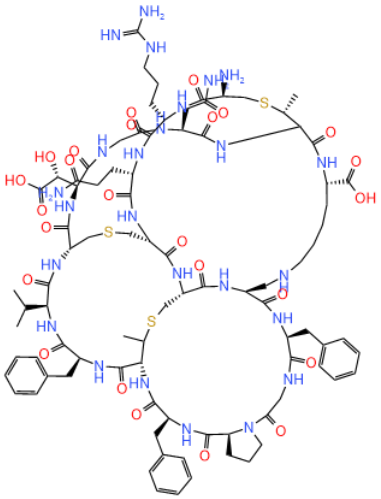
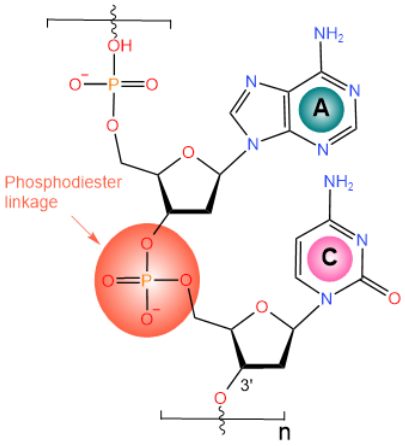


Figure 4.5.5: Synthesis of polyurethane from diisocyanate and a diol, wherein the urethane groups $-\overset{\overset{H}{|}}{N}-\overset{\overset{O}{||}}{C}-O-$ links the molecular units. (Copyright; RLM0518, CC BY-SA 3.0, via Wikimedia Commons)

Polyurethane makes foam, e.g., mattresses foam.

Nucleic acids, i.e., DNA and RNA are biopolymers in which phosphate diester linkages hold the monomer units together, as illustrated in the figure on the right (copyright: Public domain). Nucleic acids are described in detail in a separate chapter. Proteins are polyamide bipolymers. For example, cinnamycin (lanthiopeptin), shown in the figure on the left (copyright: Public domain), is an antibacterial peptide produced by *Streptomyces cinnamoneus* containing 19 amino acids. Proteins are also described in detail in a separate chapter.





Section of DNA

Cinnamycin (lanthiopeptin) -a polyamide

4.6: Nucleophilic Addition Reactions

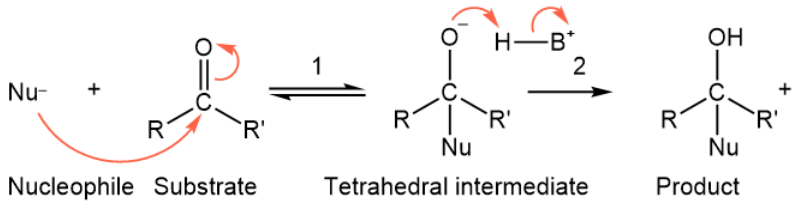
Learning Objectives

- Understand and be able to draw nucleophilic addition mechanisms in basic and acidic mediums.
- Apply nucleophilic addition reactions to simple examples, like the addition of cyanide, water, and alcohols to aldehydes and ketones.

The carbonyl group of an aldehyde or a ketone is a polar double bond just like the acyl group of a carboxylic acid derivative, i.e., $\mathrm{R-\overset{\delta+}{C}(=\overset{\delta-}{O})-R'}$ where the carbonyl $\overset{\delta+}{\mathrm{C}}$ is an electrophile. However, this electrophile does not have a leaving group on it. It can add a nucleophile but usually needs ether base-promoted or acid catalysis conditions. These two mechanisms are described next.

Base-promoted nucleophilic addition reactions

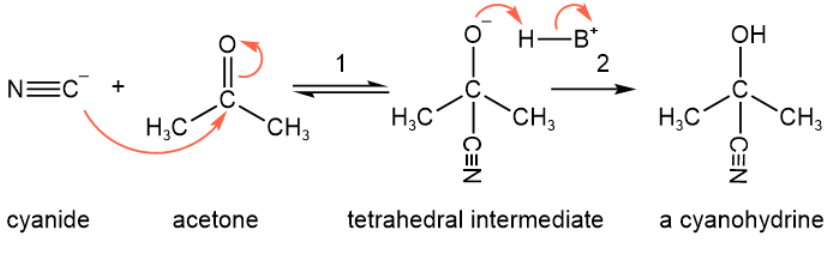
A nucleophile in its more reactive basic form $\mathrm{Nu^-}$ in basic medium attacks the electrophilic carbonyl $\overset{\delta+}{\mathrm{C}}$. Simultaneously, the π-bond bond break, as shown in step#1 of the mechanism below. The $\mathrm{O^-}$ either re-establish π-bond expelling the nucleophile, or it is neutralized by an acid-base reaction with any acid group present in the medium in step#2 of the mechanism shown below.

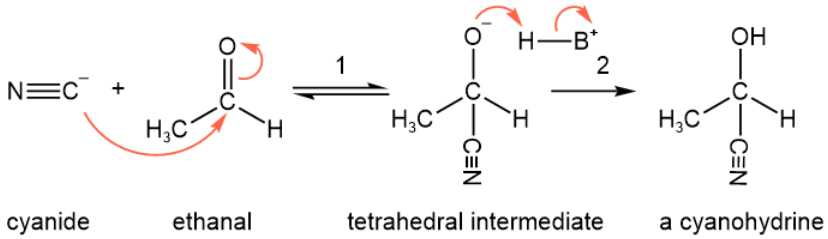


Remember, the medium is basic but amphoteric and has some acidic groups. If there is no sufficiently acidic group in the medium, an acid ($\mathrm{HB^+}$) is added after mixing the reactants for step#2 to happen. The carbonyl C changes hybridization from sp^2 to sp^3, and $\mathrm{C{=}O}$ group converts to an alcohol ($\mathrm{-OH}$) in this reaction. It is called nucleophilic addition reaction because the overall result is the addition of a nucleophile (HNu) to a carbonyl ($\mathrm{C{=}O}$) group.

Cyanohydrin formation

The addition of cyanide ($\mathrm{^-CN}$) to aldehydes or ketones is an example of base-promoted nucleophilic addition. It converts the $\mathrm{C{=}O}$ into cyanohydrin, i.e., a C with a cyanide ($\mathrm{-CN}$) and hydroxyl ($\mathrm{-OH}$ groups on it. Two examples of cyanohydrin formation are shown below.



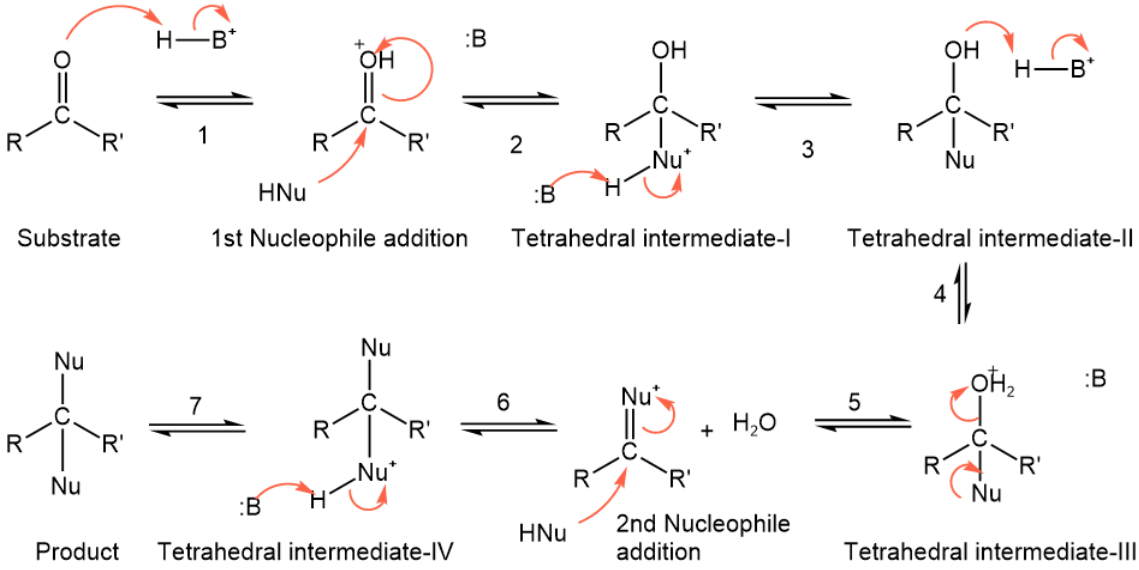


Acid-catalyzed nucleophilic addition reactions

$\overset{\delta-}{\underset{}{O}}$ $\overset{+}{O}H$

$\underset{\delta+}{\underset{||}{C}}$ $\underset{\delta+}{\underset{||}{C}}$

The acid in the medium converts less electrophilic $R-C-R'$ to a more electrophilic $R-C-R'$ form by protonating its O in step#1. Neutral nucleophile can exist only in neutral (HNu) form in the acidic medium. The neutral nucleophile (HNu) attacks the carbonyl C making a σ-bond and, simultaneously, the π-bond breaks in step#2. Although the medium is acidic, it is amphoteric and has some basic groups that receive acidic proton on the incoming $\boxed{\ce{-{\overset{+}{Nu}H}}}$ in step#3, as illustrated below.



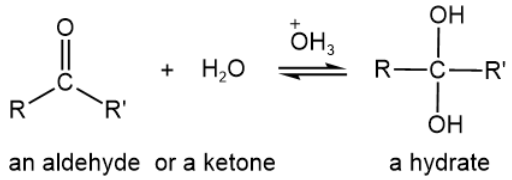
The acidic medium converts bad leaving $-OH$-group to a good leaving $-\overset{+}{O}H_2$-group in step#4. The $-\overset{+}{O}H_2$-group leaves as H_2O, pushed by lone pair on $-Nu$ in step#5. Second HNu attacks the electrophilic C in step#6, and the acidic proton on the second incoming $-\overset{+}{N}uH$ is removed by any base ($:B$) in the medium in step#7. All the steps of the mechanism are reversible. Two situations can arise:

1. if HNu is in limited supply, equilibrium exists between the substrate and tetrahedral intermediate-II of step#3, and
2. if HNu is in excess, equilibrium exists between the substrate and the product of reation#7.

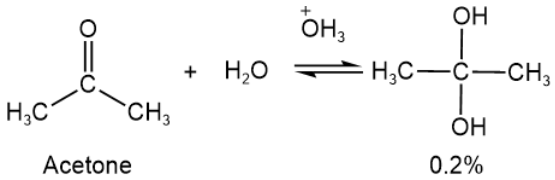
The reaction can be made irreversible by removing H_2O from the products. The reaction can be reversed by mixing the acid catalyst and excess water with the product.

Hydration of aldehydes and ketones

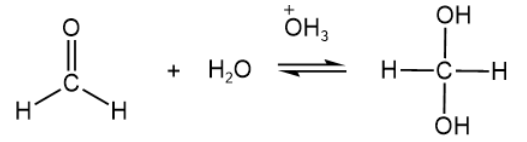
An example of nucleophilic addition is the hydration of aldehydes and ketones when diluted with water, as shown below.



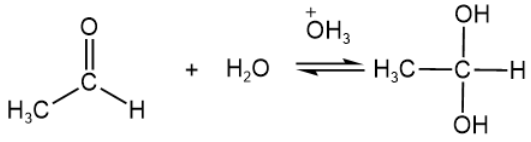
The hydrates are usually unstable and exist in a very small proportion relative to the initial aldehyde or ketone, e.g., the hydrate of acetone, shown below.



A few exceptions exist where one or both of the groups attached with the carbonyl C are H′s, hydrate form is dominant, e.g., the hydrated acetaldehyde and formaldehyde, as shown below.



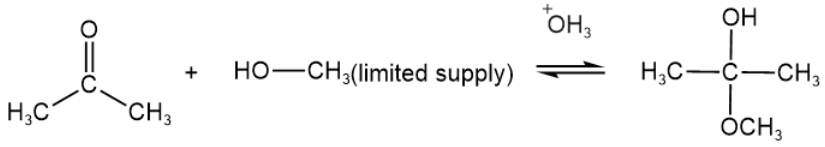
Formaldehyde 99.9%



Acetaldehyde 58%

Hemiacetals and acetals

The addition of alcohol (R−OH) to an aldehyde or a ketone is another example of a nucleophilic addition reaction. If R−OH is in limited supply, the reaction stops after the addition of one R−OH molecule resulting in a product called **hemiacetal**, as in the following example.

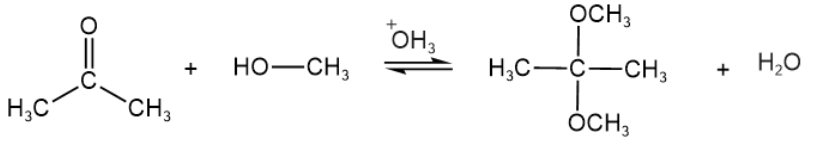


Acetone A hemiaceal

Hemiacetal has an −O−R-group and an −OH bonded to the same C. Hemiacetals are usually unstable and exist in small proportion at equilibrium. Exceptions are five-membered and six-membered hemiacetals formed by the reaction of (−OH)-group and the C=O-group on the same molecule. For example, more than 99% of glucose molecules exist in six-membered hemiacetal forms in equilibrium with open-chain aldehyde form in a water solution, as shown in Figure 4.6.1.

Figure 4.6.1: Illustration of glucose molecule interconverting between open chain aldehyde form and six-membered cyclic hemiacetal form. (Copyright; Wikimuzg, CC BY-SA 3.0, via Wikimedia Commons)

When alcohol is in excess, a second R−OH adds to the hemiacetal resulting in an **acetal**. An acetal has two −O−R-groups bonded to the same C, as shown in the example below.



Acetone Excess An acetal Water

The acetal groups in carbohydrates are called glycoside linkages. Glycoside linkages connect monosaccharide units together in polysaccharides like starch and cellulose and in disaccharides, e.g., lactose, maltose, and sucrose, as shown in Figure 4.6.2

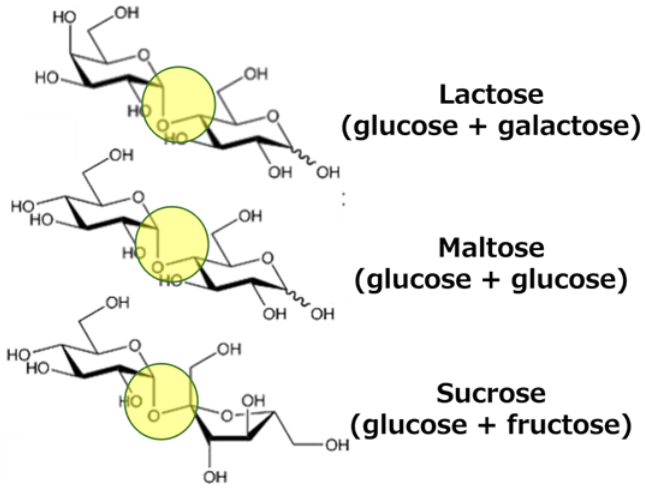


Figure 4.6.2: Lactose, Maltose, and Sucrose, each with the sugars that compose it and the glycosidic linkage highlighted. (Copyright; SrKellyOP, CC0, via Wikimedia Commons)

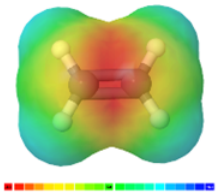
4.7: Electrophilic addition reactions

Learning Objectives

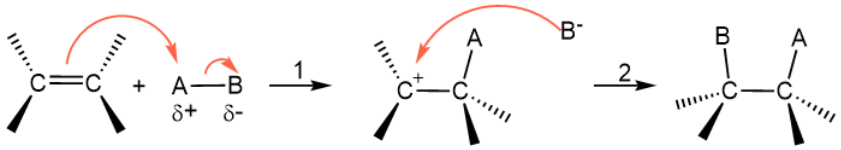
- Draw an electrophilic addition mechanism with curly arrows showing the movement of electrons.
- Apply the electrophilic addition to reactions of alkenes and alkynes with halogen acids and water.
- Define the concepts: regioselectivity, stereoselective and stereospecific reactions, and tautomerism with examples of electrophilic addition relations of alkenes and alkynes.

Electrophilic addition mechanism

The π-bond of alkenes distributes electrons above and below the σ-bond that creates a local electron-rich ($\delta-$) region, as shown by red color region in the following electrostatic potential map of ethene.



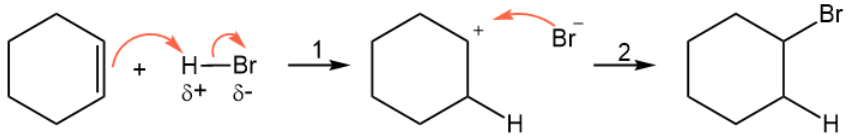
Therefore, the alkenes are nucleophiles ($\delta-$) attracted to electrophiles ($\delta+$) and capable of donating their π-electrons to make a bond with them, as shown in the following reaction mechanism.



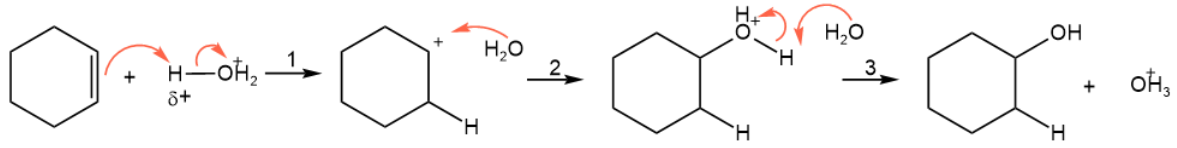
In first step of this mechanism, π-bond acts as an incoming nucleophile making a σ-bond with the electrophilic $\delta+$ end of the polar reagent $\overset{\delta+}{\mathrm{A}}-\overset{\delta-}{\mathrm{B}}$ while the $(\overset{\delta-}{\mathrm{B}})$ acts as a leaving group, leaving as a nucleophile B^-. One of the π-bonded carbon that receives the electrophile $\overset{\delta+}{\mathrm{A}}$ becomes sp^3 hybridized, and the other carbon stays sp^2 hybridized with an empty p-orbital as a carbocation. In the second step, the nucleophile (B^-) attacks the electrophilic carbocation and makes a new bond. As a result, the reagent $\mathrm{A}-\mathrm{B}$ add to the $\mathrm{C}=\mathrm{C}$-bond. Since the process begins with the addition of an electrophile $\overset{\delta+}{\mathrm{A}}$, it is called an electrophilic addition reaction.

Examples of electrophilic addition reactions of alkenes

Halogen acids ($\overset{\delta+}{\mathrm{H}}-\overset{\delta-}{\mathrm{Cl}}$, $\overset{\delta+}{\mathrm{H}}-\overset{\delta-}{\mathrm{Br}}$, and $\overset{\delta+}{\mathrm{H}}-\overset{\delta-}{\mathrm{I}}$) are among many reagents that add to alkenes by electrophilic addition mechanism, as shown in the following example.



Water ($\mathrm{H_2O}$) has more polar $\overset{\delta+}{\mathrm{H}}-\overset{\delta-}{\mathrm{O}}$ bonds than the halogen acids, but $-\mathrm{OH}$ is a bad leaving group and usually do not act as a leaving group. The bad leaving $-\mathrm{OH}$ is first converted to a good leaving $-\overset{+}{\mathrm{O}}\mathrm{H_2}$ by adding stronger acid like sulfuric acid ($\mathrm{H_2SO_4}$) in the water. Then, the hydronium ion ($\mathrm{H}-\overset{+}{\mathrm{O}}\mathrm{H_2}$) enters the first step of the electrophilic addition reaction, as shown below.

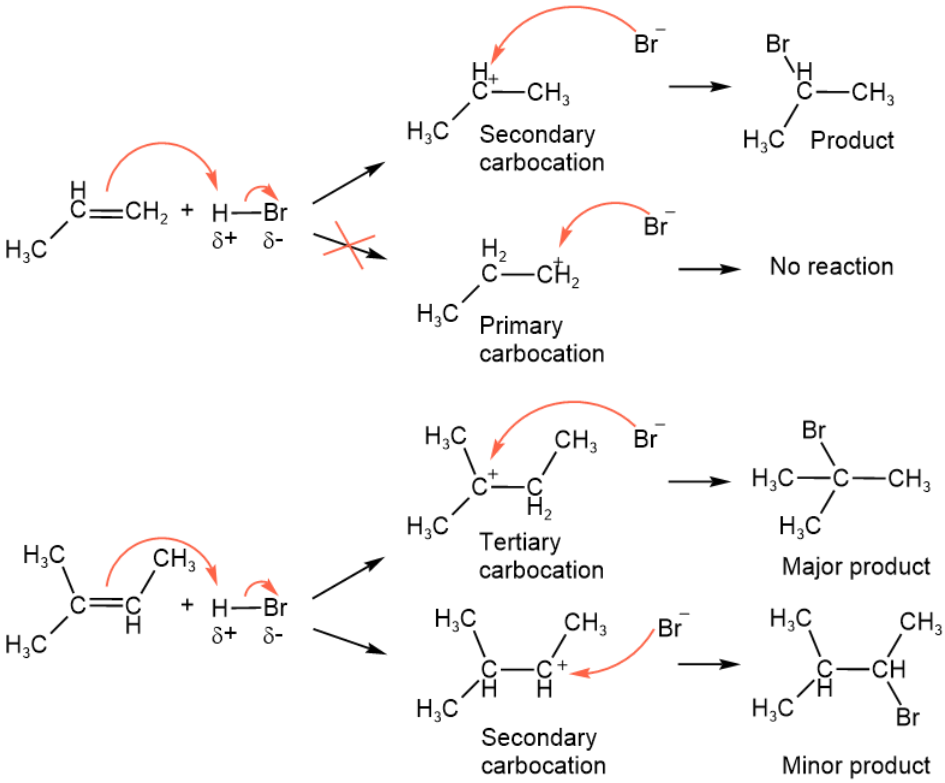


H_2O acts as an incoming nucleophile in step#2, another water molecule removes its acidic proton in step#3. This reaction is called hydration of alkene because the result is the addition of a water molecule $\overset{\delta+}{H}-\overset{\delta-}{OH}$ to a C=C-bond of an alkene.

Sulfuric acid is used as an acid catalyst because its conjugate base HSO_4^- is a poor nucleophile. H_2O is also a poor nucleophile, but being in higher concentration wins the completion in step#2. Even if HSO_4^- add to the carbocation in the second step, it is substituted by H_2O under the reaction conditions.

Regioselectivity of electrophilic addition reactions

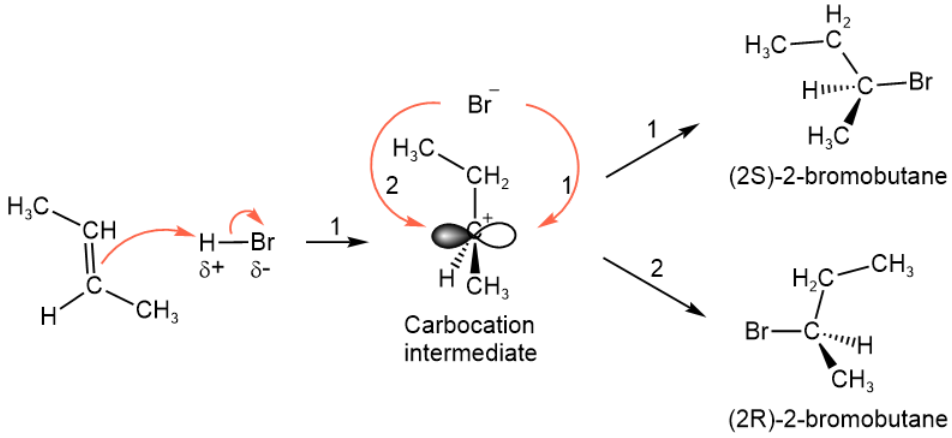
In the cases of unsymmetrical alkenes, two constitutional isomers are possible, but one is produced exclusively or as a major product. The reason is that the two potential carbonation intermediates have unequal stability. The stability of carbocations increases in this order: primary $(R\overset{+}{C}H_2)$ < secondary $(R_2\overset{+}{C}H)$ < tertiary $(R_3\overset{+}{C})$. The following examples show that the more stable carbocation is formed exclusively or predominantly when there is a choice.



Chemical reactions in which two constitutional isomers are possible, but one is formed exclusively or predominantly are called **regioselective reactions**. Electrophilic additions are regioselective reactions.

Stereoselectivity of electrophilic addition reactions

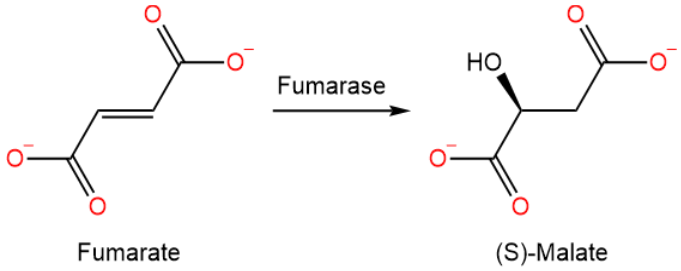
When reactants and catalysts are achiral, but the product is chiral, it is produced equally in both (R) and (S) configurations. For example, the carbocation intermediate in electrophilic addition reactions is trigonal planer. The nucleophile can attack it from either side with about equal probability resulting in a racemic mixture, as illustrated in the following example.



However, suppose one of the reactants or the catalyst is chiral and a pure enantiomer. In that case, one of the enantiomers (R or S) is usually formed exclusively or as a major product. Enzymes in biochemical systems are chiral catalysts that usually produce one enantiomer exclusively.

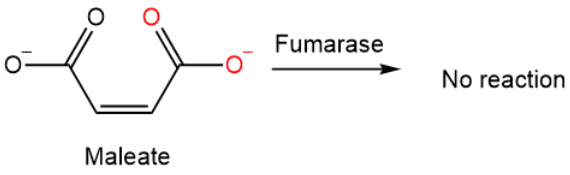
> Chemical reactions that produce one stereoisomer exclusively or preferentially relative to the others are called **stereoselective reactions**.

Enzyme-catalyzed reactions are stereoselective reactions. For example, hydration fumarate, catalyzed by an enzyme fumarase, produces (S)-malate, as shown below.



The R-enantiomer is not produced in the above reaction.

Enzymes usually also react with one stereoisomer in the reactant exclusively. For example, fumarase catalyzes the hydration of fumarate but does not react with its enantiomer maleate, as illustrated below.

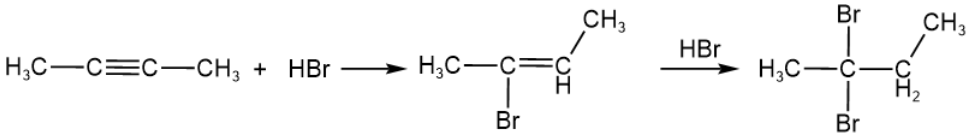


> Chemical reactions that preferentially react with one stereoisomer among reactant and preferentially produce one stereoisomer in the products are called **stereospecific reactions**.

Enzyme-catalyzed reactions are often stereospecific reactions. For example, the fumarase enzyme selectively reacts with fumarate (and not with its stereoisomer maleate) and selectively produces (S)-malate (and not its stereoisomer (R)-malate).

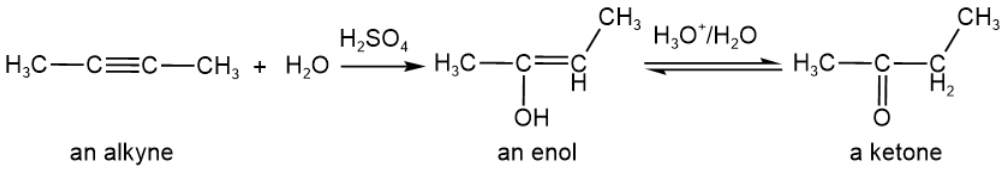
Electrophilic addition reactions of alkynes

Alkynes have a triple $\mathrm{C \equiv C}$-bond, i.e., one σ-bond and two π-bonds. The electrophilic addition reactions happen on π-bonds of alkynes. If the reagent is in a one-to-one mole ratio, it adds to one of the π-bonds. If the reagent is in excess, a second addition reaction happens on the product of the first addition, as shown below.



In step#2, the nucleophile (\(\ce{Br} in this case) adds to the carbon carrying the nucleophile of step#1. This is because the carbonation intermediate of step#2 is stabilized by resonance with the lone pair of electrons on the nucleophile added in the first step.

Hydration of alkyne follows the same step#1 as for alkenes, but the intermediate enol containing $\mathrm{C{=}C}$ and $-\mathrm{OH}$ together exists in equilibrium with its structural isomer ketone, as shown in the following example.



> **Tautomers** are constitutional isomers that are readily interconvertible. **Tauomerization** is the chemical reaction that interconverts the tautomers.

Enol and its isomer ketone, e.g., but-2-ene-2-ol and but-2-one in the above example, are called **tautomers,** and the equilibrium between the two in step#2 of the above reaction is **tautomerization**.

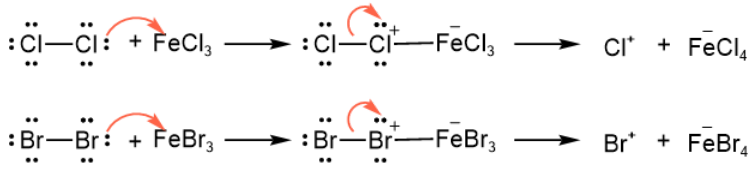
4.8: Electrophilic aromatic substitution reactions

Learning Objectives

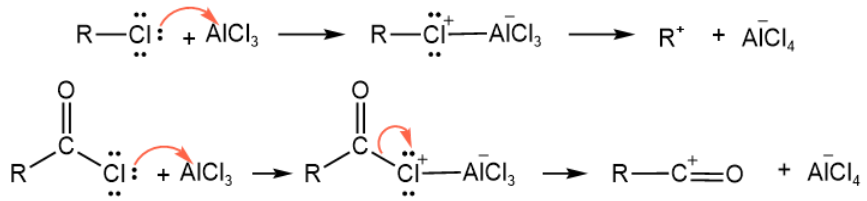
- Understand the difference in the electrophilicity of π-bond of benzene and alkenes.
- Draw the electrophilic aromatic substitution mechanism with curly arrows showing the flow of electrons.
- Apply the electrophilic aromatic substitution to some reactions of benzene, including halogenation, nitration, sulfonation, alkylation, and acylation reactions.

Which electrophiles can react with an aromatic substrate?

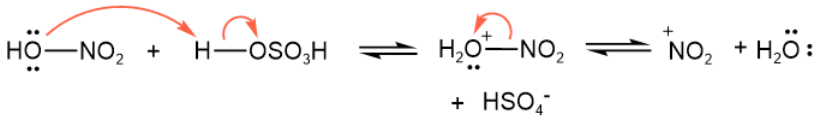
The π-bonds in a benzene ring of aromatic compounds are weaker nucleophiles than the π-bonds in alkenes. This is because breaking a π-bond of alkene costs about 260 kJ/mole energy, but breaking a π-bond in an aromatic substrate costs an additional 208 kJ/mol because the aromatic stabilization is lost. Unlike alkenes, the aromatic substrates do not react with partial positive ($\overset{\delta+}{A}-\overset{\delta-}{B}$) electrophiles. The aromatic substrates react with electrophiles in their most reactive cation E^+ form. The cation electrophiles are generated in situ by acid-base or Lewis acid-Lewis base reactions. For example, halogens ($X-X$ react as Lewis bases with Lewis acids like AlX_3 or FeX_3, where X is a halogen atom (Cl or Br). The Lewis acid receives a lone pair from one halogen atom causing a heterolytic breaking of $X-X$. The other halogen leaves as X^+, as shown below.



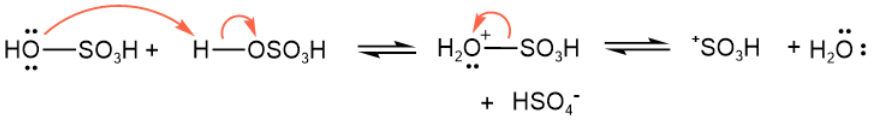
Similar reaction happens when an alkylhalide ($R-X$ or an acyl halide ($R-\overset{O}{\overset{\|}{C}}-X$) reacts with Lewis acid, as shown below.



An $-OH$ bonded with a potential electrophile E^+ can be converted into a better leaving $-\overset{+}{O}H_2$ group by adding a stronger acid to the substrate. The $-\overset{+}{O}H_2$ leaves as neutral neucleophile H_2O, leaving behind the E^+. For example, protonation of nitric acid with sulfuric acid generated nitronium ion ($\overset{+}{N}O_2$), as shown below.

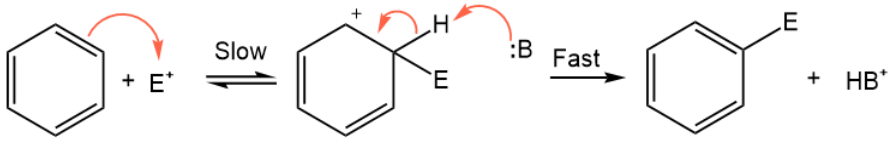


Like the auto-ionization of water, the autoionization of sulfuric acid followed by elimination of H_2O generates protonated sulfur trioide ($\overset{+}{S}O_3H$), as shown below.



Electrophilic aromatic substitution mechanism

The nucleophilic π-bond of an aromatic compound attacks the cation electrophile (E^+), as shown in step#1 in the mechanism illustrated below. Any base group in the medium removes the acidic proton that re-establishes the π-bond in Step#2.

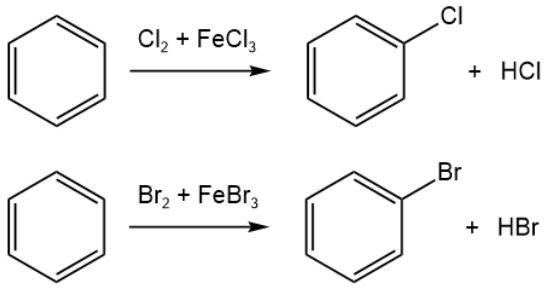


Removal of the proton by a base is preferred over electrophile attacking the carbonation intermediate in step#2, because aromatic stabilization decreases the energy barrier for the former. It is called **electrophilic aromatic substitution reaction** because an electrophile E^+ substitutes another electrophile H^+ from an aromatic substrate.

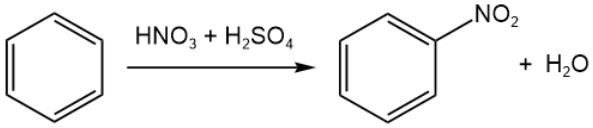
Examples of electrophilic aromatic substitution reactions

Some fo the important electrophilic aromatic substitution reactions of benzene are listed below.

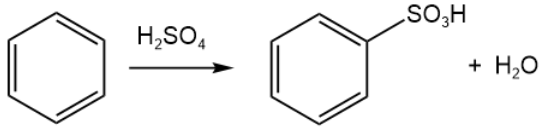
- **Halogenation of benzene** substitutes a $-H$ with a halogen ($-Cl$ or $-Br$), as shown below.



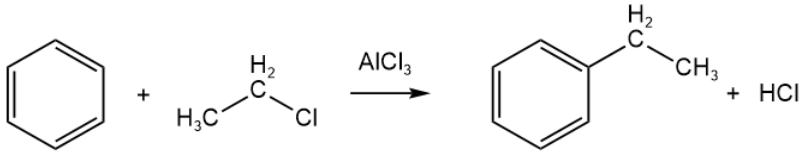
- **Nitration of benzene** substitutes a $-H$ with nitro group ($-NO_2$) by the following reaction.



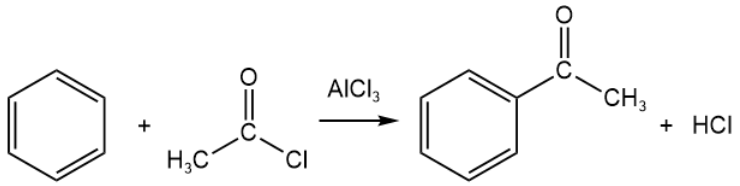
- **Sulfonation of benzene** substitutes a $-H$ with sulfonic acid group ($-SO_3H$) by the following reaction.



- **Alkylation of benzene** substitutes a $-H$ with alkyl acid group ($-R$), e.g.,:



- **Acylation of benzene** substitutes a $-H$ with acyl group ($-\overset{\overset{O}{\|}}{C}-R$), e.g.,:



4.9: Reduction and oxidation (redox) reactions

Learning Objectives

- Define the redox process and calculate the oxidation number of an atom in an organic compound.
- Learn catalytic reduction of alkenes and carbonyl compounds.
- Learn reduction of carbonyl compounds by $NaBH_4$ or $LiAlH_4$, and by $NADH$ in living things.
- Learn oxidation of alcohols and aldehydes by oxidizing agents in laboratories and by NAD^+ in living things.
- Learn oxidation of thiol ($-SH$) to disulfide ($-S-S-$) and its revers, in proteins.

What are oxidation and reduction?

Oxidation is i) the loss of electrons, ii) the gain of O, or iii) the loss of H. **Reduction** is the opposite, i.e., i) the gain of electrons, ii) the loss of O, or iii) the gain of H. The reduction and oxidation reactions are coupled, called **redox** reactions. Electrons' loss or gain is easily recognized in inorganic chemical reactions. For example, consider the following reaction:

$4\,Fe + 3\,O_2 \longrightarrow 2\,Fe_2O_3$,

Fe lost three three electrons to become Fe^{3+} and O gained two electrons to become O^{2-} in this reaction.

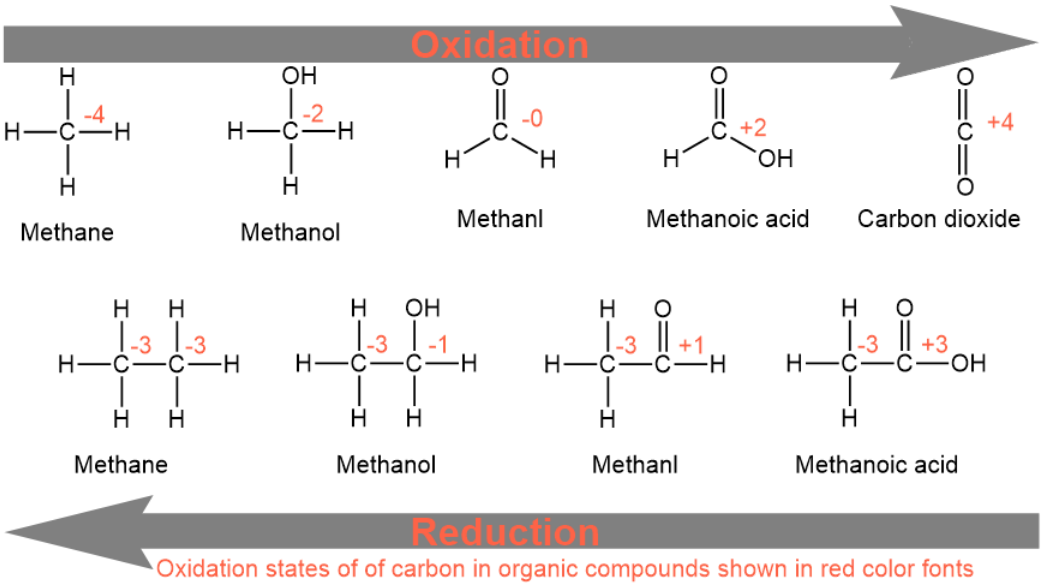
Oxidation number

Organic chemicals usually have covalent bonds where the loss or the gain of electrons from an atom is accounted for by **oxidation number.** An increase in the oxidation number is oxidation, and a decrease in the oxidation number is reduction. The following is a simple procedure for assigning an oxidation number to an atom in a neutral organic compound. Start with oxidation number zero for a free atom,

1. a bond with a more electronegative atom increases the oxidation number by one, e.g., a $C-O$ bond increases the oxidation number of C by +1;
2. a bond with a less electronegative atom decreases the oxidation number by one, e.g., a $C-H$ bond decreases the oxidation number of C by -1; and
3. a bond with the same atom does not change the oxidation number, e.g., a $C-C$-bond does not change the oxidation number of C.

For example, in CH_4 four bonds of $C$with less electronegative H's decrease its oxidation number to -4. H's. In CH_3-OH three bonds of $C$with less electronegative H's decrease its oxidation number to -3 and one bond with more electronegative O increases to +1, with the overall oxidation number of the C = -3 +1 = -2. So, conversion of methane (CH_4) to methanol (CH_3-OH) is oxidation.

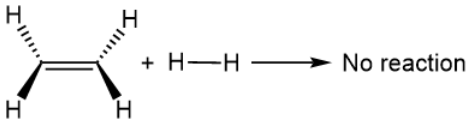
The following chart illustrates the oxidation states of C in organic compounds and their changes with organic transformations.



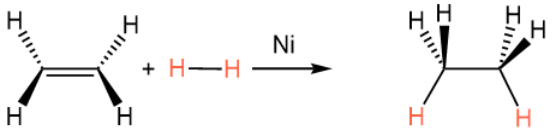
Examples of organic reduction reactions

Catalytic hydrogenation of alkenes

When an alkene and hydrogen are mixed, no reaction happens, as shown in the following example.



This is because alkene is an electrophile but $\mathrm{H-H}$ is a nonpolar bond that is not a nucleophile. Finely divided nickel (Ni), platinum or palladium supported on carbon, i.e. $\mathrm{Pd/C}$ or $\mathrm{Pt/C}$ catalyze the reaction, as shown below.



Mechanism of catalytic hydrogen of alkenes

In step#1, the catalyst absorbs $\mathrm{H-H}$ and $\mathrm{RR'C{=}CR''R'''}$ on the surface, as illustrated in Figure 4.9.1. In step#2, $\mathrm{H-H}$-bond breaks. $\mathrm{H's}$ and the π-bond of alkene become bonded with the catalyst surface. The surface-bonded species can migrate along the surface. In step#3, the π-bond attacks and makes a bond with H at the expense of breakage of $\mathrm{H-}$catalyst bond. The second C of the π-bond stays bonded with the catalyst. In step#4, $\mathrm{C-}$catalyst surface bonding electrons make a bond with the second H at the expense of breakage of $\mathrm{H-}$catalyst and $\mathrm{C-}$catalyst bonds. The alkane product of step#4 is no more attached to the catalyst and departs. The refreshed catalyst surface repeats the process.

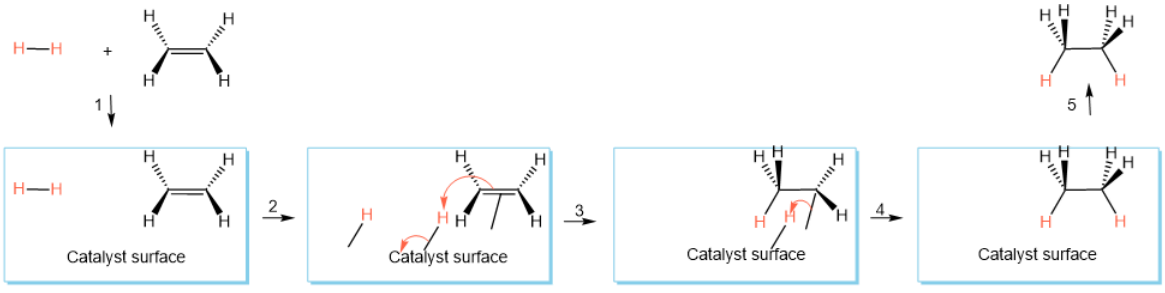


Figure 4.9.1: Mechanism of catalytic hydrogenation of alkene. (Copyright; Public domain)

Alkenes are planer around $\mathrm{C{=}C}$ bond. Both $\mathrm{H's}$ add from the same face of alkene facing the catalyst surface.

Syn addition is the addition of two substituents from the same side (same face) of the π-bond. Addition of $H-H$ to $C=C$ of an alkene is an example of syn addition.

Catalytic hydrogenation of vegetable oils

Vegetable oils are fats of tri-ester of glycerol with long chain fatty acids containing one or more $C=C$ bonds. The $C=C$ bonds create kinks in the long, preventing the chains' packing and resulting in lower melting points. Partial hydrogen of $C=C$ bonds in vegetable oil is carried out to convert it to semisolid margarine, as illustrated in Figure $4.9.2$. Hydrogenation makes them straight-chain alkanes that pack nicely and increases their melting points.

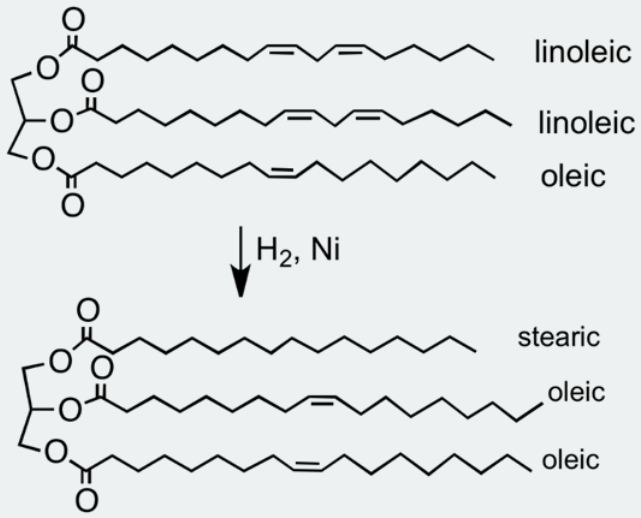
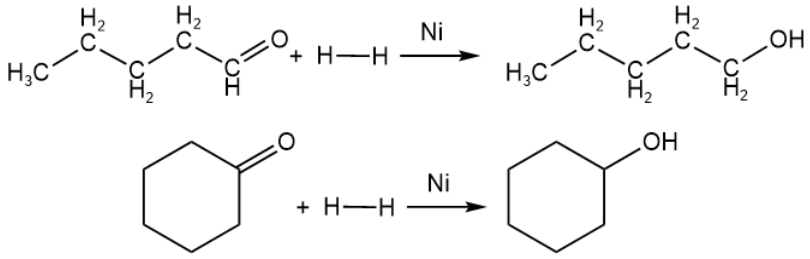


Figure $4.9.2$: Illustration of partial hydrogenation of vegetable oil (rapeseed oil in this example) to convert it to semisolid margarine. (Copyright; Smokefoot, CC BY-SA 3.0, via Wikimedia Commons)

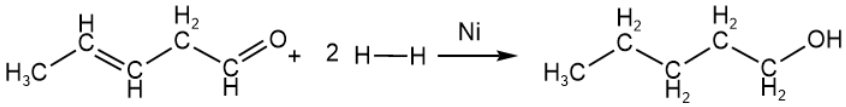
Reduction of aldehydes and ketones

Catalytic hydrogenation of aldehydes and ketones

Hydrogen (H_2) reduces aldehydes and ketones to alcohols in the presence of a catalyst like finely divided Ni, Pd/C, or Pt/C. as shown in the following examples.

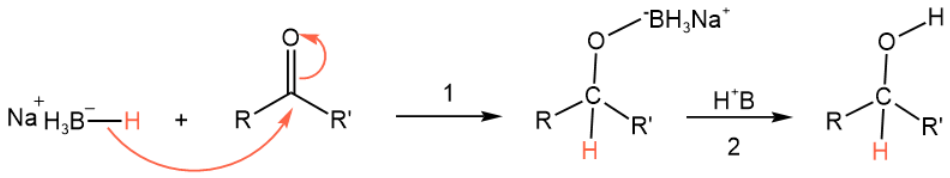


If $C=C$-bond and a $C=O$-bond are present in the same molecule, catalytic hydrogen reduces both, as shown below.



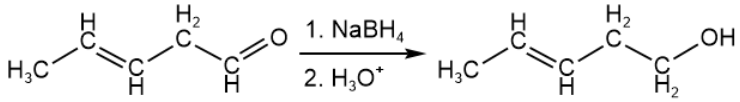
Reduction of aldehydes and ketones by nucleophilic addition reaction

The $C=O$ group can be selectively reduced to $C-OH$ group by nucleophilic addition of hydride (H^-) ion to the electrophilic C of $\overset{\delta+}{C}=\overset{\delta-}{O}$ group, by simplified mechanics shown below.



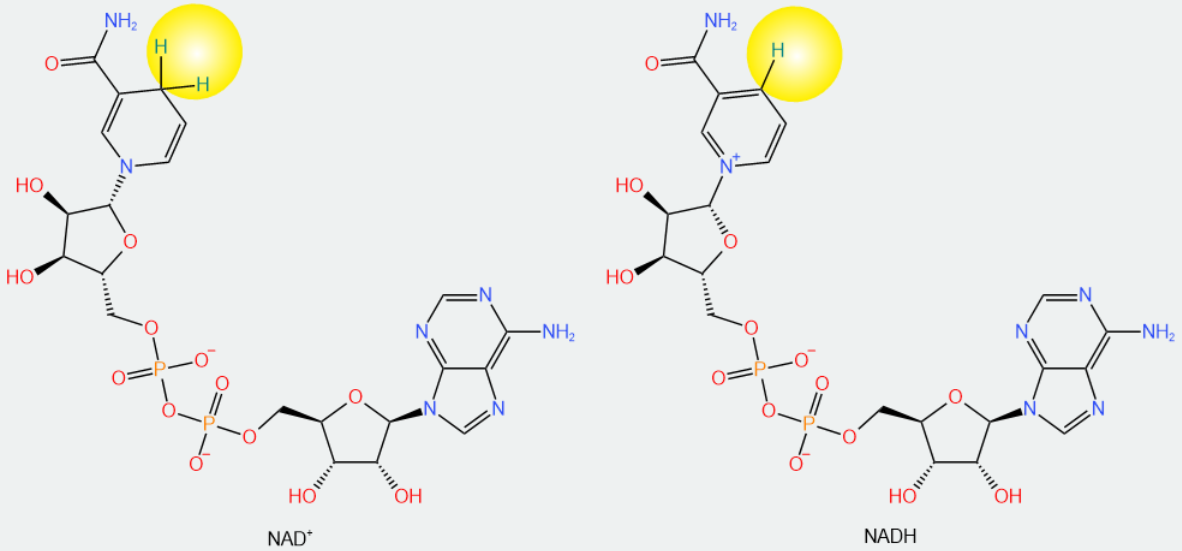
The H^- is not employed in free form, but it is donated by a reducing agent like sodium borohydride $NaBH_4$ or lithium aluminum hydride $LiAlH_4$. Note that the H^- do not survive in an acid medium, so the H^- reacts in the first step, and then acid is added to donates a proton to O in the second step.

The nucleophilicity of H^- ion depends on the donor reagent. For aldehydes and ketones, $NaBH_4$ and $LiAlH_4$ work, but usually $NaBH_4$ is employed as it is tolerant to water. $NaBH_4$ reduces aldehydes, ketones, and acid halides but does not reduce other carboxylic acid derivatives and C=C, as shown in the example below.

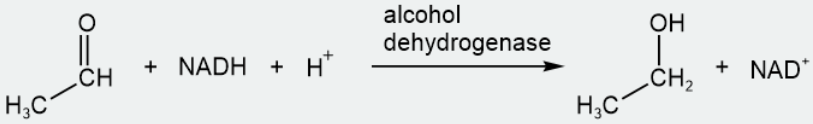


Reduction of aldehydes and ketones in biological systems

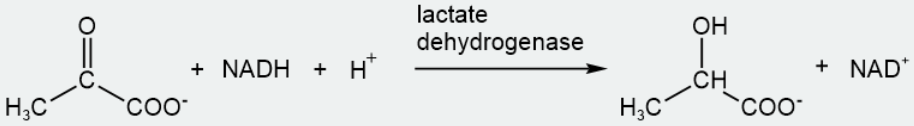
$NaBH_4$ and $LiAlH_4$ are used as H^- donors in laboratories. **Nicotinamide adenine dinucleotide** (NADH) coenzyme in its reduced form NADH is used as a hydride donor in biological systems. NADH is oxidised to NAD^+ after donating H^-, as illustrated below.



Enzyme-catalyzed conversion of acetaldehyde to ethanol during fermentation is an example of reduction using NADH.



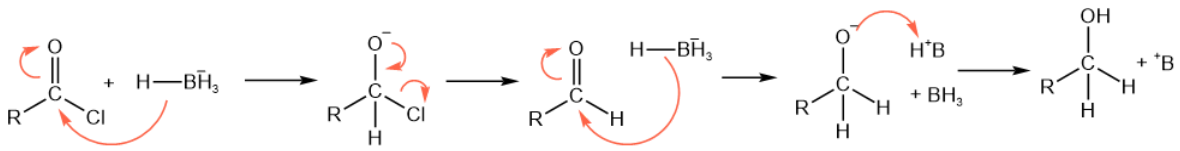
Conversion of pyruvate to lactate during glycolysis under anaerobic (absence of oxygen) conditions is another example of reduction using NADH, as shown below.



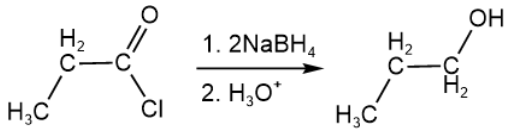
Cumulation of lactate after thought exercises causes muscle fatigue.

Reduction of carboxylic acid their derivatives

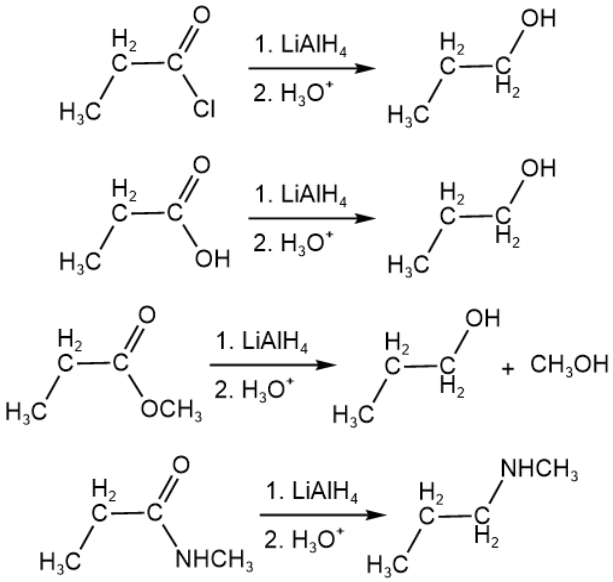
The $NaBH_4$ is less reactive and reduces only acid halids among carboxylic acids and their derivatives. $LiAlH_4$ is more reactive and reduces carboxylic acids and their derivatives. The mechanism is nucleophilic acyl substitution by H^- followed by nucleophilic addition, as shown below in the simplified form.



The reduction of propanoyl chloride by $NaBH_4$ is shown below.



$NaBH_4$ does not reduce the rest of the carboxylic acids and their derivatives. $LiAlH_4$ reduces carboxylic acids and their derivatives, as illustrated below with examples of reduction of acid halide, carboxylic acid, ester, and amide.



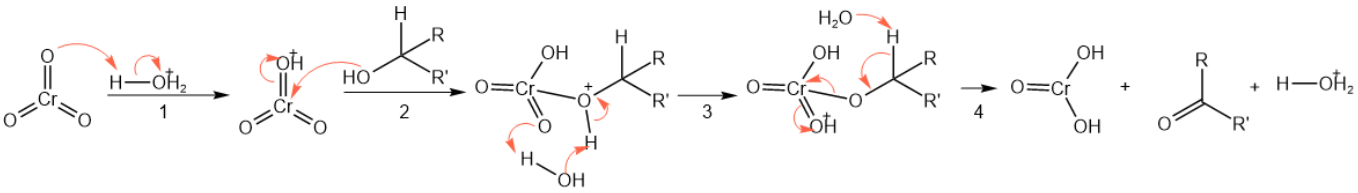
Reduction of disulfide $(S-S)$ bond in biological systems

The S−S bond commonly occurs in the protein tertiary structure. It is responsible for straight or curly hair shapes. The S−S bond can be reduced to −SH group by a variety of reducing agents, including $NaBH_4$ and Na metal. A thiol-disulfide exchange reaction often achieves this conversion in biological systems, as in the following example.

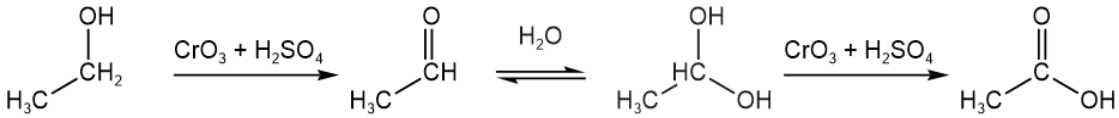
$$R-S-S-R + 2\,HOCH_2CH_2SH \rightleftharpoons 2\,R-SH + HOCH_2CH_2S-SCH_2CH_2OH$$

Oxidation of alcohols and aldehydes

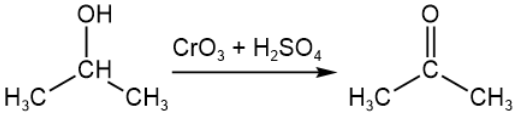
Primary and secondary alcohols are oxidized by several oxidizing agents, including Jones reagent (a mixture of chromium trioxide (CrO_3) and sulfuric acid (H_2SO_4)), chromic acid ((H_2CrO_4), or potassium dichromate (KCr_2O_7) in sulfuric acid (H_2SO_4) acid. The mechanism with Jones reagent is shown below.



Primary alcohols are oxidized to aldehydes. Aldehydes convert to hydrates by adding water, and the hydrates are oxidized to carboxylic acids, as illustrated below.



Secondary alcohols are oxidized to a ketone, as shown in the following example. Ketone is not oxidized further as they do not have hydrogen on the carbonyl carbon needed to carry out step#4 of the mechanism.



Tertiary alcohols are not oxidized for the same reason, i.e., there is no hydrogen on the carbon to which alcohol is attached to carry out step#4.

Tests for aldehydes and reducing sugar based on easy oxidation of aldehydes

The fact that an aldehyde is oxidized further easily while ketone is not was used as a test to distinguish between aldehydes and ketones. In this test **Tollens' reagent** ($[Ag(NH_3)_2]OH$) is mixed with the test sample. Ketones do not react but aldehydeis oxidized by the following reaction:

$$2\,[Ag(NH_3)_2]OH + R{-}CHO \longrightarrow 2\,Ag(s) + 4\,NH_3 + R{-}COOH + 2\,H_2O$$

Ag produced in the above reaction deposits and forms silver mirror on the wall of the test tube as shown in Figure 4.9.3

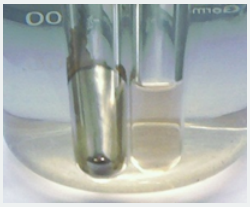


Figure 4.9.3: Tollen's test for aldehydes: left side positive (silver mirror formed) and right side negative. (Copyright; FK1954, Public domain, via Wikimedia Commons)

Benedict's reagent is a mixture of sodium carbonate (Na_2CO_3), sodium citrate ($Na_2C_6H_5O_7$), and copper(II) sulfate pentahydrate ($CuSO_4 \cdot 5\,H_2O$). Benedict's reagent produces red-precipitate (Cu_2O) as a positive test of aldehyde by the following reaction.

$$2\,Cu(C_6H_5O_7)^- + 5\,OH^- + R{-}CHO \longrightarrow Cu_2O(s, red) + 2\,C_6H_5O_7^{3-} + 3\,H_2O$$

This test distinguishes aldehydes from ketones and is also positive for reducing sugars (carbohydrates) that contain an aldehyde group, as shown in Figure 4.9.4.

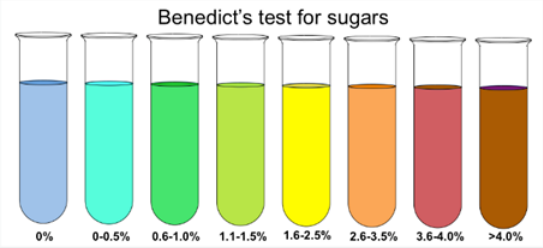


Figure 4.9.4: Benedict's test for aldehydes and reducing sugars that contain an aldehyde group. (Copyright; Thebiologyprimer, CC0, via Wikimedia Commons)

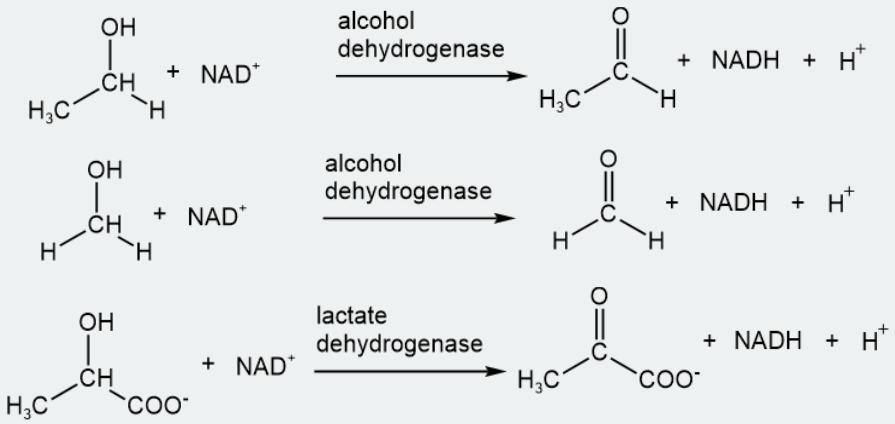
Alcohol exhaled in breath is oxidized by dichromate $Cr_2O_7^{2-}$ ions along with producing green color chromium(III) ions (\ce{Cr^{3+}}\) by the following reaction.

$$CH_3CH_2OH + Cr_2O_7^{2-}(\text{reddishorange}) \xrightarrow{H2SO4+H2O} CH_3COOH + Cr^{3+}(\text{green})$$

This reaction is used to test alcohol in drunk-and-drive cases by law enforcement.

Oxidation of alcohols in living systems

Alcohols are oxidized by nicotinamide adenine dinucleotide (NAD^+) to aldehydes and ketones in the cells. For example, ethanol, methanol, and lactate are oxidized to ethanal, methanal, and pyruvate, respectively, as shown below.



Unpleasant effects of drinking alcohol, i.e., flushing, nausea, dizziness, sweating, headache, and low blood pressure, are caused by ethanol (acetaldehyde). Antabuse drugs prohibit aldehyde dehydrogenase from converting acetaldehyde to acetic acid, causing the unpleasant effects of acetaldehyde to persist. Methanal produced by the oxidation of methanol is a poisonous substance that damages many tissues and causes blindness. It is used to poison ethanol meant to be used as a solvent. In methanol poison medical cases, ethanol is injected intravenously into the patient to compete with methanol for alcohol dehydrogenase enzyme. This way, methanol is excreted before converting to poisonous methanal in the body.

Oxidation of thiol ($-SH$) to disulfide ($-S-S-$) linkage

Thiol ($-SH$) is present in aminoacid cysteine. Two cysteine units in the same or different protein chains can oxidize their thiol groups to form disulfide ($-S-S-$) linkage. The disulfide linkage plays an important part in defining the tertiary structure of proteins. This oxidation reaction can be carried out easily by a variety of oxidizing agents, including iodine (I_2), bromine (I_2), and oxygen (O_2), as shown in an example below.

$$2\,R-SH + \tfrac{1}{2}O_2 \longrightarrow RS-SR + H_2O$$

4.10: Reactions with cyclic transition state

Learning Objectives

- Learn examples of reactions that involve five- or six-member transition state, including cyclonic hemiacetal formation of monosaccharides, Diels-Alter reactions producing six-membered cyclic products, and decarboxylation of β-keto acids.

Intramolecular reactions happen if the two reacting groups are on the same molecule and can come to a bonding distance through a five- or six-member cyclic transition state. Some examples of it are described in the next sections.

Cyclic hemiacetal formation of monosaccharides

Monosaccharides, like glucose, fructose, galactose, etc., have a $\mathrm{C{=}O}$-group on one C and $-\mathrm{OH}$-group on every other C. $-\mathrm{OH}$-group can add to $\mathrm{C{=}O}$-group forming a hemiacetal. Monosaccharides exist primarily in a five- or six-membered hemiacetal form because one of their $-\mathrm{OH}$-group can form a five- or six-membered transition state for the reaction, as shown in Figure $4.10.1$ for the case of D-glucose and D-fructose.

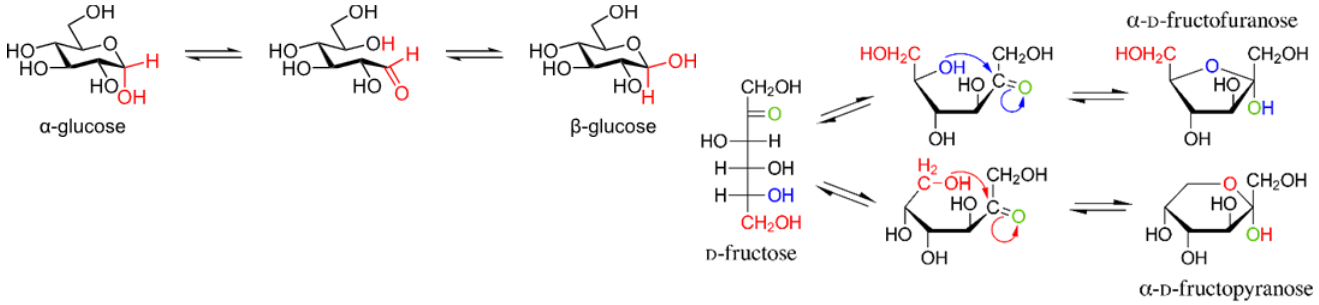


Figure 4.10.1: Illustration of five- or six-member cyclic hemiacetal formation of D-glucose and D-fructose through a cyclic transition state. (Copyright; Glucose: Calvero (talk · contribs), Public domain, via Wikimedia Commons, and Fructose: Vaccinationist, Public domain, via Wikimedia Commons)

Diels–Alder reaction

A conjugated diene, e.g., butadiene, and an alkene, e.g., ethene, make a cyclic six-member transition state. They react by the Diels-Alder reaction mechanism and produce a six-member cyclic product. This reaction can be intermolecular, e.g., between butadiene and then, or intramolecular, e.g., in the biosynthesis of antibiotic lovastatin, illustrated in Figure $4.10.2$

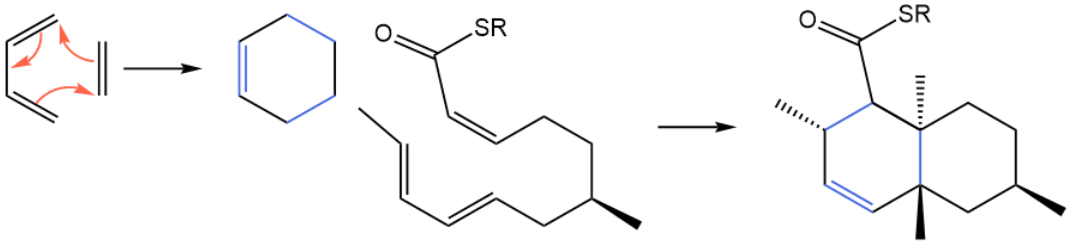


Figure 4.10.2: Mechanism of Diels-Alter reaction illustrated with the example of reaction between butadiene and ethene (left) and biosynthesis of lovastatin (right). Three π-bonds break, and two σ-bonds and one π-bond form, shown in blue. (Copyright: Public domain)

Decarboxylation

Decarboxylation is the removal of carbon dioxide ($\mathrm{CO_2}$) from a carboxylic acid ($\mathrm{R{-}COOH}$), as in this example: $\mathrm{R{-}COOH} \xrightarrow{\Delta} \mathrm{R{-}H} + \mathrm{CO_2}$.

This reaction requires high temperatures, such as in the thermal decomposition process. However, if there is a second carbonyl ($\mathrm{C{=}O}$) group β} to the $-\mathrm{COOH}$ group, it can easily acquire a six-member transition state and decarboxylate at moderate temperatures, as illustrated in Figure $4.10.3$

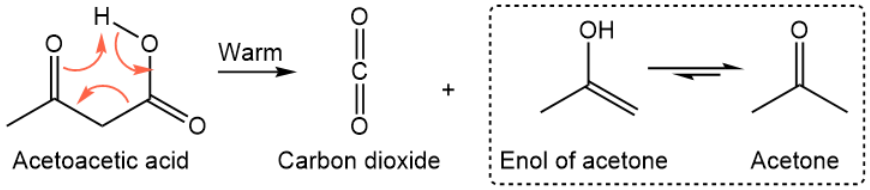
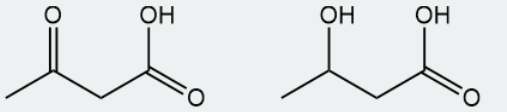


Figure 4.10.3: Decarboxylation mechanism of β-keto carboxylic acids illustrated with the example decarboxylation of acetoacetic acid. (Copyright: Public domain).

Ketone bodies and diabetes mellitus

Acetoacetic acid and its reduced product β-hydroxybutyric acid, shown below, are produced in the liver as a result of the metabolism of fatty acids and some amino acids.



Acetoacetic acid and β-hydroxybutyric acid are called **ketone bodies**. Their concentration in the blood of healthy persons is about 0.01 mmol/L but in persons suffering from starvation or diabetes mellitus may be up to 500 times higher.

Carboxylic acids exist as carboxylate anions under physiological conditions. Decarboxylation of the β-keto carboxylates happens spontaneously under physiological conditions. For example, acetoacetate decarboxylates and produces carbon dioxide and acetone, as illustrated in Figure 4.10.4

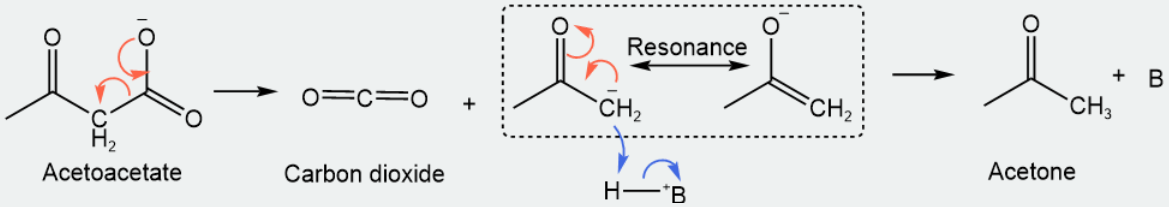


Figure 4.10.4: Illustration of decarboxylation mechanism with the example of acetoacetate converting into carbon dioxide and acetone. (Copyright; Public domain)

Carbon dioxide leaves under moderate conditions in this case because the anion left behind is in resonance with the β-C=O group. The body does not metabolize acetone but exhales through the lungs. Acetone is responsible for its characteristic sweet smell in the breath of people with swear diabetes.

CHAPTER OVERVIEW

5: Carbohydrates

5.1: What are carbohydrates

Learning Objectives

- Define carbohydrates, monosaccharides, polysaccharides, and write their general formulas.
- Draw and interpret fisher projections of monosaccharides.

What are carbohydrates

Carbohydrates are primary energy storage compounds, e.g., glucose ($C_6H_{12}O_6$) synthesized by using carbon dioxide (CO_2) from the air, water (H_2O) from the soil, and energy from sunlight, along with the release of oxygen (O_2), as illustrated in Figure 5.1.1.

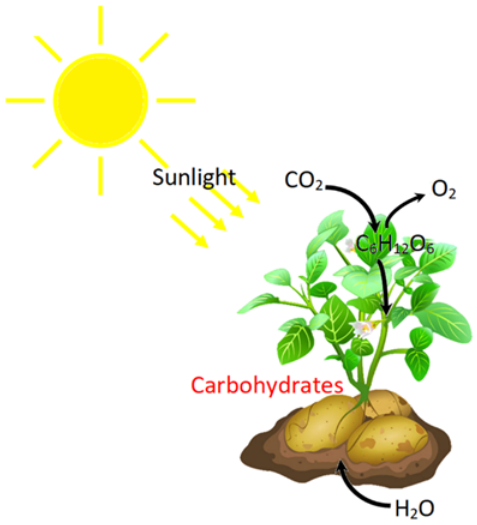


Figure 5.1.1: Illustration of photosynthesis by plants. (Copyright; Public domain)

Oxidation of carbohydrates, e.g., the reverse of the photosynthesis reaction, release energy that plant and animals use for activities. Carbohydrates are the support structure material in plants, e.g., cell-wall and wood, shell material of crustaceans, and connective tissues in animals.

Glucose is a polyhydroxy aldehyde, and fructose, i.e., another carbohydrate found in honey, is a polyhydroxy ketone, as illustrated in Figure 5.1.2.

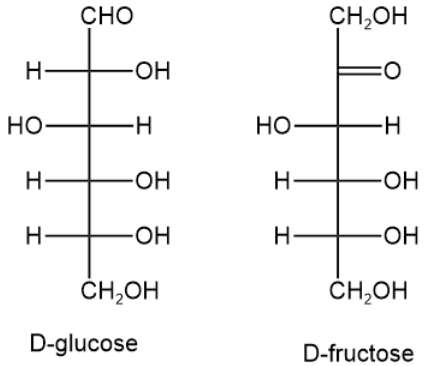


Figure 5.1.2: Fisher projections of D-glucose and D-fructose. (Copyright; Public domain)

Carbohydrates

Carbohydrates are polyhydroxy aldehydes, polyhydroxy ketones, or other compounds that hydrolyze to polyhydroxy aldehydes or polyhydroxy ketones.

The general formula of simple carbohydrates is $C_nH_{2\,n}O_n$, which can also be written as $C_n \cdot (H_2O)_n$ which is the origin of the name carbohydrates, i.e., hydrates of carbon.

Fisher projections

Fisher projections are two-dimensional representations of molecules for showing the configuration of chiral centers:

- the chiral center is in the plane of the page,
- horizontal lines represent bonds projecting toward the viewers,
- and vertical lines represent bonds projecting away from the viewer.
- The parent C chain is placed on the vertical line, with the most oxidized C, i.e., the C=O in carbohydrates, at the top or near the top end, and the numbering starts from the top most C, as illustrated in Figure 5.1.2.

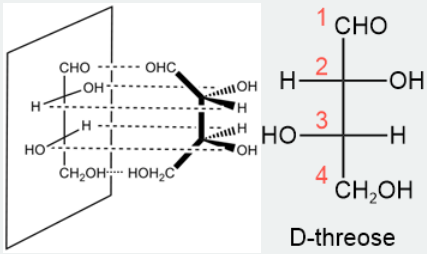


Figure 5.1.2: Illustration of creating a Fisher projection (a) and numbering the C chain of carbohydrates using D-threose as an example. (Copyright: (a) Olion17, Public domain, via Wikimedia Commons, (b) Public domain)

Monosaccharides and polysaccharides

Monosaccharides

Simple carbohydrates that can not be hydrolyzed to more simple ones are called **monosaccharides**. For example, D-glucose, D-fructose, and D-threose shown in previous figures are monosaccharides.

Polysaccharides

Linear or branched chain polymers comprised of monosaccharide repeat units (monomers) are called **polysaccharides**. For example, starch is a polysaccharide with a D-glucose monomer, illustrated in Figure 5.1.3.

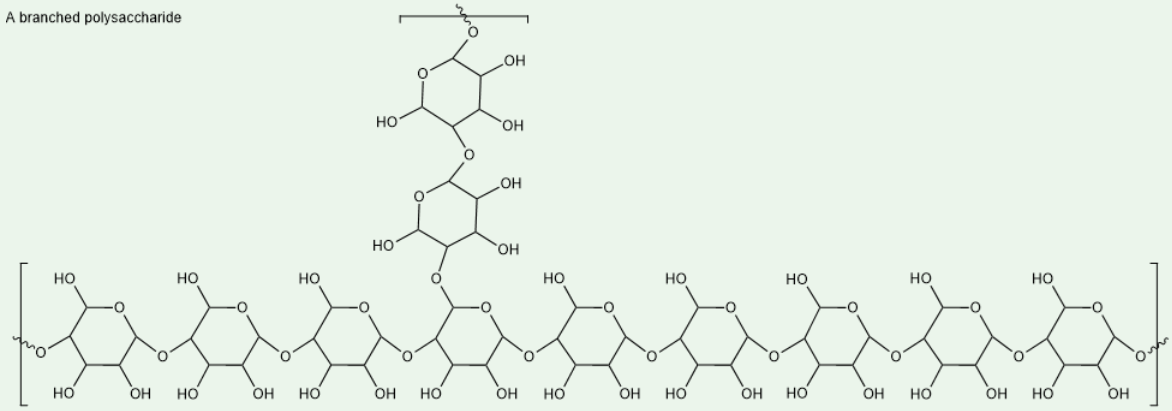
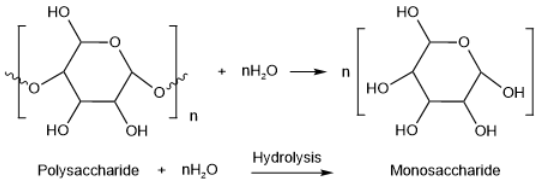


Figure 5.1.3: General structural formula of a starch -a polysaccharide D-glucose. (Copyright; Public domain)

Polysaccharides hydrolyze to monosaccharides, as illustrated in the following general hydrolysis reaction.



5.2: General class names and Common names of monosaccharides

Learning Objectives

- Assign and interpret class names of monosaccharides.
- Assign D/L stereodescriptors to the common names and define epimers.
- Draw structure from the name and vice versa for important monosaccharides, including D-ribose, D-glucose, D-mannose, and D-galactose.

General class names of monosaccharides

Monosaccharides are either polyhydroxy aldehydes that take *aldo* as prefixes or polyhydroxy ketones that take *keto* as prefixes in their general name. The general formula of monosaccharides is $C_nH_{2\,n}O_n$ where n can be 3, 4, 5, 6, 7, or 8 representing triose, tetrose, pentose, hexose, heptose, or octose, respectively, in the general name. For example, D-glucose belongs to aldohexose, where the aldo- prefix tells it is an aldehyde, -hex- in the middle of the name tells it has six C's, and -ose suffix denotes it is a carbohydrate. D-fructose belongs to ketohexose, i.e., it is a monosaccharide with a ketone group and six C's. D-glyceraldehyde is an aldotriose, i.e., a monosaccharide with an aldehyde group and three C's.

Common names of monosaccharides

Common names are specific for each monosaccharide. All C's in a monosaccharide are chiral centers except the two terminal C's and the C of ketone group if it is a ketose. The absolute configuration of the **penultimate** C, i.e., the second-last C, is explicitly expressed by D- or L-stereochemical descriptors and the absolute configuration of all other chiral centers is implicit in the common name of the monosaccharide. There is one set of common names for all D-isomers, and their mirror images (enantiomers) have the same common name with D- replaced with L-. For example, D-glyceraldehyde and L-glyceraldehyde, D-glucose, and L-glucose are enantiomer pairs shown in Figure 5.2.1.

D/L Stereochemical descriptors

If the $-OH$ group on the second-last carbon (penultimate C or the second C from the bottom end) in Fisher projection of a monosaccharide is on the right side, it is assigned D- and if it is on the left side, it is assigned L-configuration. Monosaccharides that are enantiomer pairs have the same common name, but D- is replaced with L- or vice versa. For example, D-glyceraldehyde and L-glyceraldehyde, D-glucose and L-glucose enantiomer pairs are shown in Figure 5.2.1. with the penultimate C defining D- or L-configuration shown in red color.

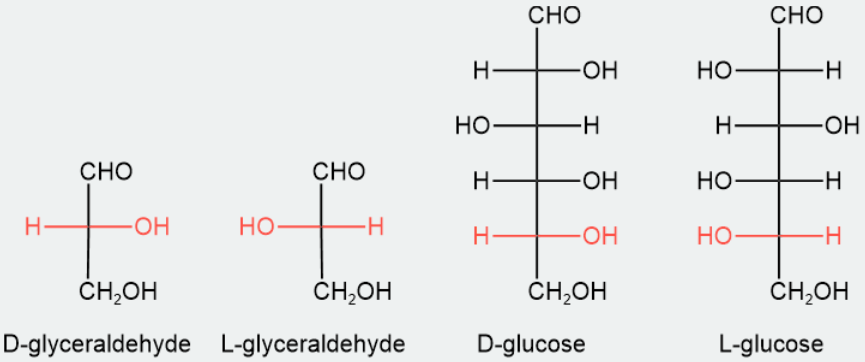


Figure 5.2.1: Assigning D- or L-configuration based on orientation of $(-OH$ on penultimate C in Fisher projection. (Copyright; Public domain)

D-configuration of monosaccharides is commonly found in nature. The D/L stereodescriptors do not indicate the rotation of the plane polarized light, i.e., the enantiomer's dextro/levo rotatory nature. However, if one enantiomer is dextro (d-), the other is levo (l-) to the same degree, and vice versa.

The structures as Fisher projections and common names of D-aldoses are shown in Figure 5.2.2. and those of D-ketoses are shown in Figure 5.2.3. These are the most common monosaccharides found in nature.

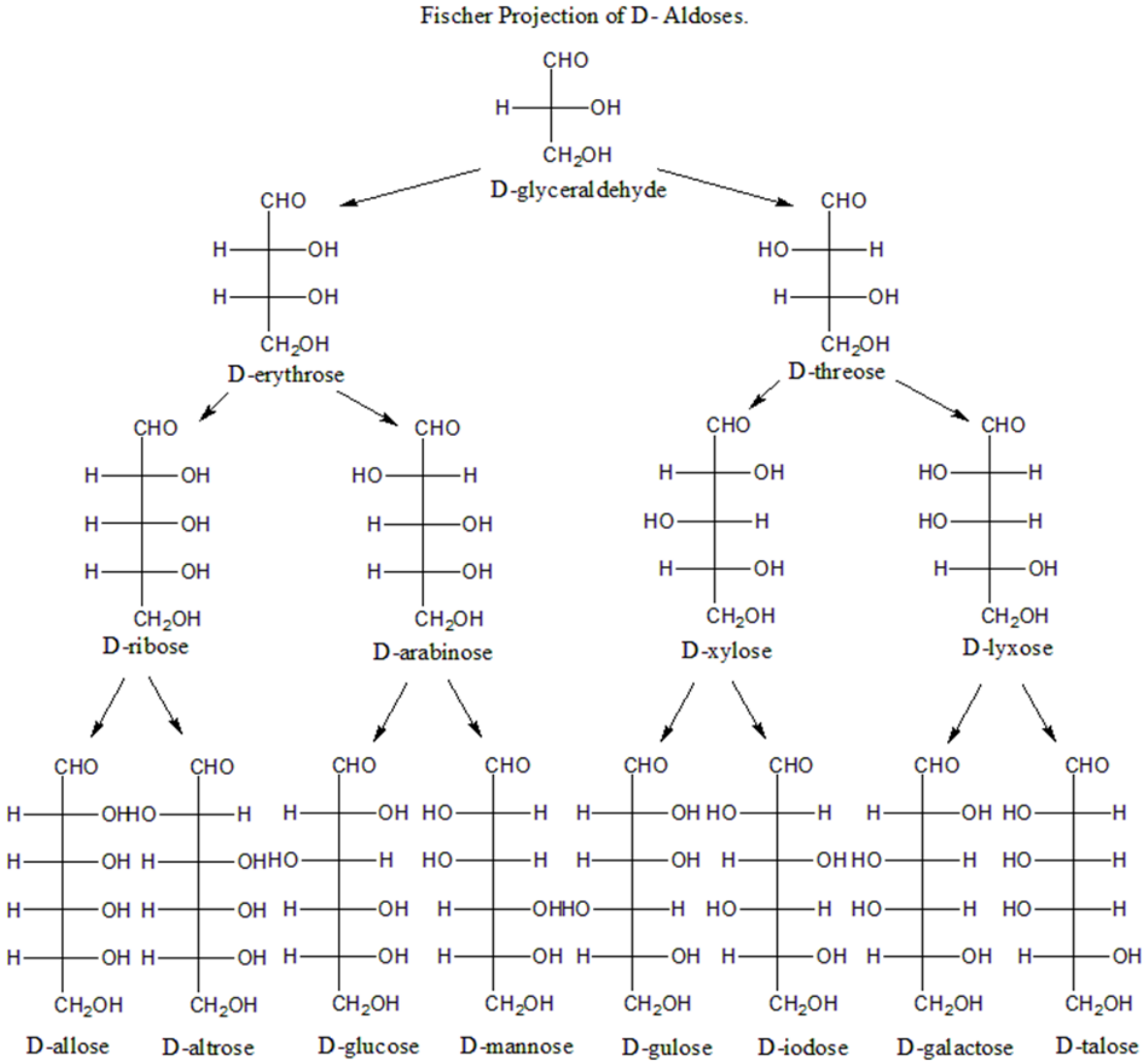


Figure 5.2.2: Fisher projections and common names of D-aldoses. (Copyright; Dineshts, Public domain, via Wikimedia Commons)

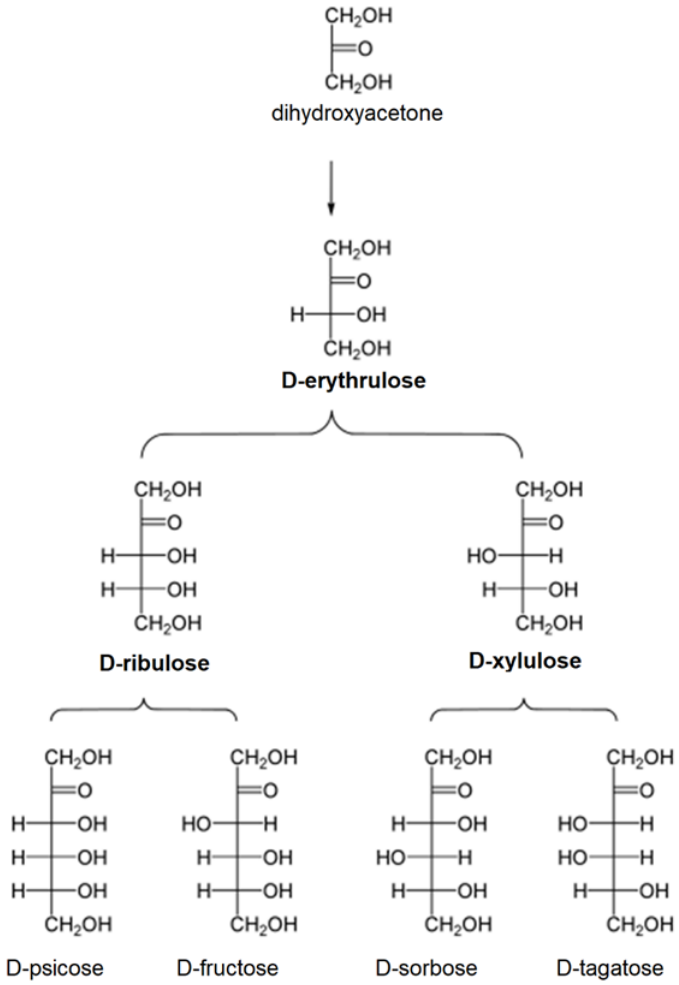


Figure 5.2.3: Fisher projections and common names of D-ketoses. (Copyright;modified from: Yikrazuul, Public domain, via Wikimedia Commons)

Some important monosaccharides

D-Glucose is the most abundant monosaccharide in nature. Plants produce it in a photosynthesis process. D-Galactose and D-mannose are two important diastereomers of D-glucose that differ from D-glucose in the configuration of only one chiral center.

Epimers

Epimers are diastereomers that differ in absolute configuration of only one chiral center. For example, D-galactose configuration is different from D-glucose only at C#4, i.e., D-galactose is a C4-epimer of D-glucose. Similarly, D-mannose is C2-epimer of D-glucose.

D-Fructose is another important monosaccharide that differs at C#1 and C#2 from glucose. That is, the C=O is an aldehyde group at C#1 in D-glucose, but it is a ketone at C#2 in D-fructose. Fisher projections of D-glucose, D-galactose, D-mannose, and D-fructose are shown in Figure 5.2.4. with the differences from D-glucose highlighted by red-color fonts. D-ribose is another important monosaccharide present in RNA.

Drawing structures of important monosaccharides

D-ribose is aldopentose, i.e., an aldehyde with five C′s. All chiral C′s have −OH groups oriented towards the right in the Fisher projection. D-allose is aldohexose with the same structural features as D-ribose, i.e., all chiral C′s have −OH groups oriented towards the right in the Fisher projection. D-glucose is C3 epimer of D-allose. The other three important monosaccharides can be drawn by relating them to D-glucose, i.e., D-galactose is a C4 epimer, D-mannose is a C2 epimer of D-glucose, and D-fructose has a ketone group at C2 in the place of aldehyde group of D-glucose at C1.

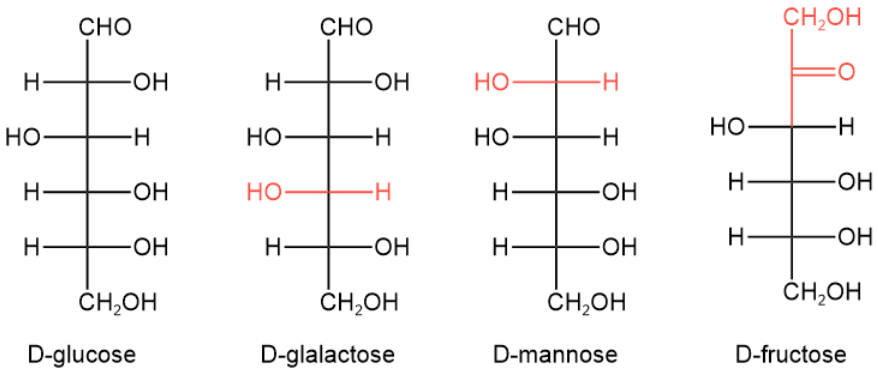


Figure 5.2.4: Fisher projections of four important monosaccharides with their differences from D-glucose highlighted by red-color fonts. (Copyright; Public domain)

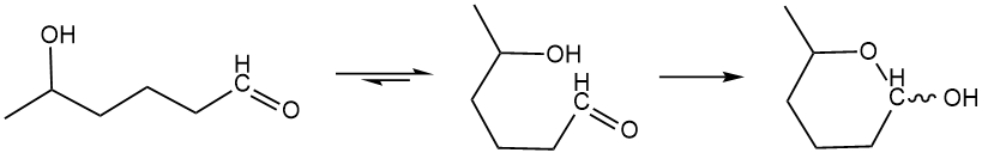
5.3: Cyclic structures of monosaccharides

Learning Objectives

- Draw and inter-convert open-chain and cyclic hemiacetal forms of compounds containing carbonyl and alcohol groups in the same molecule.
- Convert Fisher projections to Howarth projections and to chair conformation in the cases of pyranose forms of monosaccharides.
- Define Howarth projection, anomeric carbon, α and β configuration of the anomeric carbon, furanose, and pyranose forms, and mutarotation of monosaccharides.

Cyclic hemiacetal

Alcohol ($\mathrm{R-OH}$) and a carbonyl ($\mathrm{C=O}$) groups react with each other and form a hemiacetal group. The hemiacetals are usually unstable and revert to the reactants. However, suppose the alcohol and the carbonyl groups are on the same molecule and can react to form a five- or six-member cyclic hemiacetal. The product is more stable than the reactants, as shown in the following example reaction.



5-hydroxyhexanal (Linear conformation) | 5-hydroxyhexanal (Cyclic conformation) | A cyclic hemiacetal (Racemic mixture)

Note that the hemiacetal C is a new chiral center that is produced as a racemic mixture, which is represented by a wavy line in the $\mathrm{-OH}$ group. Monosaccharides are polyhydroxy aldehydes or polyhydroxy ketones that often exist as hemiacetal as the dominant form in equilibrium with the open-chain form, as described in the next section.

Fisher projections to Haworth projections

Fisher projections represent the open-chain forms of monosaccharides where the C-chain is a vertical line. Hemiacetals are five or six-member ring structures represented by pentagon or hexagon shapes.

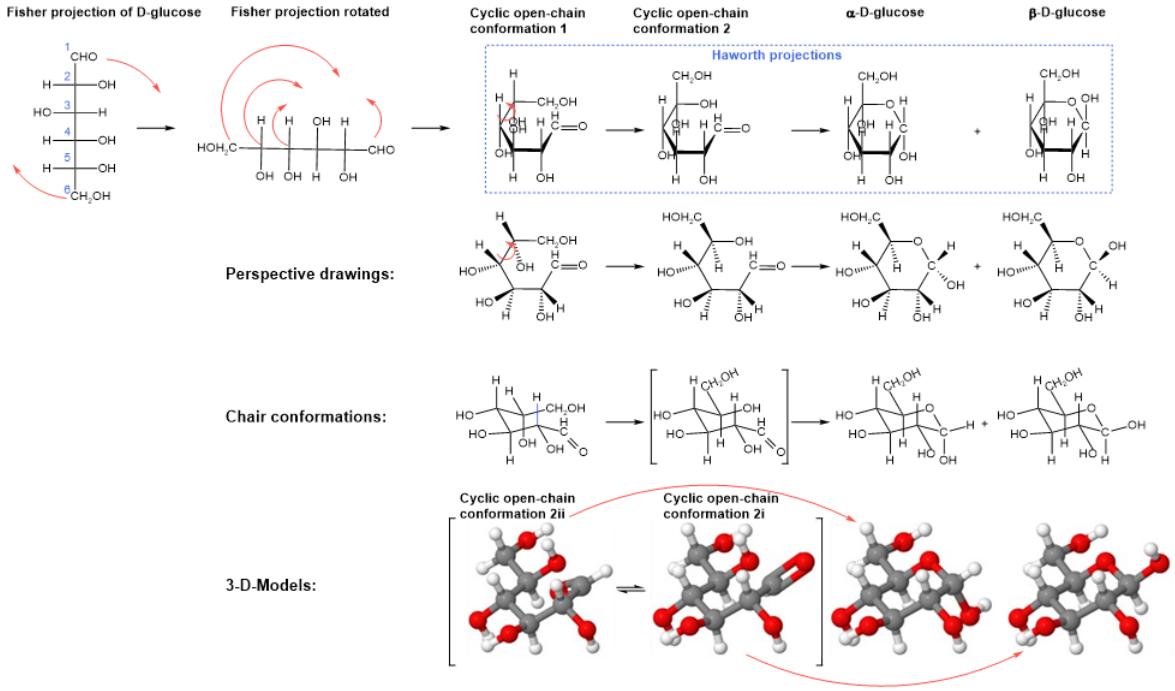


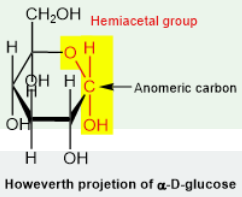
Figure 5.3.1: Fisher projections to Haworth projections explained. Follow the movement of atoms or rotations around single bonds, represented by curvy arrows in red color (Copyright; Public domain)

Figure 5.3.1 explains the steps to re-draw Fisher projection in cyclic conformation.

1. Rotate the Fisher projection by 90^o clockwise from vertical to horizontal orientation. The groups on the right of Fisher's projection end up above, and those to the left end below in the horizontal orientation.
2. Move C$'$s#1, 4, 5, and 6 up by 60^o each to arrive at the cyclic open-chain conformation#1.
3. Rotate the C4-C5 bond by 60^o counterclock wise to bring the C#5$-$OH group in the cycle of C$'$s and C#6 pointing above, i.e., cyclic conformation#1 in Figure 5.3.1.
4. C#5$-$OH and a carbonyl (C$=$O) of C#1 react to produce the hemiacetal group with the new $-$OH group on C#1 either pointing away from the hemiacetal O, i.e., α- or point towards the hemiacetal O, i.e., β-orientation in the cyclic hemiacetal.

Anomeric C

Anomeric C is the C of carbonyl (C$=$O) group in the open chain conformation that becomes the C of the hemiacetal group in the cyclic structure of a monosaccharide. For example, C#1 of D-glucose is the anomeric C , as highlighted in the figure on the right.



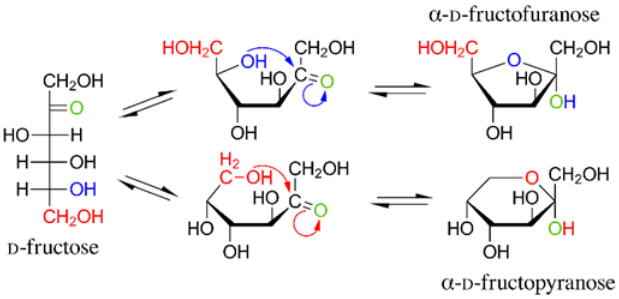
Haworth projection

Haworth projection is five or six-membered cyclic hemiacetals of monosaccharides, presented as pentagon or hexagon shape. It is viewed through its edge perpendicular to the plane of the page, such that the anomeric C s the right corner in the plane of the page and ether group (R$-$O$-$R) is the top right corner in the rear, as shown in the figure on the right. Bold lines usually show the ring bonds toward the viewer (above the page) and those below the page (away from the viewer) as solid lines. Bonds outside the ring are shown as solid lines pointing above or below it.

Assigning α- or β- to the cyclic structures of monosaccharides

Conversion of monosaccharides from open chain to cyclic hemiacetal produced a new chiral center at the anomeric C. It is a mixture of two conformations: the $-$OH group on anomeric C going away from the ether (R$-$O$-$R) group is α- and towards it is β-configuration, as illustrated in the figure on the right with the help of D-glucose example.

D-Fructose exists in five-membered and six-membered hemiacetal forms, as shown in Figure 5.3.2.



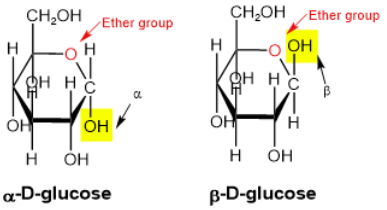
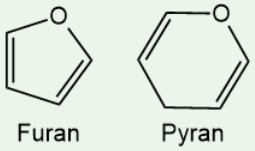


Figure 5.3.2: Equilibrium reaction between the acyclic and the cyclic hemiketal forms of D-fructose (note: hemiacetal group from a ketone is called hemiketal. α-forms of the hemiketal are shown and the corresponding β-forms are not shown in this figure). (Copyright; Vaccinationist, Public domain, via Wikimedia Commons)

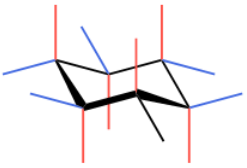
Furanose and pyranose forms of monosaccharides

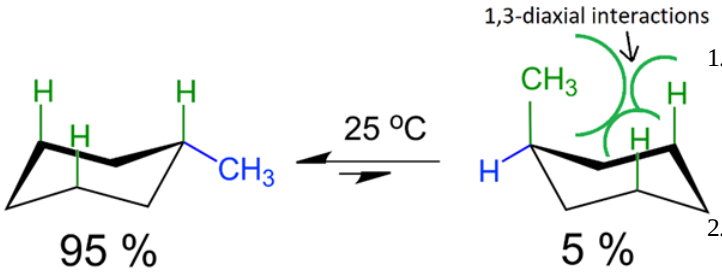
Furanose is a five-membered, and **pyranose** is a six-membered cyclic hemiacetal of monosaccharide. These names originate from the names of cyclic ethers, furan, and pyran, shown in the figure on the right. Based on this nomenclature, the class name of α-D-Glucose is α-D-glucopyranose and of β-D-glucose is β-D-glucopyranose. D-Fructose exists in pyranose and furanose forms, as shown in Figure 5.3.2.



Chair conformations of pyranose forms

Haworth projections assume a flat ring structure. It is close to the structure of furanose forms (five-membered rings), but pyranose forms (six-membered rings) exist in chair conformations. The figure on the right shows a chair conformation of a six-membered cyclic structure. The points to note are the following.





1. Each C has two bonds (shown in black) in the cycle, one axial bond (shown in red) pointing along the axis of the cycle and one equatorial bond (shown in blue) pointing approximately along the equator of the cycle.
2. The axial bonds point in opposite directions, up and down, on neighboring C′. The same applies to equatorial bonds.
3. Bulky group, i.e., any group other than −H, is more stable at the equatorial than axial position because of steric strain (push away) from the other two axial positions on the same face of the ring., as illustrated in the figure on the left.

How to draw chair conformations of monosaccharides

Alpha (α) and axial start with the letter *a*. Remember α- means an axial −OH at anomeric C in the case of pyranose chair conformation.

Since $\alpha-$OH is axial, $\beta-$OH is equatorial at anomeric C. The Howarth projection of β-D-glucose in the figure on the right shows that the bulky groups alternate between up and down positions on C#1 to C#5. It means all of them are either axial or equatorial. Since $\beta-$OH is equatorial, all bulky groups in β-D-glucose are equatorial, as shown in Figure 5.3.3.

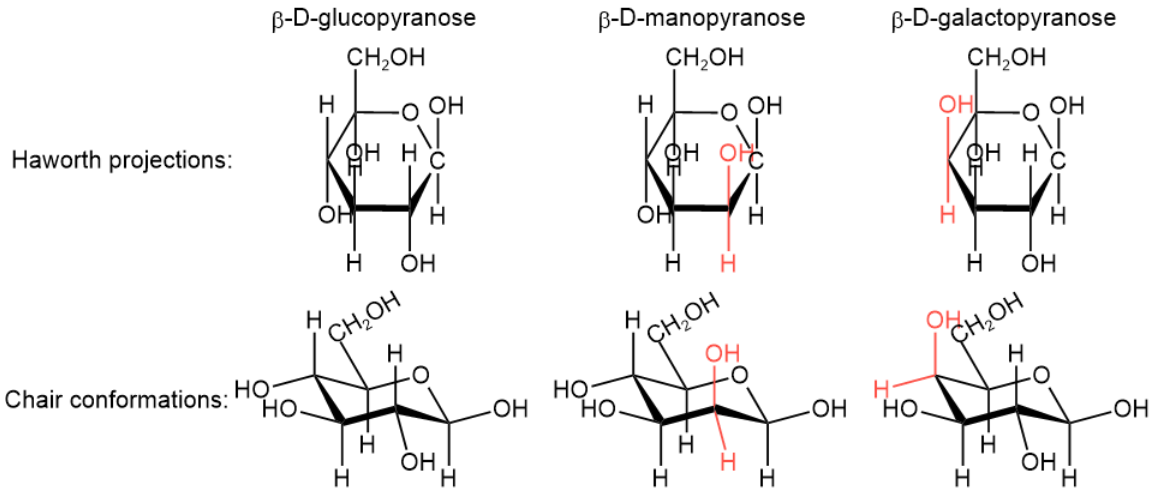


Figure 5.3.3: Haworth projections and the corresponding chair conformations of β-D-glucopyranose, β-D-manopyranose, and β-D-galactopyranose with the differences form β-D-glucopyranose shown in red color. The corresponding α forms (not shown) have $ce-OH$ on C#1 (right-most) on axial positions . (Copyright; Public domain)

After learning how to draw β-D-glucopyranose, the other monosaccharides can be drawn easily based on the knowledge of the difference from β-D-glucopyranose. For example, β-D-manopyranose is C#2 epimer, β-D-galactopyranose is C#4 epimer, and their α forms have an axial $-\mathrm{OH}$ at anomeric C, as shown in Figure 5.3.3.

Mutarotation

D-Glucose is a chiral compound. A freshly prepared aqueous solution of β-D-glucopyranose has a specific rotation of +112.2. A freshly prepared solution of α-D-glucopyranose has a specific rotation of +18.7. However, the specific rotation of β-D-glucopyranose and α-D-glucopyranose gradually changes to +52.5. This is because the open chain and the two cyclic forms establish equilibrium in the solution, as illustrated in Figure 5.3.4.

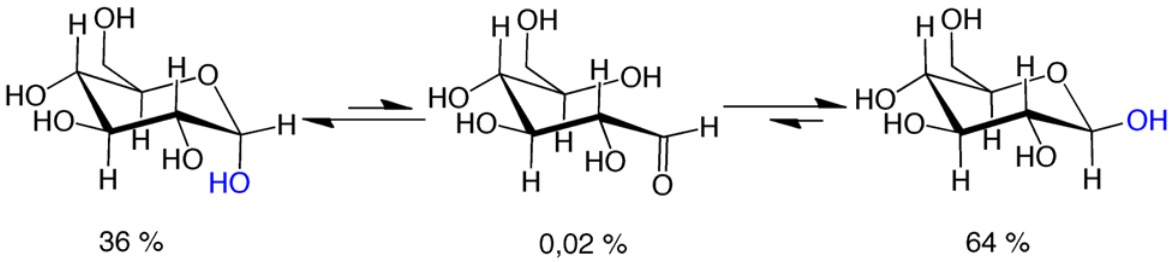


Figure 5.3.4: Equilibrium between three forms of D-glucose: open-chain D-glucose in the middle, β-D-glucopyranose on the right, and α-D-glucopyranose on the left side, with their % ratios shown below the structure. (Copyright; Jü, CC0, via Wikimedia Commons)

Open chain and the two cyclic forms of D-glucose exist in equilibrium in solution where β-D-glucose has all bulky group in equatorial position is the most stable, α-D-glucose with one bulky group in axial position is the second most stable and, and open chain for is the least stable, as reflected by their proportions at equilibrium shown in Figure 5.3.4.

> Mutarotation is the change in the optical rotation of carbohydrate solution over time due to the change in the composition of different compound forms to achieve equilibrium composition.

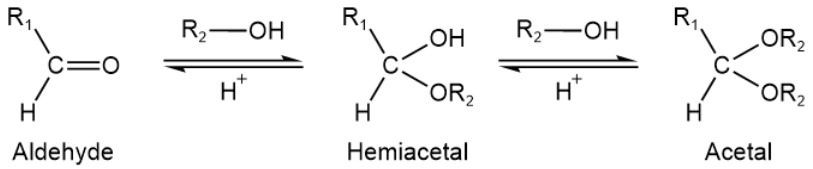
5.4: Reactions of monosaccharides

Learning Objectives

- Write and understand chemical reactions of monosaccharides, including conversion of cyclic hemiacetal to glycosides and N-glycosides, oxidation of open-chain aldehyde form to aldonic acids, and reduction to alditols.
- Define reducing sugars with the examples of D-glucose and D-fructose.

Glycosidic bond formation

Aldehydes and ketones react with alcohols to form hemiacetals. If the alcohol reagent is in excess, a second molecule of the alcohol reacts and converts hemiacetal to acetal. Acetals can be isolated. Acetals convert to alcohol and the aldehyde or ketone when their aqueous solution is acidified, as shown in the general reaction below.



Cyclic (pyranose and furanose) forms of monosaccharides are hemiacetals that react with alcohols to form acetals.

Glycosides are the acetals of monosaccharides. The glucose formed by a reaction with alcohol is also called **O-glycoside**.

For example, β-D-glucopyranose reacts with methanol to form glycosides, shown in Figure 5.4.1.

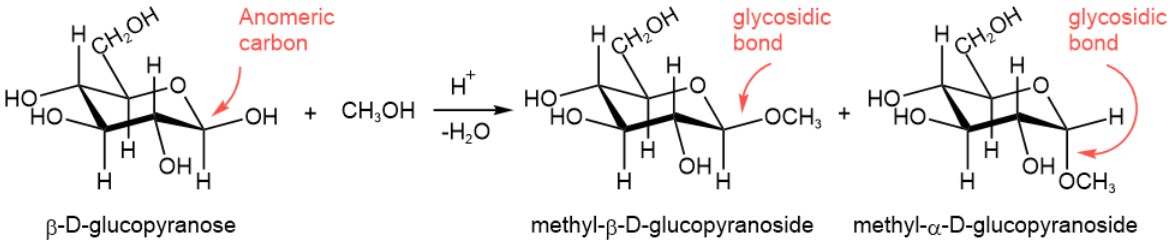


Figure 5.4.1: Conversion of β-D-glucopyranose to its methyl glycosides. Mixture of α and β glycosides is produced, irrespective of whether the reactant is α or β form of the carbohydrate.(Copyright; Public domain)

Glycoside and glycosidic bond

An acetal of a monosaccharide is a glycoside. The bond between anomeric C to the new ether $-$OR group in the glycoside is called the glycosidic bond. Anomeric C and glycosidic bonds are indicated by red arrows in Figure 5.4.1.

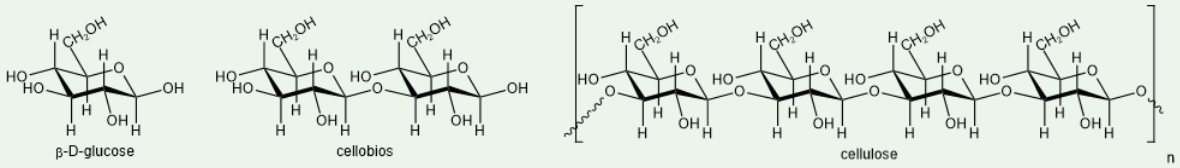
Glycoside is named by replacing the last e of the carbohydrate name with ide and adding the name of the alkyl part of the new ether $-$OR group as a prefix separated by a hyphen. For example, β-D-glucopyranose reacts with methanol to form methyl-β-D-glucopyranoside and methyl-α-D-glucopyranoside, as shown in Figure 5.4.1.

Since monosaccharides have many alcohol groups, two monosaccharides can react with each to form a glycoside, called a **disaccharide**, three can form a **trisaccharide**, and many can form a **polysaccharide**. The di-, tri-, and polysaccharides can be hydrolyzed to monosaccharides in an acidic aqueous solution.

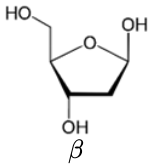
Monosaccharide, disaccharide, oligosaccharide, and polysaccharide

A carbohydrate that can not be hydrolyzed to a simpler carbohydrate is a **monosaccharide**, e.g., D-glucose. A glycoside of two monosaccharides is a **disaccharide**, e.g., cellobiose. A glycoside of three to ten monosaccharides is an **oligosaccharide**, e.g., a

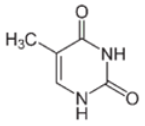
fragment of cellulose. A glycoside of more than ten monosaccharides is a **polysaccharide**, e.g., cellulose. The figure below illustrates the structures of D-glucose -a monosaccharide; cellobiose -a disaccharide; and cellulose, -a polysaccharide.



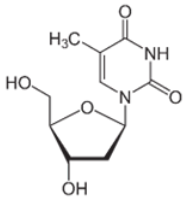
Amines react with hemiacetals forms of monosaccharides the same way as alcohols and produce N-glycosides. For example, β-D-2-deoxyribose reacts with thymine and produced an N-glycoside called deoxythymidine, a part of a monomer of DNA molecule, as shown in the figure below.



-D-2-Deoxyribose (a monosaccharide)



Thymine (an amine)



Doxythymidine (an N-glycoside)

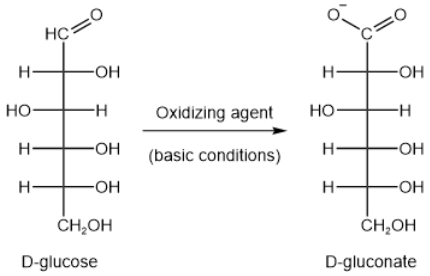
Exceptions to the general formula of carbohydrates

The general formula of carbohydrates is $C_m(H_2O)_n$, and usually for monosaccharides m is equal to n, i.e., $C_n(H_2O)_n$ or $C_nH_{2n}O_n$. However, there are exceptions.

- Carbohydrates do not necessarily conform to this formula. For example, β-D-2-Deoxyribose $C_5H_{10}O_4$ has less $O's$ than what is dictated by the general formula.
- Compounds following this general formula are not necessarily always carbohydrates. For example, acetic acid $C_2H_4O_2$ follows the general formula, but it is not a carbohydrate.

Oxidation to aldonic acids

Aldehydes are oxidized by a variety of reagents to carboxylic acids. Similarly, the aldehyde group ($-CHO$) of the open-chain form of aldoses is oxidized to a carboxylic acid ($-COOH$) group by oxygen O_2 in the presence of enzymes called oxidases. The acid is named by replacing the suffix -ose of the class name aldose with --onic acid. For example, D-glucose is oxidized to D-gluconic acid. Since, $-COOH$ exists in ionized ($-COO^-$) form, the suffix -onic acid is replaced with -onate for the ionized form. For example, D-glucose is oxidized to D-gluconate, as shown in the figure below.



Reducing sugars

A carbohydrate that reacts with a mild oxidizing agent under alkaline conditions to form an aldonic acid (or aldonate anion in physiological conditions) is called a reducing sugar. For example, D-glucose is a reducing sugar, as shown in the figure above. The carbohydrate reduces the oxidizing agent.

Common reagents used to test the presence of a reducing sugar include Tollen's reagent in which Ag^+ is reduced to Ag that forms a silver mirror on the glass, and Benedict's reagent in which Cu^{2+} is reduced to Cu^+ that forms red color precipitate

Cu_2O, as shown in Figure 5.4.2.

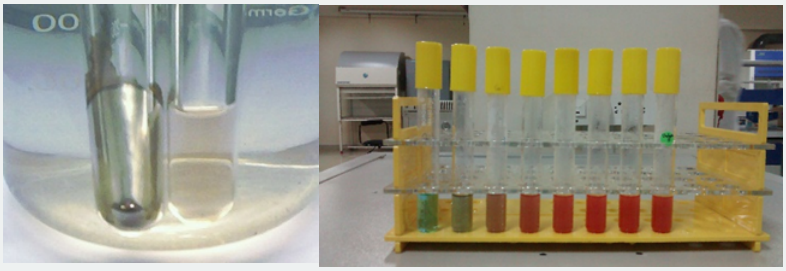
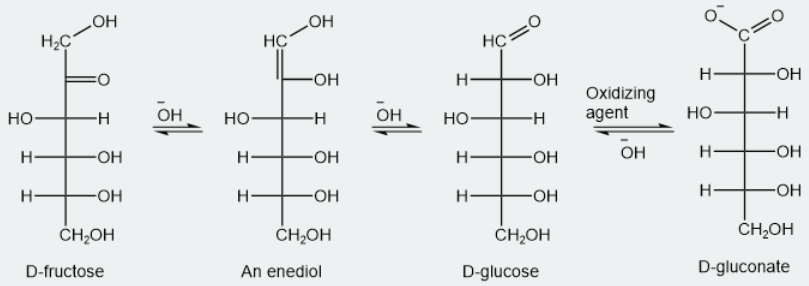


Figure 5.4.2: Tests for reducing sugars: silver mirror test forming silver mirror as a positive test (left) and Benedict's test forming red color precipitate as a positive test (right). (Copyright; silver mirror test: K1954, Public domain, via Wikimedia Commons, Benedict's test: Psgs123xyz at English Wikipedia, CC0, via Wikimedia Commons)

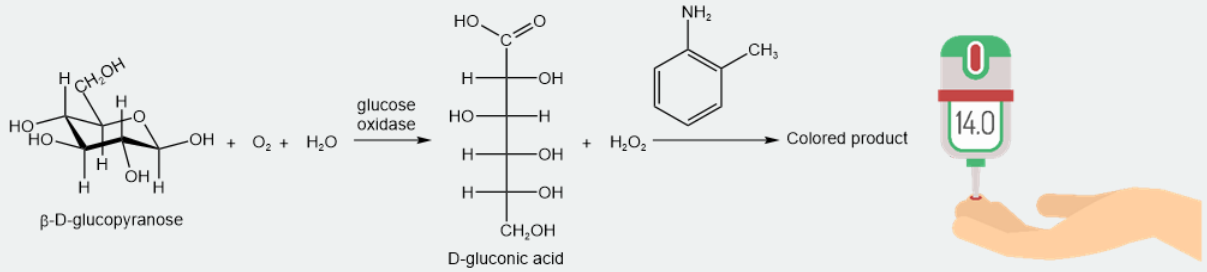
Although open-chain form is in small concentration at equilibrium, it is continuously formed from the hemiacetal forms as the oxidation reaction removes it. The 2-ketoses are also reducing sugars. This is because 2-ketoses exist in equilibrium with the corresponding aldoses under the basic conditions. The aldose form is the actual reducing agent in this mixture. For example, D-fructose is a reducing sugar. D-fructose exists in equilibrium with D-glucose in basic conditions, where D-glucose is the reducing agent, as illustrated in the figure below.



Diabetes and blood-sugar test

D-Glucose is also called blood sugar, as it is normally present in blood at about 70 mg/dL to 130 mg/dL. It may rise up to 140 mg/dL after eating food, but it returns to the normal range in healthy persons. Enzyme insulin controls blood sugar. If blood glucose control is not functioning properly, the blood glucose may stay higher than normal -a condition called hyperglycemia or diabetes; or it may stay below the normal level -a condition called hypoglycemia.

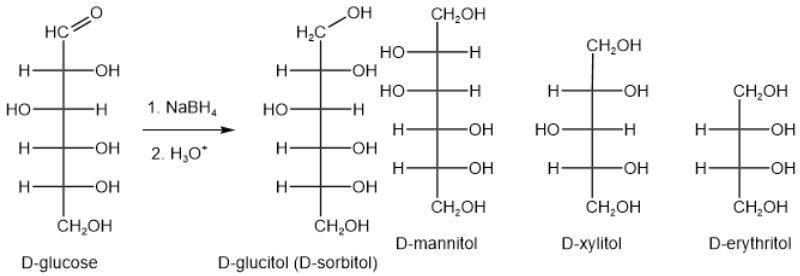
Blood glucose is usually tested based on enzyme-catalyzed reactions. Glucose oxidase enzyme oxidizes β-D-glucose using O_2 to D-gluconate and hydrogen peroxide H_2O_2. α-D-glucose does not react directly, but it converts to β-D-glucose as the latter is consumed. A second enzyme, peroxidase, causes H_2O_2 to react with 2-methylaniline and produces a colored product that is monitored to measure blood glucose, as illustrated in the figure below.



Reduction to alditols

Aldehydes are reduced to alcohols by a variety of reducing agents. Aldoses in the open-chain form are reduced by reducing agents like sodium borohydride ($NaBH_4$), H_2 in the presence of Pt, Pd, or Ni. Oxidases enzymes are reducing agents in biochemical systems. Although open-chain aldehyde form is in small concentration in the equilibrium mixture, cyclic hemiacetal forms convert

to open-chain form as the latter is consumed. The product alcohol is named by replacing -ose suffix of the aldose name with -itol. For example, D-glucose is reduced to D-glucitol, as shown below.



D-glucitol is found in berries, cherries, plums, pears, seaweed, and algae. It is commonly known as D-sorbitol and is a sugar substitute for diabetes. Other alditol examples include D-erythritol, D-mannitol, and D-xylitol. D-xylitol is used in cereals, sugarless candies, and gums.

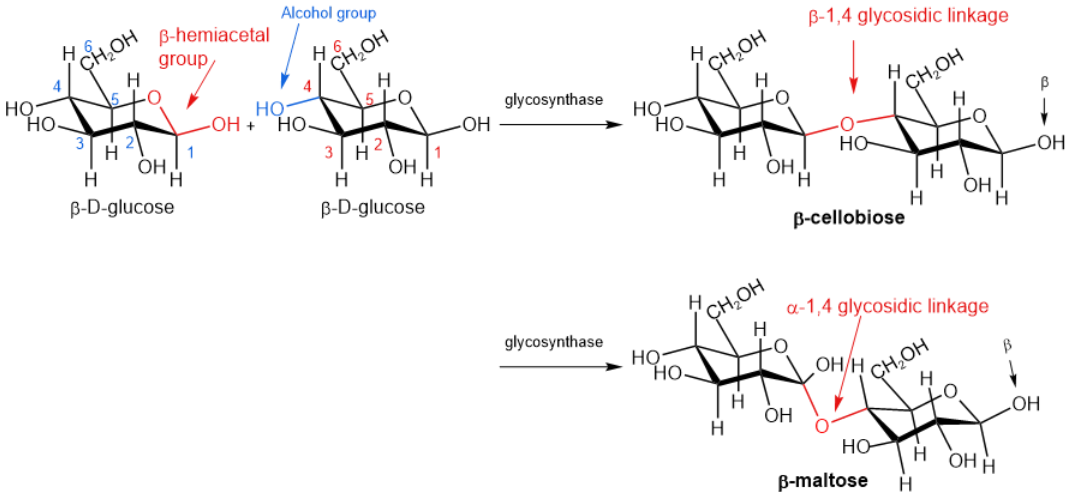
5.5: Disaccharides

Learning Objectives

- Be able to draw and name the glycosidic linkage in disaccharides.
- Know common disaccharides' structures, sources, and properties, including maltose, cellobiose, lactose, and sucrose.

Nomenclature of glycosidic linkage

Disaccharides are glycosides of two monosaccharides. Glycosidic linkage is formed by a reaction between the hemiacetal group of one monosaccharide and an alcohol group of the other, as illustrated in the figure below.

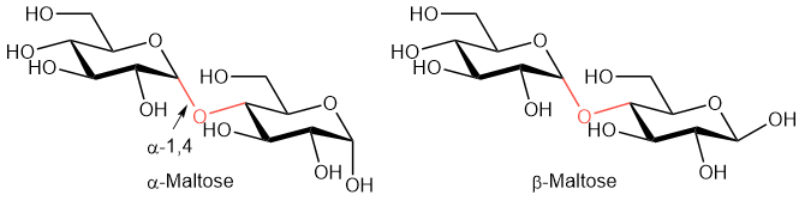


The glycosidic linkage is named by listing the number of the C′s linked, separated by a comma. In the case of anomeric carbon, α/β orientation is also indicated, separated by a hyphen from the number of the anomeric carbon. For example, β-cellobiosee is linked by β-1,4 glycosidic linkage and β-maltose by α-1,4 glycosidic linkage. The β in the name β-cellobiose and β-maltose indicates the orientation of the free hemiacetal group in the second saccharide unit. It is not related to the orientation of the glycosidic linkage, as indicated by the black arrow in the figure. β-Cellobiose and β-maltose are configurational isomers having different physical and chemical properties.

Examples of disaccharides include maltose, cellobiose, lactose, and sucrose, described next.

Maltose

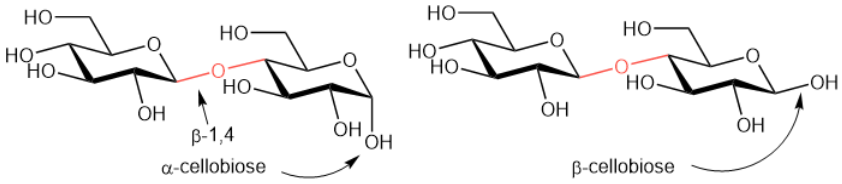
Maltose is a disaccharide of two D-glucose units connected through α-1,4-glycosidic linkage, as shown below.



The second D-glucose unit with free hemiacetal group can be either in α- or β-configuration, as indicated by α/β prefix to the name. An aqueous solution exists in equilibrium between α, β, and open-chain aldose forms. Since common oxidizing agents can reduce the open-chain aldose form, maltose is a reducing sugar. Maltose is found in the juice of sprouted barley grains and other grains.

Cellobiose

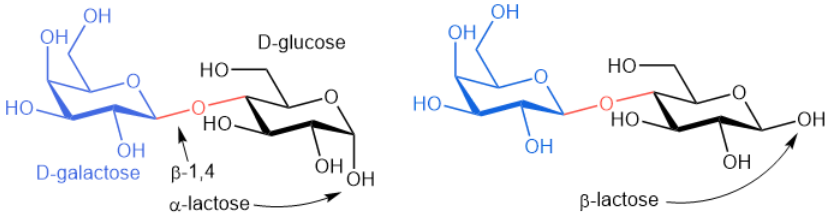
Cellobiose is a disaccharide of two D-glucose units connected by β-1,4-glycosidic linkage, as shown below. Hemiacetal forms of cellobiose are shown below.



Cellulobiose is a reducing sugar because, like maltose, the D-glucose unit with free hemiacetal group can exist in α, β, or open-chain aldose forms in solution. Cellobiose is obtained by hydrolysis of cellulose.

Lactose

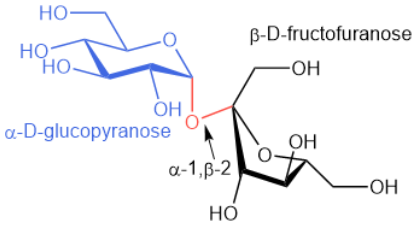
Lactose is a disaccharide of D-galactose joined by β-1,4-glycosidic linkage with D-glucose, as shown below.



Lactose is a reducing sugar because, like maltose, the D-glucose unit with free hemiacetal group can exist in α, β, or open-chain aldose forms in solution. Lactose is present in milk, up to 5% to 8% in human milk and 4% to 6% in cow milk. Lactase is the enzyme that hydrolysis lactose in the digestion process. Many adults develop lactose intolerance due to a lack of or insufficient lactase production. Without lactase, lactose enters the colon undigested and is fermented by bacteria causing bloating and abdominal cramps. Therefore, some products that use lactose also add lactase to avoid the problem.

Sucrose

Sucrose is a disaccharide of D-glucose (in pyranose form) and D-fructose (in furanose form) joined by α-1,\(beta)-2-glycosidic linkage, as shown below.



Sucrose

Both anomeric C, s in the glycosidic linkage can not exist in open-chain aldose or ketose forms. Therefore, sucrose is a non-reducing sugar. Sucrose is commonly used as table sugar. It is the most abundant disaccharide found in sugar cane and sugar beats.

Relative sweetness

Monosaccharides and disaccharides are sweet, but sweetness varies. Fructose is the sweetest monosaccharide, and lactose is the least sweet, as shown in Table 1. Honey is a mixture of fructose and glucose and has a sweetness about the same as table sugar (sucrose). Relative sweetness in the table is based on the results of a group of people tasting and ranking them in order of taste relative to sucrose (table sugar), rated 1. Some artificial sweeteners listed in the table are commonly used as sugar-substitute in diet foods.

Table 1: Relative sweetness of some carbohydrates and artificial sweeteners (Based on the results of a group of people tasting and ranking them in order of taste relative to sucrose (table sugar) rated 1 as reference).

Carbohydrates	Sweetness relative to sucrose	Sugar alcohols	Sweetness relative to sucrose	Artificial sweeteners	Sweetness relative to sucrose
Fructose	1.74	Xylitol	1.00	Advantame	2,000

Carbohydrates	Sweetness relative to sucrose	Sugar alcohols	Sweetness relative to sucrose	Artificial sweeteners	Sweetness relative to sucrose
Sucrose (table sugar)	1.00	Maltitol	0.80	Saccharin	450
Honey	0.97	Sorbitol	0.60	Acesulfame-K	200
Glucose	0.74			Aspartame	180
Maltose	0.33			Stevia	150
Galactose	0.65			Sucralose	60
Lactose	0.16				

5.6: Oligosaccharides

Learning Objectives

- Differentiate between blood types based on the oligosaccharides attached to the red blood cell surface and blood compatibility related to the blood type.

Oligosaccharides are polymers containing a small number, typically three to ten of monosaccharides. Oligosaccharides include dextrins which are microbial breakdown products of long polysaccharides like starch or cellulose and other glycans linked to lipids or proteins through N- or O-glycosidic bonds. The primary role of the last class is cell recognition and cell adhesion. For example, oligosaccharides attached to red-blood cells define the blood type A, B, AB, or O, depending on the linked oligosaccharides, as described next.

A, B, AB, or O blood types

The membrane of animal plasma cells typically has oligosaccharides comprising 4 to 17 monosaccharides attached to them. Their purpose is cell recognition, i.e., if there is a foreign cell or an object in the biochemical system, it is identified through the surface oligosaccharides and either destroyed or flagged for destruction by the body's immune system. For example, red blood cells have different types based on the oligosaccharides attached to their cell membrane, as shown in Figure 5.6.1.

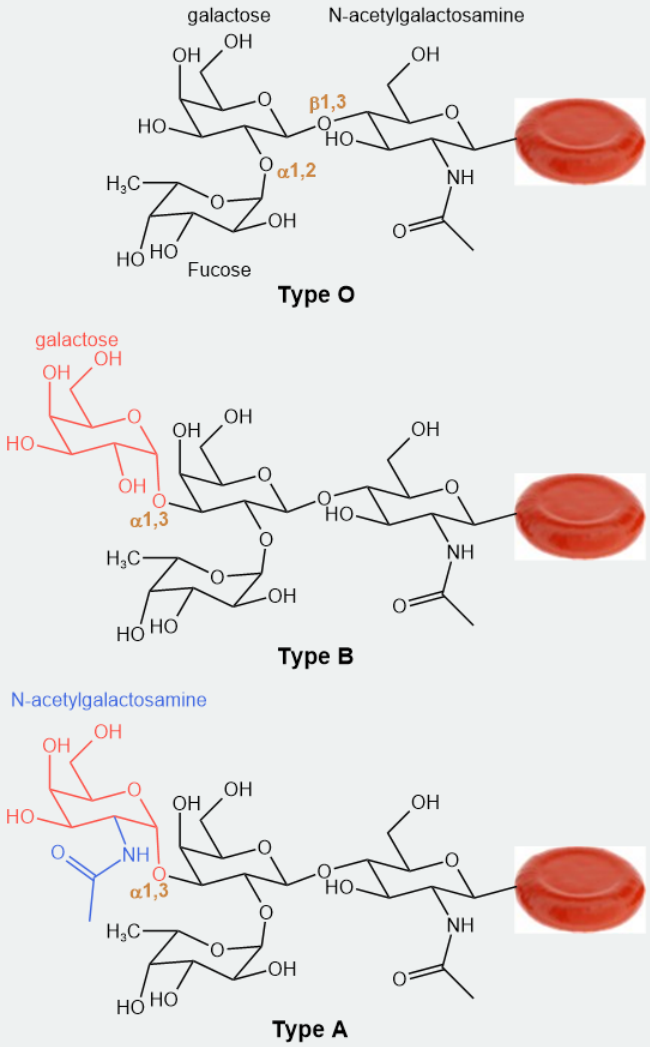


Figure 5.6.1: Oligosaccharide structures of blood type O, type B, and type A. Type AB has type A + type B in it. (Copyright; Public domain)

Blood type O has a combination of N-acetylgalactoseamine-galactose-Fucose. This combination is present in all other blood types also. So, Blood type O is not recognized as foreign in any blood type. A person with blood type O can donate blood to every type, i.e., type O is a universal donor. Any class other than type O has some additional monosaccharide attached which is recognized as foreign by blood type O. So, a person with blood type O can accept only type O blood.

Blood type B has an additional monosaccharide, i.e., galactose attached to galactose of type O oligosaccharide. Type B can receive blood from type O or type B and donate to type B or type AB, as shown in Figure 5.6.2.

Blood type A has an additional monosaccharide, N-acetyl galactosamine, attached to galactose of type O oligosaccharide. Type A can receive blood from type O or type A and can donate to type A or type A.B

Blood type AB has a mix of type A and type B. Type AB can accept all types, i.e., it is a universal recipient but can donate only to type AB.

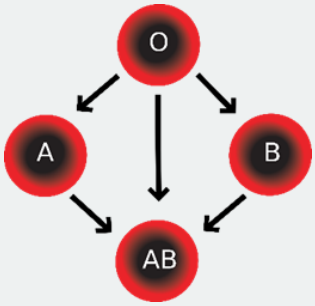


Figure 5.6.2: Diagram showing the blood group compatibility for transfusion purposes. (Copyright; InvictaHOG, Public domain, via Wikimedia Commons)

5.7: Polysaccharides

Learning Objectives

- Learn the structures and some characteristics of polysaccharides, including starches, cellulose, and chitin.

Polysaccharides are long polymers from ten to thousands of monosaccharides joined by glycosidic linkages. The most abundant polysaccharides are starch, glycogen, cellulose, and chitin. Except for chitin, all others are composed of D-glucose. Chitin is composed of a modified form of glucose. These are described next.

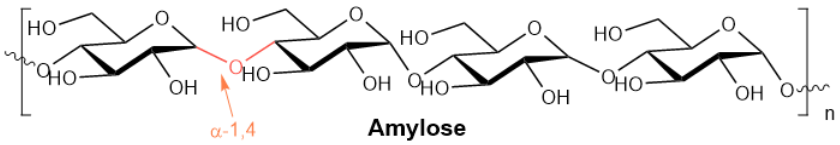
Starch

Starch is a storage form of D-glucose in plants. It is found in potatoes, beans, rice, wheat, and other grains and roots, as illustrated in Figure 5.7.1. Starch is a mixture of two forms, 20% to 25% amylose and 75 to 80% amylopectin.

Figure 5.7.1: Starch food: Sindhi Biryani is a delicious Dish of Sindh. (Copyright; Miansari66, Public domain, via Wikimedia Commons)

Amylose

Amylose is an unbranched chain of up to 4,000 D-glucose units joined by α-1,4-glycosidic linkage, as shown below.



Amylose

The orientation of bonds with α-1,4-glycosidic linkage makes the amylose chain exist in a helical structure, as shown in Figure 5.7.2, that allows water molecules to access and establish hydrogen bonds with $-\mathrm{OH}$ groups. Starch is a water-soluble polymer. It can dissolve in water due to extensive hydrogen bonding with the water molecules when mixed with water.

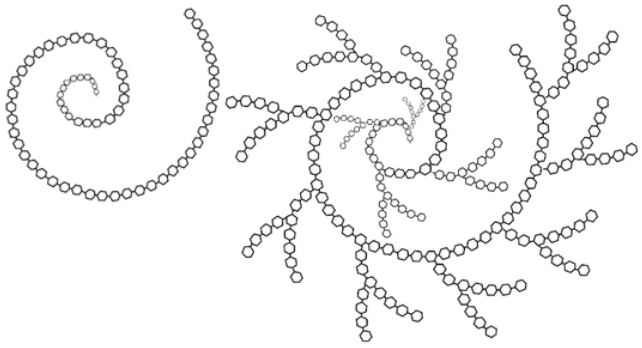
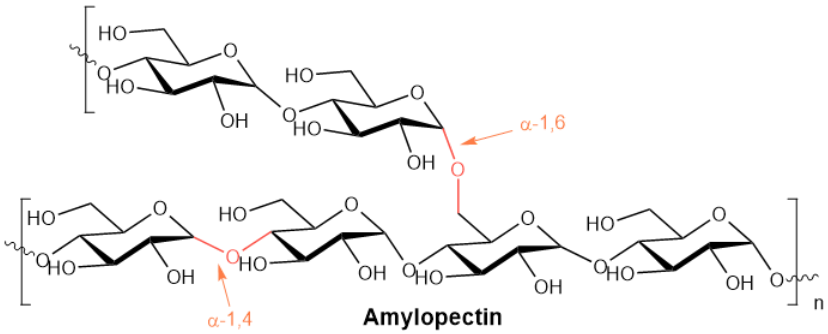


Figure 5.7.2: Illustration of a spiral structure of amylose (left) and branched helical structure of amylopectin (right). (Copyright; modified the free download from: https://www.hiclipart.com/free-trans...lumuh/download)

Amylopectin

Amylopectin is a branched chain polymer of D-glucose, as shown below.



The main chain comprises 10,000 D-glucose units joined by α-1,4-glycosidic linkage. The main chain exists in the form of a helix. Branches originate at every 24 to 30 unit intervals on the main chain connected to the main chain by a (\alpha\)-1,6-glycosidic linkage, as shown in Figure 5.7.2. This helix with branches going outwards makes amylopectin more accessible to water for hydrogen bonding and easier to digest.

Glycogen

Glycogen is an energy-storage polysaccharide in animals with the same structure as amylopectin. it has up to 10^6 D-glucose units joined by α-1,4-glycosidic linkages and branching through (\alpha\)-1,6-glycosidic linkages. The main difference from amylopectin is that glycogen has more frequent branching at 10 to 15 D-glucose units interval, as illustrated in Figure 5.7.3. More branching makes it more soluble in water and easier to hydrolyze multiple D-glucose units when the body needs D-glucose.

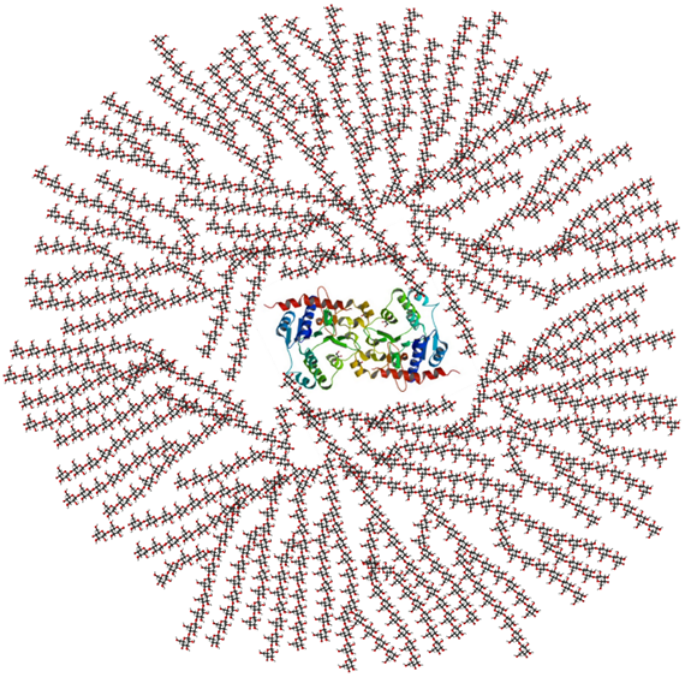


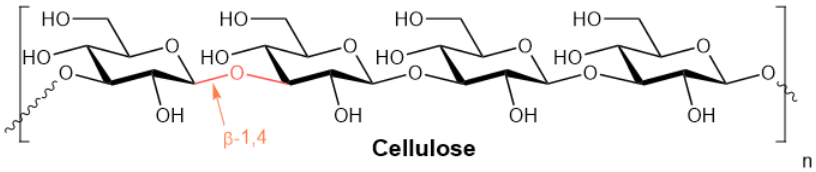
Figure 5.7.3: A 2-D cross-sectional view of glycogen -A core glycogenin protein surrounded by branches of glucose units. (Copyright; Mikael Häggström. Public domain, via Wikimedia Commons)

Cellulose

Cellulose is the most abundant polysaccharide in nature that makes up about 50% of the cell wall of plant cells. Cotton, shown in Figure 5.7.4, is almost pure cellulose.

Figure 5.7.4: Cellulse: Cotton is almost pure cellulose. (Copyright; United States Department of Agriculture, Public domain via Wikimedia Common)

Cellulose is a linear polysaccharide of about 2200 D-glucose units joined by β-1,4-glycosidic linkage, as shown below.



The β-1,4-glycosidic linkage allowes glucose units to adopt a linear structure with intra-molecule hydrogen bonding. The linear polymer packs nicely with inter-molecular hydrogen bonding and the London-dispersion forces, as shown in Figure 5.7.5. It gives cellulose its mechanical strength and makes it insoluble in water. Due to the extensive inter- and intra-molecular hydrogen bonding within cellulose, water can not establish enough hydrogen-bonds to dissolve it.

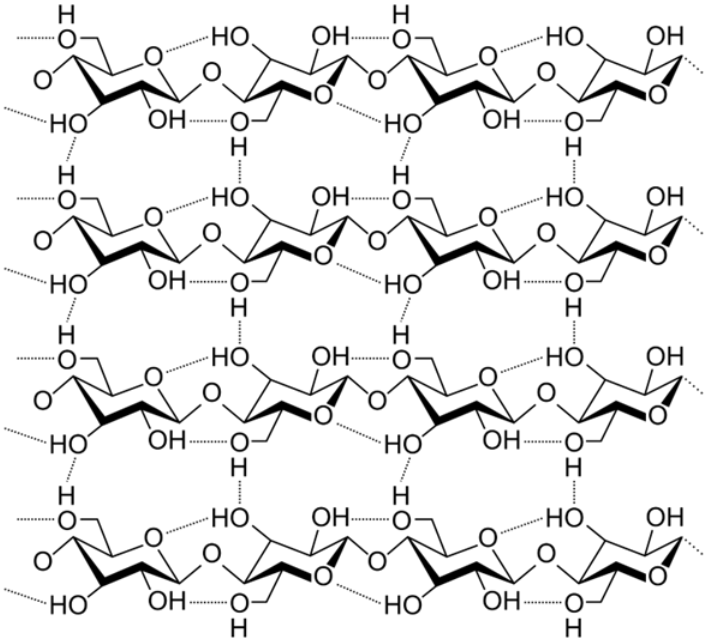


Figure 5.7.5: Illustration of linear chains of cellulose polymer with intra- and inter-molecular hydrogen bonding shown by dotted lines. (Copyright; Laghi.l, CC BY-SA 3.0, via Wikimedia Commons)

Animals have α-glucosidase enzymes that allow them to hydrolyze starch and glycogen to D-glucose, but they do not have β-glucosidase enzymes needed to hydrolyze cellulose. Grazing animals and termite hose bacteria in their stomach that have the β-glucosidase enzymes. Therefore, grazing animals and termites can digest cellulose.

Chitin

Chitin is the second most abundant polysaccharide, second only to cellulose. It is found in the exoskeleton of crustaceans and insects, as shown in Figure 5.7.6.

Figure 5.7.6: Blue crab, an example of chitin in the exoskeleton of crustaceans and insets. (Copyright; Blue Crab. USEPA Environmental-Protection-Agency, Public domain, via Wikimedia Commons)

Chitin has the same structure as cellulose except that it is composed of N-acetylglucosamine, an amide derivative of D-glucose, in place of D-glucose in cellulose, as illustrated below Figure 5.7.7.

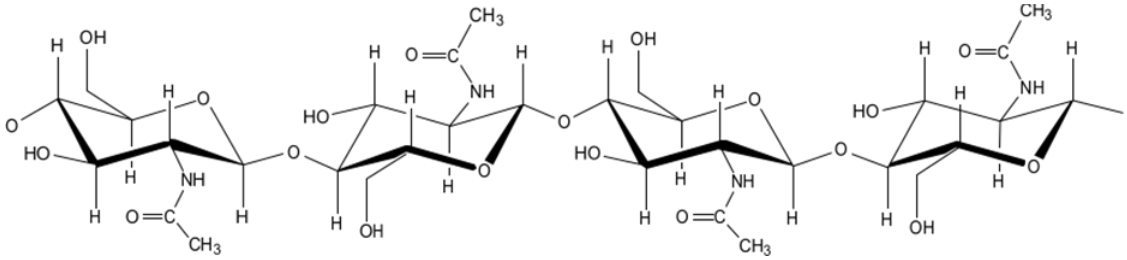


Figure 5.7.7: Structure of chitin. (Copyright; Ronny Flores, Public domain, via Wikimedia Commons)

CHAPTER OVERVIEW

6: Lipids

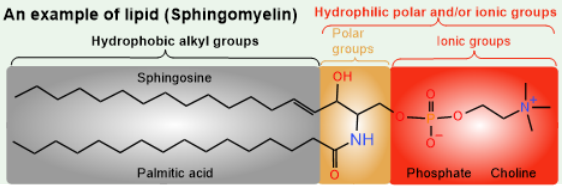
6.1: What are lipids?

Learning Objectives

- Define lipids and understand the basic characteristics of their structural components.

Lipids

Lipids can be defined as hydrophobic or amphiphilic biochemicals, where **hydrophobic** means water-hating or non-polar and **amphiphilic** means having both **hydrophilic**, i.e., water-loving or polar and **lipophilic**, i.e., fat-loving or non-polar components within the same molecule, but overall water insoluble.



Lipids have a long alkyl group that imparts hydrophobic or lipophilic character to the lipids and, usually, a polar and/or ionic group that is a hydrophilic component of the molecule, as illustrated in the figure on the right.

Lipids are a diverse group of biomolecules, including waxes, fats, fat-soluble vitamins (such as vitamins A, D, E, and K), phospholipids, steroids, terpenes, and others. They have a common property of being water insoluble. They can be dissolved in nonpolar solvents like ether or chloroform. Lipids have different functions in living things, including energy storage, signaling, hormonal activities, acting as structural components of cell membranes, etc.

Classification of lipids

Lipids are divided into eight categories: 1) fatty acyls; 2) glycerolipids; 3) glycerophospholipids; 4) sphingolipids; 5) saccharolipids; 6) polyketides; 7) steroids; and 8) prenol lipids. Six of the lipid categories are illustrated in Figure 6.1.1.

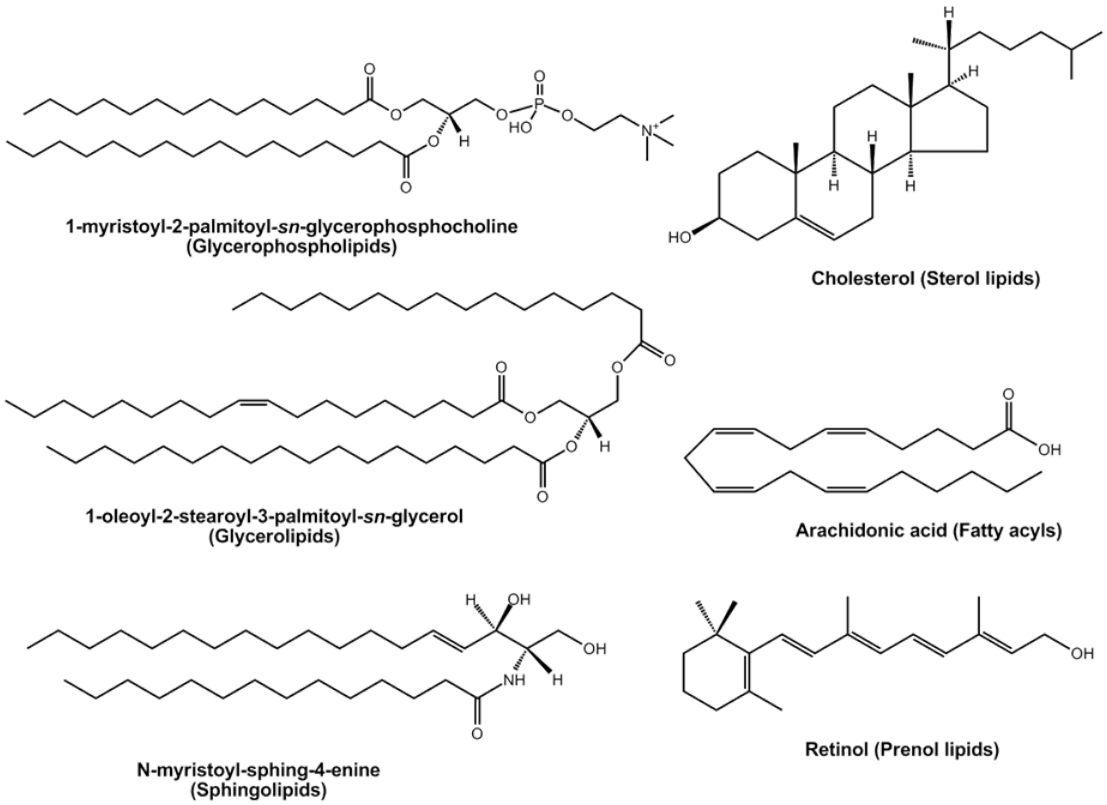
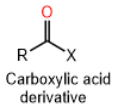


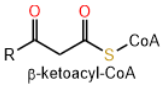
Figure 6.1.1: Examples of some lipid categories. (Copyright; Lmaps, GFDL 1.2, via Wikimedia Commons)

Five lipid categories contain long alkyl chain carboxylic acids or carboxylic acid derivatives. These include fatty acyls, glycerolipids, glycerophospholipids, sphingolipids, and saccharolipids. Polyketides are derived from condensation reactions of β-



Carboxylic acid derivative

ketoacyl subunits. Steroids and prenols are derived from condensation reactions of isoprene subunits. These categories of lipids are described in the following sections.



β-ketoacyl-CoA


Isoprene

6.2: Fatty acyls

Learning Objectives

- Understand the basic structural features of fatty acids, the notations reflecting them, and their relationship to the melting points and health effects.
- Define essential fatty acids, omega fatty acids, and their importance.
- Understand the basic structure features, nomenclature, and functions of prostaglandins.
- Understand the structures and uses of important waxes.

Fatty acyls are a group of lipids that contain a fatty acid or its derivative.

Fatty acids

Fatty acids

Fatty acids are carboxylic acids with an alkyl chain that is usually unbranched and containing an even number of C's, usually between 8 to 20 C's.

Figure $6.2.1$ shows some examples of fatty acids that have 18 C's in the chain. Carboxylic acid group ($-COOH$ is hydrophilic, and the alkyl chain is hydrophobic, making fatty acids amphiphilic. The hydrophobic property of the long alkyl chain dominates, making fatty acids insoluble in water.

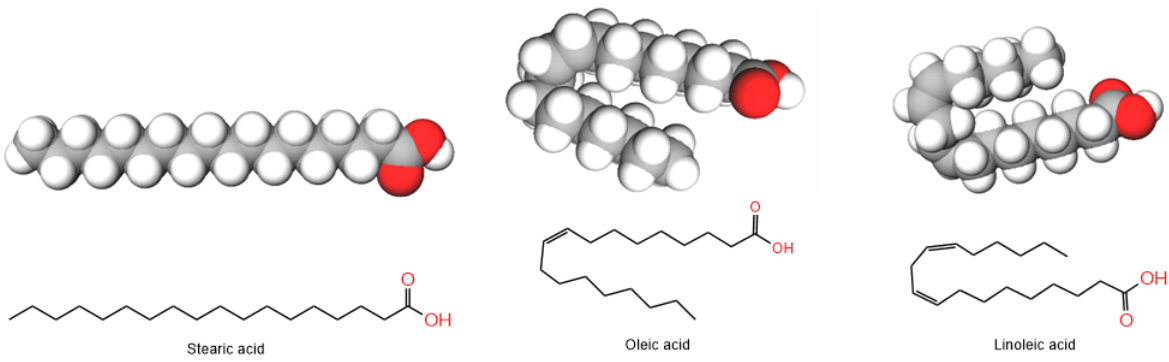


Figure $6.2.1$: Examples of fatty acids: a saturated steric acid (18:0), a monounsaturated oleic acid (18:1), and a polyunsaturated linoleic acid (18:2). Skeletal formulas are shown with the corresponding space-filling model shown on the top of each of the skeletal formula. (Copyright; Public domain)

Unsaturated, monounsaturated, and polyunsaturated fatty acids

- The alkyl chain can be saturated with all $C-C$ σ-bonds, called **saturated fatty acid (SFA)**, e.g., steric acid;
- or may contain some double bonds ($C=C$), called **unsaturated fatty acid (USFA)**., e.g., oleic acid and linoleic acid Figure $6.2.1$.
  - The unsaturated fatty acid may have one $C=C$ bond called **monounsaturated fatty acid (MUFA)**, e.g., oleic acid;
  - or more than one $C=C$ bonds called **polyunsaturated fatty acid (PUFA)**, e.g., linoleic acid.

Notations for and labelling C's in fatty acids

Short notations

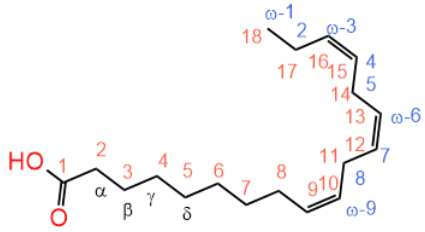
Carboxylic acids are usually referred to by their trivial names. **Short notation** is used to indicate # of C's:# of $C=C$ bonds, e.g., steric acid Figure $6.2.1$ is 18:0 meaning eighteen carbons and zero double bonds. Similarly, oleic acid is 18:1 meaning eighteen carbons and one double bond, and linoleic acid is 8:2 meaning eighteen carbons and two double bonds.

Labelling C′s

Systematic IUPAC nomenclature starts numbers from carbonyl-carbon (C=O) as shown in Figure 6.2.2, red numbers. Trivial nomenclature uses Greek alphabets starting from C#2 as α, next one as β, than γ, and so on, as shown in Figure 6.2.2, black numbers. **IUPAC short** notation shows the position labels of the double bonds' first C′s in a bracket, separated by commas, next to the regular short notation. The numbering begins from ($-COOH$ end. For example, steric acid 18:0, i.e., eighteen carbons and zero double bonds, oleic acid is 18:1(9), i.e., eighteen carbons and one double bond at C# 9, and linoleic acid is 18:2(9, 12), i.e., eighteen carbons and bonds, one at C# 9 and the other at C# 12.

Omega-minus (ω-#) labeling

In omega-minus (ω-#) labeling the last C, i.e., the $-CH_3$ at the end opposite to carboxyl ($-COOH$) group is labeled ω-1, the C next to it as ω-2, than ω-3 and so on, as as shown in Figure 6.2.2, in blue fonts. Recall that ω is the last alphabet in Greek, so, ω-1 is the last carbon. **The "omega-#"** (ω-#) label is usually used to refer to the position of the first C of the first C=C bond from the ω-end, i.e., from the $-CH_3$ end, opposite to the $-COOH$ end. For example, α-linolenic acid in Figure 6.2.2 is an ω-3 fatty acid.



alpha-Linolenic acid

Figure 6.2.2: Labelling of C′s in fatty acids: systematic IUPAC numbers are in red, nonsystematic Greek letter labels are in black, and omega-minus (ω-#) are in blue. (Copyright; Public domain)

Table 1 lists common fatty acids with their IUPAC short notations, ω-#, and melting points.

Table 1: Some of the common fatty acids

Common name	Condensed formula	IUPAC notation*	Melting point (^oC)	ω-#	Found in
Saturated fatty acids					
Caprylic acid	$CH_3(CH_2)_6COOH$	8:0	17		Milk
Capric acid	$CH_3(CH_2)_8COOH$	10:0	32		Coconut
Lauric acid	$CH_3(CH_2)_{10}COOH$	12:0	44		Coconut
Myristic acid	$CH_3(CH_2)_{12}COOH$	14:0	55		Nutmeg
Palmitic acid	$CH_3(CH_2)_{14}COOH$	16:0	63		Palm
Stearic acid	$CH_3(CH_2)_{16}COOH$	18:0	69		Animal fat
Arachidic acid	$CH_3(CH_2)_{18}COOH$	20:0	75		Corn
Monounsaturated fatty acids					
Myristoleic acid	$CH_3(CH_2)_3$**CH=CH** $(CH_2)_7COOH$	14:1(9)	0	$\omega-5$	Butter
Oleic acid	$CH_3(CH_2)_7$**CH=CH** $(CH_2)_7COOH$	18:1(9)	14	ω-9	Olives, pecan, grapeseed
Elaidic acid	$CH_3(CH_2)_7$**CH=CH** $(CH_2)_7COOH$	18:1(9t)	45	ω-9	Milk
Erucic acid	$CH_3(CH_2)_7$**CH=CH** $(CH_2)_{11}COOH$	22:1(13)	34	ω-9	Rapeseed
Polyunsaturated fatty acids					

Common name	Condensed formula	IUPAC notation*	Melting point (^oC)	ω-#	Found in
Linoleic acid	$CH_3(CH_2)_4$**CH=CH**CH_2**CH=CH**$(CH_2)_7$COOH	18:2(9,12)	-5	$\omega-6$	Soybean, safflower, sunflower
α-Linolenic acid	CH_3CH_2**CH=CH**CH_2**CH=CH**CH_2**CH=CH**$(CH_2)_7$COOH	18:3(9,12,15)	-11	$\omega-3$	Corn
Arachidonic acid	$CH_3(CH_2)_4$**CH=CH**CH_2**CH=CH**CH_2**CH=CH**CH_2**CH=CH**$(CH_2)_3$COOH	20:4(5,8,11,14)	-50	$\omega-6$	Meat, eggs, fish
Eicosapentaenoic acid	CH_3CH_2**CH=CH**CH_2**CH=CH**CH_2**CH=CH**CH_2**CH=CH**CH_2**CH=CH**$(CH_2)_3$COOH	20:5(5,8,11,14,17)	-65	$\omega-3$	Fish
Docosahexaenoic acid	CH_3CH_3**CH=CH**CH_2**CH=CH**CH_2**CH=CH**CH_2**CH=CH**CH_2**CH=CH**CH_2**CH=CH**$(CH_2)_2$COOH	22:6(4,7,10,13,16,19)	-44	$\omega-3$	Fish

* IUPAC notation shows Numbers of C′s:number of double bonds(position of *cis*-double bond, except when there is symbol t for *trans*-configuration)

Essential fatty acids

α-Linolenic acid (18:3(9,12,15)) -an $\omega-3$ fatty acid, and linoleic acid (18:2(9,12)) - an $\omega-6$ fatty acid are **essential fatty acids (EFAs)** because humans and other animals cannot synthesize them. They must be included in the diet as they are needed for good health. Although eicosapentaenoic acid (20:5(5,8,11,14,17)) and docosahexaenoic acid (22:6(4,7,10,13,16,19)), called **long-chain unsaturated fatty acids**, can be synthesized from α-linolenic acid, the conversion efficiency is low, and it is recommended that these should also be included in the diet. Arachidonic acid (20:4(5,8,11,14)) is another long-chain unsaturated fatty acid that can be synthesized from linoleic acid.

Shapes and physical properties of fatty acids

Shapes

The saturated fatty acids are cylindrical that pack nicely in the solid state with a large contact area between the molecules. The double bonds in unsaturated fatty acids are usually in the *cis* configuration. A *cis*-C=C introduces a kink in the structure that does not let molecules pack with a large contact area between neighboring molecules, as illustrated in Figure 6.2.3. Some unsaturated fatty acids have *trans*-double bonds and are cylindrical, like saturated fatty acids, but they are rare.

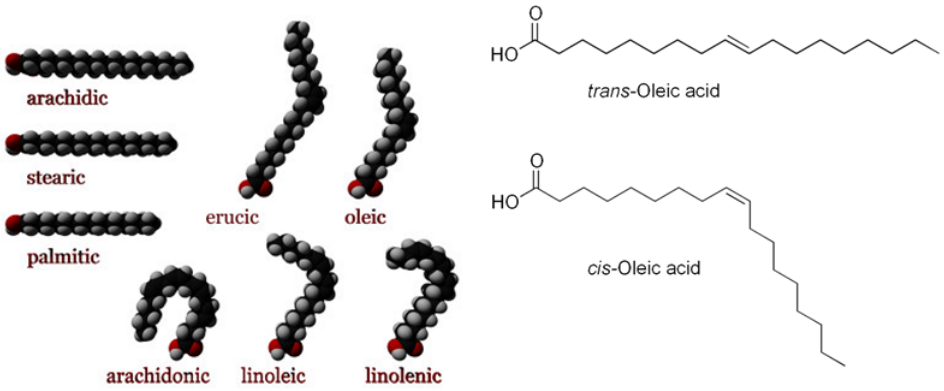


Figure 6.2.3: Three-dimensional representations of several fatty acids. Saturated fatty acids (arachidic 20:0, steric acid 18:0. palmitic acid 16:0) are cylindrical. Unsaturated ones (erucic 21:1, oleic 18:1, arachidonic 20:4, linoleic 18:2, and linolenic (18:3) are typically bent unless they have a trans configuration (Copyright: Automated conversion, CC BY-SA 3.0, via Wikimedia Commons), right: Oleic acid and its trans isomer (Copyright: Edgar181, Public domain, via Wikimedia Commons)

Melting points

The melting point of fatty acids is affected by two factors, as shown in Table 1:

1. an increase in the number of C′s increases the melting point because it increases London dispersion forces holding the molecules together, and
2. the presence of *cis*-C=C bond decreases the melting point due to the kink in the structure that reduces the contact area between molecules and, consequently, London dispersion forces are less.

Fats containing saturated fatty acids are considered unhealthy as they tend to deposit in arteries forming plaque, i.e., atherosclerosis. It can lead to high blood pressure, rupture of arteries, or heart attack. Saturated fatty acids are more common in animal fats, while unsaturated fatty acids, considered healthier, are more common in vegetable oils.

It has been observed that people from Alaska eat more unsaturated fats and have a low occurrence of atherosclerosis and heart attack. Fats in Alaskan food is primarily from fish, such as salmon, tuna, and herring, as shown in Figure 6.2.4. Vegetable oils have higher contents of ω-6 fatty acids, such as linoleic acid and arachidonic acid. Fish oil mainly contains ω-3 fatty acids, such as linolenic acid, eicosapentaenoic acid, and docosahexaenoic acid.

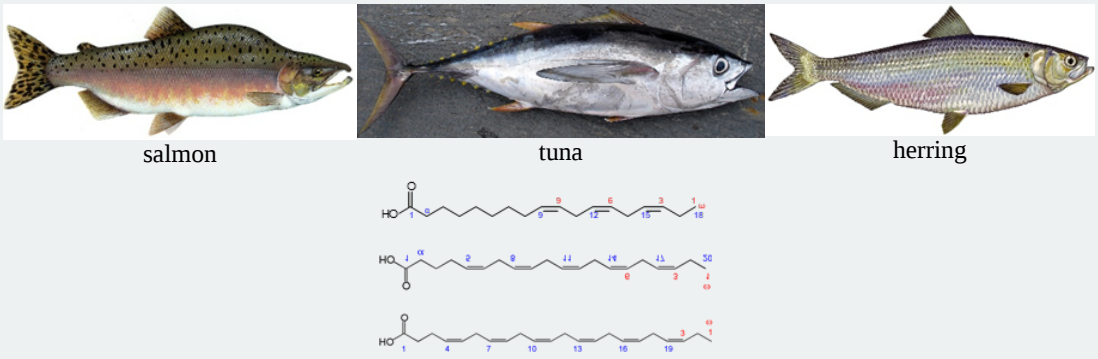


Figure 6.2.4: Three common fish in the diet of people of Alaska and three common ω-3 fatty acids found in the fish: linolenic acid (top), eicosapentaenoic acid (middle), and docosahexaenoic acid (bottom). (Copyright: Public domain).

Prostaglandins

Prostaglandins are lipids containing 20 C′s that are derived from arachidonic acid (20:4(5, 8, 11, 14)) and act like hormones. They are found in small concentrations in almost all body tissues and cells. Their activities include blood pressure regulation, blood clotting, inflammation, gastric secretions, kidney functions, and reproductive activities. They are usually short-lived and produced locally where they perform their function.

Prostaglandins synthesis

Prostaglandins synthesis starts from arachidonic acid using enzymes like cyclooxygenase (COX-1 and COX-2) and prostaglandin synthase, like PGE-synthase and PGF-synthase, as illustrated in Figure 6.2.5. The prostaglandins have a five-membered ring and two side chains. The prostaglandin PGG_2 is synthesized from arachidonate and quickly converts to PGH_2. PGH_2 regulates constriction and dilation of blood vessels and stimulates platelet aggregation. PGH_2 converts to PGE_2 and PGF_2 which induce labor by stimulating uterine contraction. PGF_2 also lowers blood pressure by relaxing smooth muscle cells in blood vessels. PGH_2 is also a precursor of several other Prostaglandins and related biomolecules.

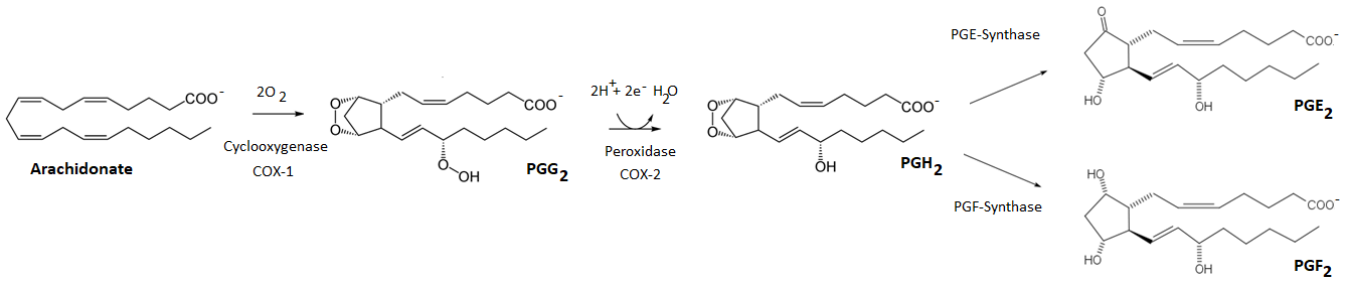


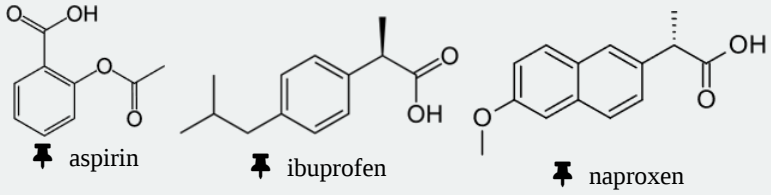
Figure 6.2.5: Synthesis of some prostaglandins from arachidonate. (Copyright; Public domain)

Prostaglandins names

The names of prostaglandins are abbreviated as three letters followed by a subscript, where the first two letters, i.e., PG, stands for prostaglandin, the 3rd letter indicates the substitution pattern, and the subscript tells the number of C=C bonds in the side chains. For example, the 3rd letter G has endoperoxide in the ring and a hydroperoxide group at C′s 15, H has endoperoxide in the ring, and a hydroxyl group at C′s 15, E has a β-hydroxyketone, and F has a 1,3-diol groups in the ring. The subscript 2 shows two C=C bonds in the side chains. The letter E in PGE_2 also indicates that it is soluble in ether and F in PGF_2 indicates it is soluble in phosphate buffer.

Nonsteroidal anti-inflammatory drugs (NSAIDs) action on prostaglandin synthesis

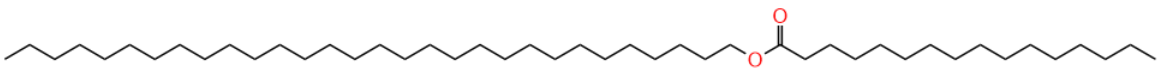
When tissues are injured, arachidonic acid is converted into prostaglandins that cause inflammation and pain. Nonsteroidal anti-inflammatory drugs, like aspirin, ibuprofen, and naproxen, shown below, inhibit the conversion of arachidonic acid into prostaglandins. Ibuprofen and naproxen inhibit the cyclooxygenase (COX) enzyme by binding with it. Aspirin deactivates the cyclooxygenase (COX) by transferring an acetyl group to the hydroxyl group of the active site of the cyclooxygenase. Care must be taken using NSAIDs as their long-term use can cause gastrointestinal, liver, or kidney damage.



Waxes

Lipid waxes are esters of one long-chain fatty acid with unbranched long-chain alcohol containing 14 to 30 C′s per chain. For example, myricyl palmitate found in beeswax can be obtained by the esterification of palmitic acid and myricyl alcohol:

$$CH_3(CH_2)_{14}-\overset{\overset{O}{\|}}{C}-OH + HO-(CH_2)_{29}CH_3 \longrightarrow CH_3(CH_2)_{14}-\overset{\overset{O}{\|}}{C}-O-(CH_2)_{29}CH_3 + H_2O$$



Functions of waxes

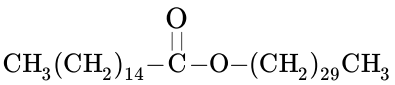
Waxes form a waterproof coating on plants' leaves, fruits, and stems to minimize water evaporation and protect against parasites. Examples include palm trees and jojoba wax from jojoba bushes. The wax coating on animals' skin, fur, and feathers protects water

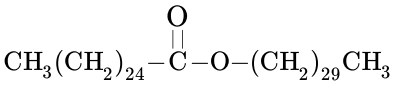
birds from water wetting and rainwater and shelters them from rainwater. Bees use wax to construct protection and housing for honey, as illustrated below. Sperm whales use spermaceti wax in their heads as a part of their eco analysis, i.e., navigation system.

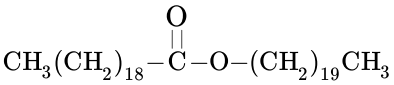
Honeycomb made of beeswax:

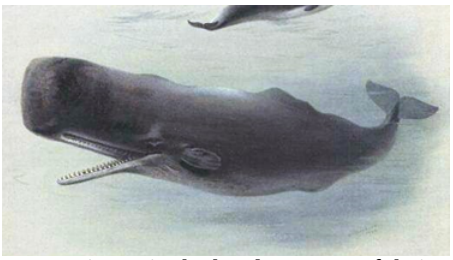
Carnauba wax coating on palm tree leaves:

Jojoba wax on leaves and fruits of jojoba bush:









Sperm whale has spermaceti wax in the head as a part of their eco-analysis syterm:

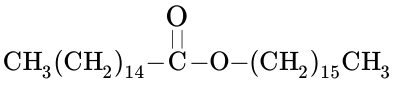


Figure 6.2.1: Examples of waxes, their uses, and formulas (Copyright: Public domain).

Beeswax and carnauba wax protective coatings on cars, furniture, floors, etc. Jojoba wax and spermaceti are used to make candles and cosmetics. A mixture of waxes obtained from wool is used in lotions for face and skin softening.

Petroleum waxes and earwax

Petroleum-derived waxes or paraffin are also called waxes, but they are not esters but mixtures of long-chain hydrocarbons. Earwax is a mixture of glycerol esters, cholesterol, phospholipids, etc., from dead skin and secretions of cerumen glands. Like other lipid waxes, it protects the ear canal against bacteria, fungi, and water. Excess earwax can cause blockage in the ear canal and hearing loss.

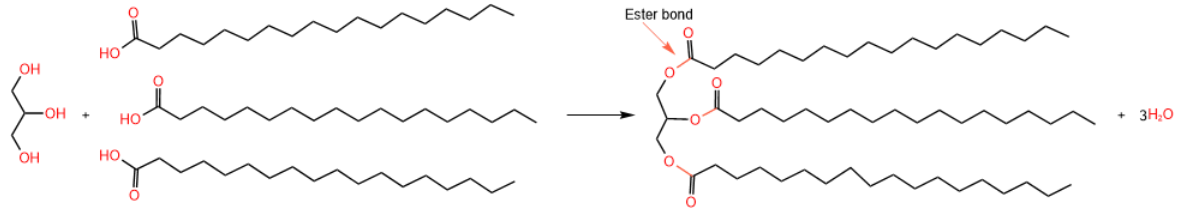
6.3: Glycerolipids

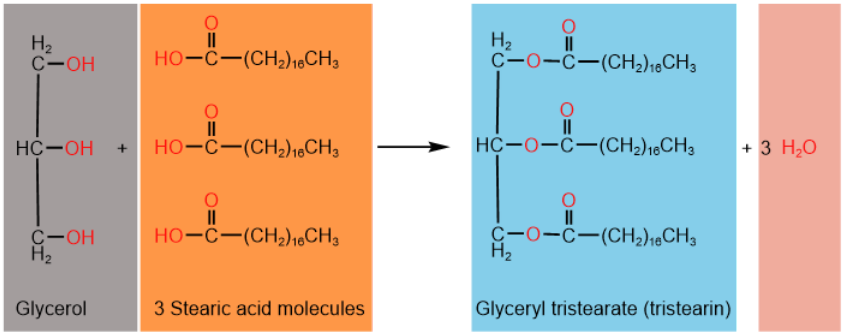
Learning Objectives

- Define glycerolipids and triglycerides and understand their structures.
- Understand the difference in composition and properties of fats and oils and their primary role as energy storage molecules.
- Understand the chemical properties of fats and oils, including hydrolysis, saponification, and partial or complete hydrogenation.

What is a glycerolipid?

Glycerolipids have two components: propane-1,2,3-triol, also called glycerol, and one, two, or three fatty acids. Recall that a condensation reaction between an alcohol and a carboxylic acid forms an ester by eliminating a water molecule. The following example shows the reaction of glycerol with three molecules of stearic acid, creating a triester.



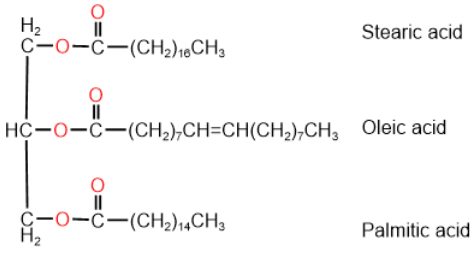


A mono-, di-, or tri-ester of glycerol is called glycerolipid.

The name of a glycerolipid begins with glyceryl and is followed by carboxylate. For example, the triester of glycerol shown above is glyceryl tristearate, which is commonly known as tristearin.

Glycerolipids that are triesters of glycerol with three fatty acids are called **triacylglycerol** or **triglycerides**. For example, glyceryl tristearate shown above is a triglyceride.

Triglycerides may be esters of three molecules of the same fatty acids, e.g., tristearin. Still, often the triglycerides found in nature are esters of two or three different fatty acids. Fatty acids usually found in fats and vegetable oils include lauric, myristic, palmitic, stearic, oleic, linoleic, and linolenic acids. An example of a mixed triglyceride of stearic, oleic, and palmitic acid is shown below.



A mixed triacylglycerol

Fats and oils

Fats and oils are triglycerides used as energy storage molecules in animals and plants.

Energy storage is essential for hibernating animals that live in icy environments. They have plenty of food available during summer but no food and below-freezing temperatures in winter. They eat plants, seeds, and nuts and accumulate fat that becomes their energy source when hibernating in winter. For example, the polar bear shown on the right can accumulate up to 14 kg of fat per week in summer when the food is plenty and uses it as an energy source when the animal hibernates for 4 to 7 months in winter.

Although glycogen is a quick energy source, fats and oils are more energy dense for two reasons:

- first, fats and oils are in a more reduced form than glycogen and release more energy per mass upon oxidation, and
- second, glycogen is hydrophilic and absorbs water, increasing its mass.

Physical properties of fats and oils

Pure fats and oils are colorless, odorless, and tasteless. The color, taste, and odor associated with butter -an animal fat, and olive oil -a vegetable oil, is due to a small number of other substances mixed with the triglycerides.

Fats come from animal sources, such as meat, whole milk, lard, butter, and cheese. Vegetable oils come from plant sources, e.g., soybeans, peanuts, sunflowers, etc.

Fats are usually solid at room temperature because they have a higher proportion of long-chain unsaturated fatty acids with higher melting points, e.g., lard and butter shown in Figure 6.3.1. **Vegetable oils are usually liquid** at room temperature because they contain more unsaturated or short-chain saturated fatty acids.

Figure 6.3.1: Butter -a fat that is semisolid, and olive oil -a vegetable oil liquid at room temperature. (Copyright:Butter: Renee Comet (National Cancer Institute), Public domain, via Wikimedia Commons, Olive oil: Roberta Sorge robertina, CC0, via Wikimedia Commons)

For example, canola oil, safflower oil, flaxseed oil, sunflower oil, corn oil, olive oil, soyabean oil, peanut oil, and cottonseed oil are vegetable oils that are liquid at room temperature and have a higher proportion of unsaturated fatty acids as shown in Figure 6.3.2. Palm and coconut oils are vegetable oils with a higher balance of saturated fatty acids. Still, they are liquid at room temperature because they have short-chain saturated fatty acids. For example, coconut mainly comprises lauric acid (12:0), a short-chain fatty acid.

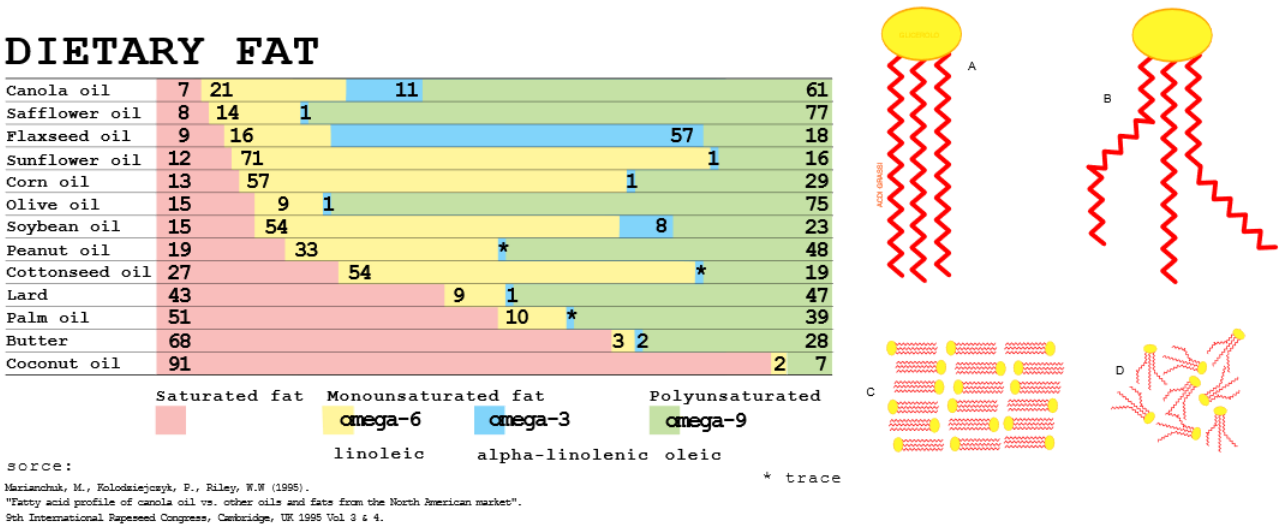


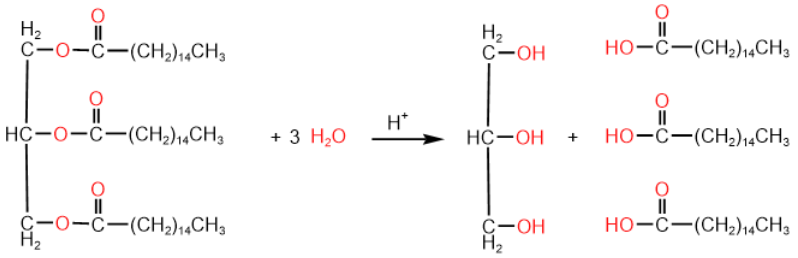
Figure 6.3.2: Right: Proportion of saturated, monounsaturated, and polyunsaturated fatty acids in fats (lard and butter) and vegetable oils. (Copyright; Diamboroid, CC BY-SA 4.0, via Wikimedia Commons). Left: Saturated triglyceride (top left) and unsaturated (top right). The drawing below shows how packing is easier in saturated triglycerides, so they tend to be solid at room temperature. (Copyright; AGeremia, CC BY 3.0, via Wikimedia Commons)

Chemical properties of fats and oils

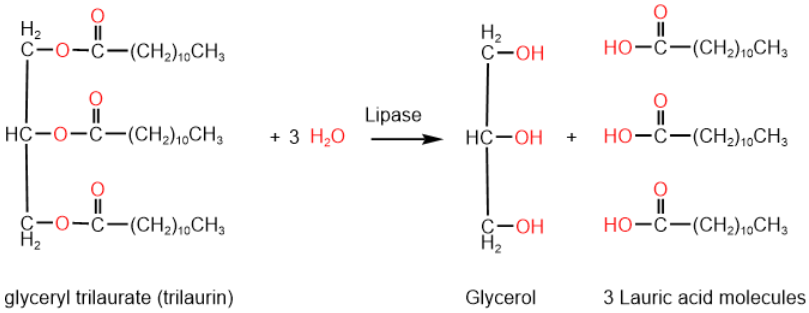
Three important reactions of triglycerides are hydrolysis by water in the presence of acid or lipase enzyme, saponification, i.e., base-promoted hydrolysis, and hydrogenation of double bonds in the fatty acids.

Hydrolysis

Easters are hydrolyzed (split) by water in the presence of an acid catalyst into alcohol and carboxylic acid. The same reaction happens with triglycerides, i.e., fats and vegetable oils—for example, tripalmitin hydrolyzes into glycerol and three palmitic acid molecules, as shown below.



Lipase enzymes do the same reaction during the digestion of triglycerides, e.g., trilaurin is hydrolyzed by lipase to glycerol and three molecules of lauric acid, as shown in the reaction below.

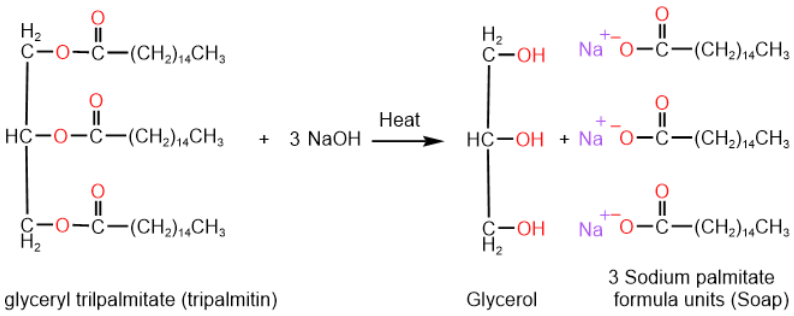


The digestion of lipids starts in the mouth with lingual lipases secreted by glands in the tongue and continues in the stomach with lingual and gastric lipases. Glycerol in the hydrolysis product is soluble in water, but fatty acids are not. Fatty acids are emulsified

in the stomach and later mixed with bile and pancreatic juice to continue the process of digestion.

Saponification

Saponification is base-promoted hydrolysis of triglycerides that produces glycerol and salts of the fatty acids, a soap, as shown below for the case of saponification of tripalmitin.



Different varieties of soaps are shown in the figure on the right. Sodium salts of fatty acids make hard soaps, while potassium salts make soft soaps. Three of the most common soaps are sodium stearate, oleate, and linoleate. Salts of saturated fatty acids make rigid, and those of polyunsaturated fatty acids make soft soap. Perfumes are added for scent and dyes for color, sand is added in scouring soap, and the air is blown into the soap to make it float on water.

Hydrogenation

Alkenes add H_2 and covert to alkane in the presence of Ni, Pt, or Pd catalyst. The same reaction happens with the C=C in unsaturated fatty acids found in triglycerides. The conversion of unsaturated fatty acids to saturated fatty acids increases the melting point of triglycerides. Therefore, vegetable oils with more unsaturated fatty acids become semisolid or solid after partial or complete hydrogenation, as illustrated by the reaction in Figure 6.3.3. Hydrogenation of vegetable oils is a commercial process to convert vegetable oils to semisolid products like margarine and shortening.

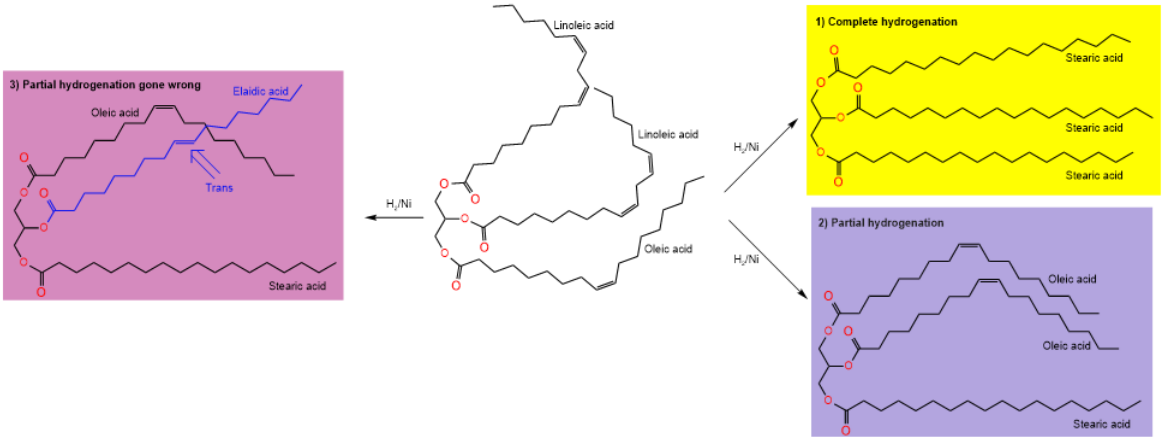


Figure 6.3.3: Hydrogenation of unsaturated fatty acids in vegetable oils: 1) complete hydrogenation, 2) partial hydrogenation, 3) partial hydrogenation that results in conversion of some *cis*-C=C bonds to *trans*-C=C bonds that are undesirable. (Copyright; Public domain)

The driving force for the industrial process of hydrogenation of vegetable oils is that margarine or shortenings are used as cheaper alternatives to butter with a longer shelf-life. Complete hydrogenation of vegetable oils converts them into hard margarine. Often partial hydrogenation is performed as the partially hydrogenated product is semisolid and mimics butter better than hard shortening. Dyes and flavors are mixed with it to mimic butter's color, taste, and odor.

Unwanted trans-fatty acids are by-products of the hydrogenation of vegetable oils.

Like all other processed food products, hydrogenated vegetable oil products are associated with health problems. Partial hydrogenation converts some of the *cis*-C=C bonds to *trans*-C=C bonds, as illustrated in Figure 6.3.3. *Trans*-C=C bonds are rare in some fatty acids. Research reports indicate that trans-fatty acids affect blood cholesterol like unsaturated fatty acids. Research reports also indicated that trans-fatty acids increase the level of low-density lipids (LDL), considered bad cholesterol,

and decrease the level of high-density lipids (HDL), regarded as good cholesterol. Health organizations force the manufacturers to show the trans-fatty acid contents on the product labels. They are trying to educate the consumers about trans-fatty acids, as shown in Figure 6.3.4 from U.S. Food and Drug Administration.

Figure 6.3.4: A variety of processed foods—including the frozen, canned, and baked goods shown—contain *trans* fat. The amount per serving is listed on the Nutrition Facts label. Including partially hydrogenated oil in the list of ingredients is another indication that *trans* fat is present. (Copyright; The U.S. Food and Drug Administration, Public domain, via Wikimedia Commons)

6.4: Glycerophospholipids

Learning Objectives

- Define glycerophospholipids and understand the structural features that allow them to organize into lipid bilayers.
- Understand the classification of glycerophospholipids and their role in infant respiratory distress syndrome.

What are glycerophospholipids?

Glycerophospholipids are triesters of propane-1,2,3-triol (glycerol): 1st and 2nd ester bonds with fatty acids, and the 3rd ester bond with phosphate, as illustrated in Figure 6.4.1. The phosphate group is often a diester: an ester with the primary alcohol of glycerol and 2nd ester bond with a small molecule, such as choline, ethanolamine, serine, or inositol.

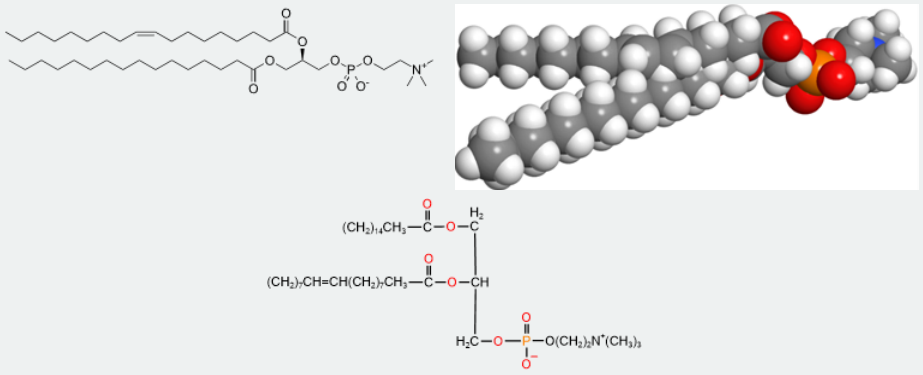
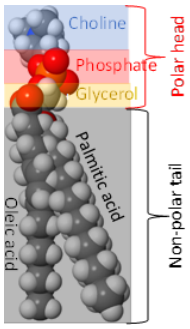


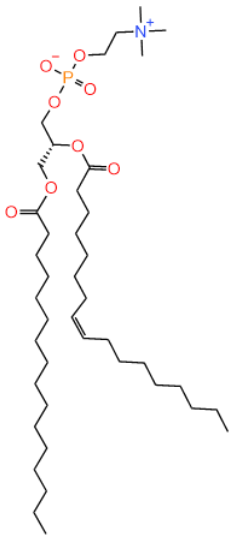
Figure 6.4.1: Chemical structure of a phospholipid (1-palmitoyl-2-oleoylphosphatidylcholine): skeletal formula (left); space filling model showing C′s gray, O′s red, N blue, and P orange (middle); and condensed formula (right). (Copyright: Public domain).

Phospholipids have two hydrophobic tails, which are the hydrocarbon groups of two fatty acids and a polar head comprising of ester groups with fatty acids, phosphate having a -ve charge, and often also have a +ve charge on the nitrogen of the small molecule attached with the phosphite, as shown in the figure on the right. The 1st fatty acid attached to the primary alcohol group is usually a saturated fatty acid, and the 2nd fatty acid attached to the secondary alcohol group of glycerol is usually an unsaturated fatty acid.

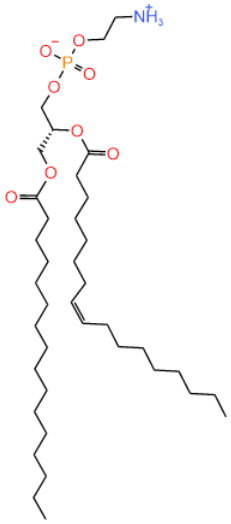


Classes of glycerophospholipids

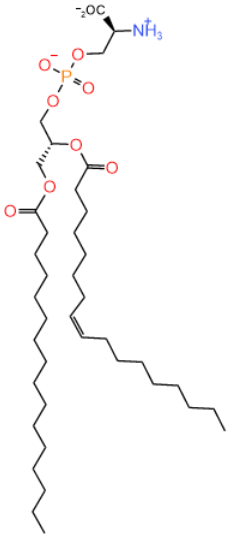
The glycerol esterified with two fatty acids and one phosphoric acid is also called **phosphatidyl**. Phosphatidyls with phosphate esterified with choline are called **phosphatidylcholines** (**lecithin**). Phosphatidyl esterified with ethanolamine, or sometimes with serine, are called phosphatidylethanolamines or phosphatidylserine (common name cephalins). Another category is phosphatidylinositols with inositol-esterified phosphate, as shown in Figure 6.4.2.



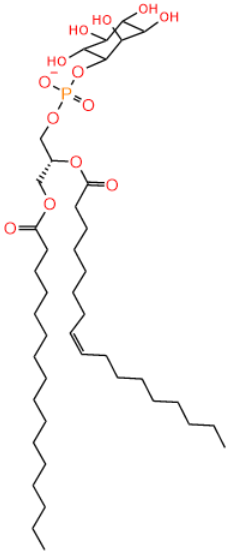
1-palmitoyl-2-oleoyl phosphatidylcholine (a lecithin)



1-palmitoyl-2-oleoyl phosphatidylethanolamine (a cephalin)



1-palmitoyl-2-oleoyl phosphatidylserine (a cephalin)



1-palmitoyl-2-oleoyl phosphatidylinositol

Figure 6.4.2: Major categories of glycerophospholipids. (Copyright; Public domain)

Lipid bilayer

Recall the general solubility rule "like dissolves like". Oil droplets that are nonpolar floating on the surface of the water, which is polar, ultimately coalesce, making bigger droplets. Glycerophospholipids are amphiphilic, having an overall cylindrical shape with nonpolar tails of fatty acid chains and polar heads of ester groups on the glycerol part and ionic groups on phosphate and amine groups. Glycerophospholipids form a bilayer in the watery environment of living things based on the principle "like dissolves like", as illustrated in Figure 6.4.3.

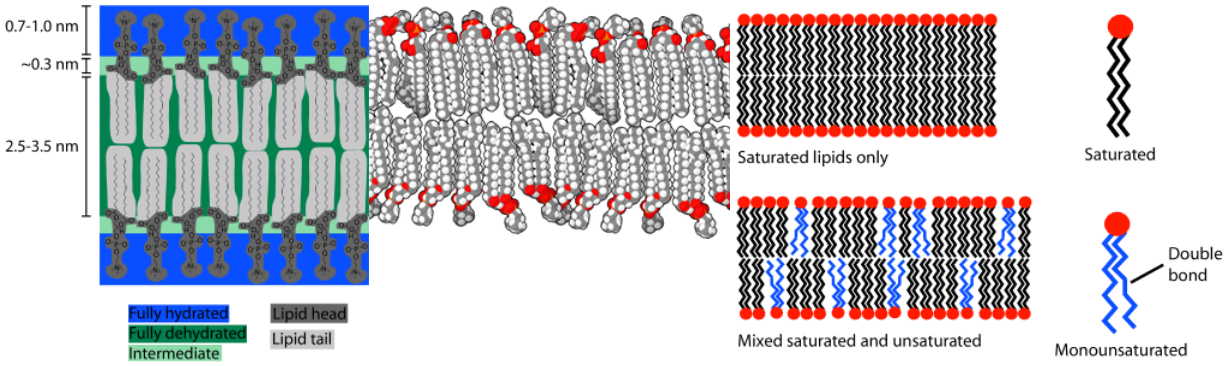


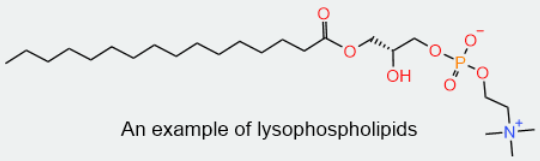
Figure 6.4.3: Lipid bilayer cross-section of phosphatidylcholine as a model phospholipid: illustration of hydrophilic and hydrophobic layers in contact with water on the interior and exterior of cells (left, Copyright: MDougM, Public domain, via Wikimedia Commons), simulation of fluid nature of the lipid bilayer (middle, Bensaccount, Public domain, via Wikimedia Commons), and illustration of the control of fluidity of the bilayer based on the proportion of unsaturated/saturated fatty acids (right, Copyright: MDougM, Public domain, via Wikimedia Commons).

The hydrophobic tails of fatty acids form the interior of the bilayer, and the hydrophilic polar heads make the outer layers in contact with water. The molecules are in a fluid (dynamic) state due to thermal energy (Figure 6.4.3 middle) but held together by intermolecular interactions: dipole-dipole interaction in the polar heads and London dispersion forces in the hydrophobic tails. Saturated fatty acids pack tightly together, making a rigid bilayer. However, adding unsaturated fatty acids with kinks makes the packaging lose, resulting in a more fluid bilayer, as illustrated in Figure 6.4.3 right.

The glycerophospholipid bilayer is a significant component of the cell membrane, the membranes around the nucleus, and some cell organelles. The hydrophobic nature of the interior of the bilayer becomes a barrier to the movement of ions and water in and out of the cell. Still, the bilayer's fluid nature allows the diffusion process to move some particles across the membrane. Other components of the cell membrane, such as cholesterol, adjust the rigidity, and proteins control the movement of particles in and out of the cell as needed. The structure of cell membranes is described in a later section.

Snake wrangler demonstrating how his venom flows from fangs at certain movement. (Copyright: Usman Ahmad from Pakistan, CC BY-SA 2.0, via Wikimedia Commons)

Snake venom

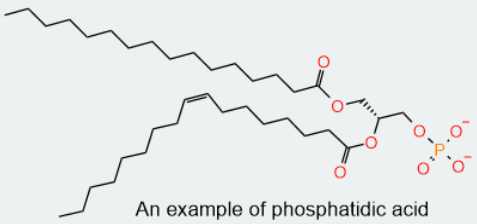


An example of lysophospholipids

Snakes inject their venom through fangs when they bite their prey, as shown in the figure on the right. The venom of eastern diamondback rattlesnakes and Indian cobra contains phospholipase enzymes that catalyze the hydrolysis of fatty acid on the secondary hydroxyl group of glycerol. The resulting phospholipids with one less carboxylate group, called lysophospholipids, is no more cylindrical molecule and do not fit correctly in the lipid-bilayer, causing the red blood cell membrane to rupture. This poses a significant health risk and may kill humans. The structure of the lysophospholipid example is shown in the figure on the left.

Infant respiratory distress syndrome (IRDS)

In the breathing process, the exchange of O_2 and CO_2 takes place in air sacs called alveoli within the lungs, as illustrated in Figure 6.4.4. Pulmonary surfactants, composed of a mixture of lecithin and sphingomyelin lipids, are released into the lungs of newborn babies that reduce the surface tension and ease the inflation of alveoli. Lecithins are mixtures of glycerophospholipids, including phosphatidylcholine, phosphatidylethanolamine,



An example of phosphatidic acid

phosphatidylinositol, phosphatidylserine (shown in Figure 6.4.2) and phosphatidic acid shown in the figure on the right.

In the cases of premature birth, before 28 weeks of gestation, the level of the surfactant is insufficient, causing the air sacs to collapse and have to reopen with each breathing, a condition called **infant respiratory distress syndrome**. This makes breathing difficult, may damage the alveoli, and lead to hypoxia and acidosis due to less oxygen inhalation and less carbon dioxide exhalation.

The lecithin to sphingomyelin ratio in mature fetal lungs is ~2.5, the ratio 2.4 to 1.6 poses a low risk, and less than 1.5 poses a high risk of infant respiratory distress syndrome. Treatment includes steroids given to the mother or infant to assist the development of lungs, the application of surfactants, and the use of a ventilator to help breathing.

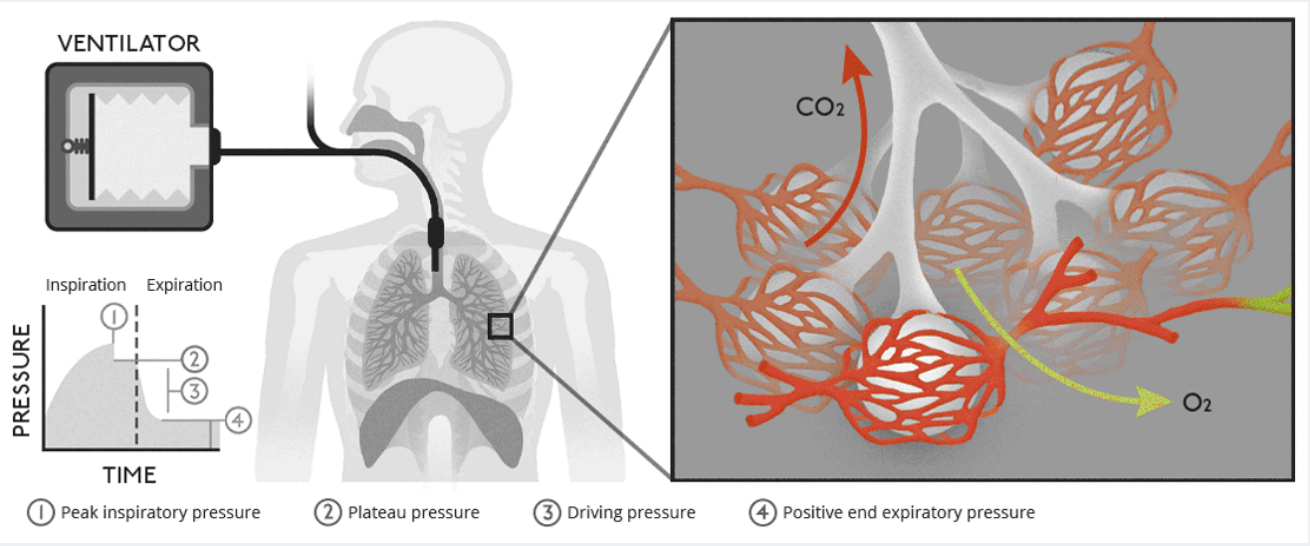


Figure 6.4.4: Illustration of the inflation of alveoli with air pressure in the lugs during inhalation accompanied by the exchange of O_2 and CO_2, and assistance of the process by a ventilator. (Copyright; Eleanor Lutz, CC0, via Wikimedia Commons)

6.5: Sphingolipids

Learning Objectives

- Define and understand the structure of sphingolipids, their subclasses, and their role in the nervous system.
- Understand the role of sphingolipids in multiple sclerosis.

Sphingolipids and their subclasses

Sphingolipids are a class of lipids having a sphingoid base that is a set of aliphatic amino alcohols, including **sphingosine,** shown in Figure 6.5.1. When the amino ($-NH_2$ group is attached to a fatty acid by an amide bond, it is called **ceramide**. When the primary alcohol of ceramide is linked to a phosphorylcholine or phosphorylethanolamine group, it is called **sphingomyelin**. When the primary alcohol of ceramide is bound to glucose or galactose by a glycosidic bond, it is called **cerebroside**. When a glycosidic bond connects the primary alcohol of ceramide to an oligosaccharide with one or more sialic acids, it is called **ganglioside**.

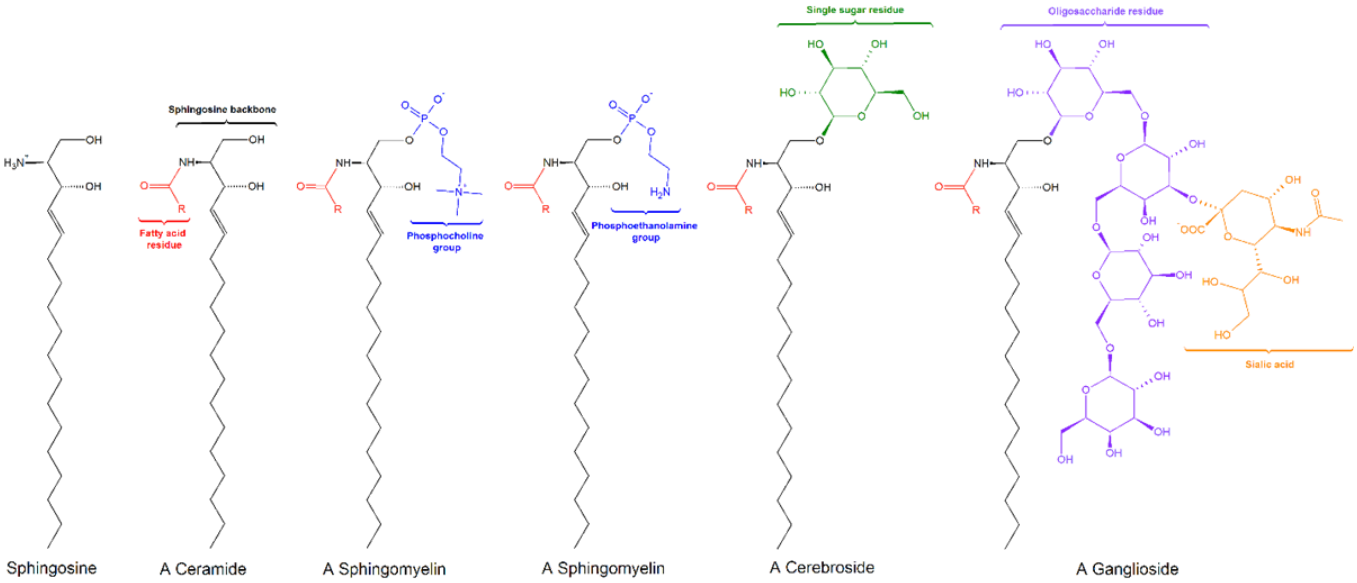


Figure 6.5.1: General structures of sphingolipids. (Copyright; LHcheM, CC BY-SA 3.0, via Wikimedia Commons)

Figure 6.5.2 illustrates two examples of sphingomyelin. Like glycerophospholipids, these sphingolipids are an essential component of the lipid bilayer. Mainly, they are abundant in the brain and nerves. They are abundant in the white matter of myelin sheath, i.e., a coating surrounding the nerve cells. Myelin sheath increases the speed of nerve impulses and is essential in protecting nerve cells, signal transduction, and cell recognition.

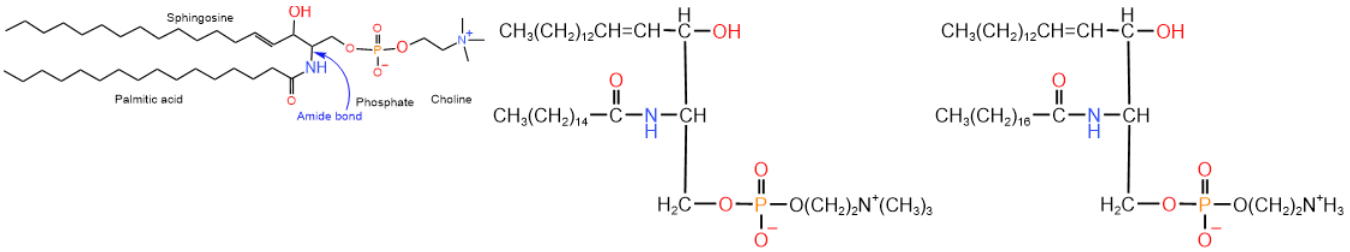


Figure 6.5.2: Two examples of sphingomyelin: one containing palmitic acid and choline (right -a skeletal formula, and middle -a condensed formula) and the other containing stearic acid and ethanolamine (right). (Copyright: Public domain).

Multiple sclerosis

Multiple sclerosis is a nervous system disease in which the myelin sheaths wrapped around axons of nerve cells are damaged, as illustrated in Figure 6.5.3. Symptoms are related to the retardation of signal conduction by the nerves that, in turn, reduces sensation, coordination, movement, cognition, muscle weakness, blindness, and other functions involving nerves. The severity

of the effects depends on the amount of damage. Studies indicate that vitamin D may lessen the severity of the disease. Nearly 1 million people in the US and about 2.8 million worldwide have multiple sclerosis.

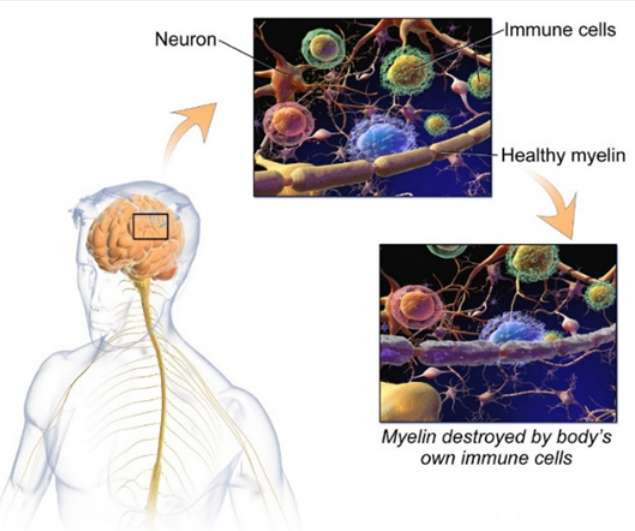


Figure 6.5.3: Multiple sclerosis is currently thought to be a disease in which autoimmune cells attack the nervous system, destroying myelin. (Copyright; BruceBlaus, CC BY-SA 4.0, via Wikimedia Commons)

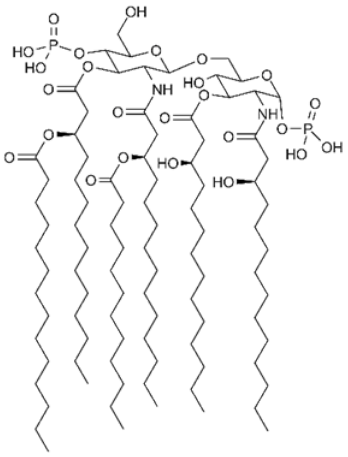
6.6: Saccharolipids, polyketides, and prenols

Learning Objectives

- Define and understand the basic structural features of saccharolipids, polyketides, and prenols.
- Define mono-, sesqui-, di-, and tri-terpenes and recognize isoprene units in terpenes, terpenoids, and steroids.

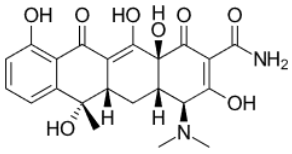
Saccharolipids

Saccharolipids are fatty acids linked to a saccharide (carbohydrate) backbone by linkages other than glycosidic linkages. They are compatible with membrane bilayers. For example, saccharolipid lipid A found in E. Coli is shown in the figure on the right.

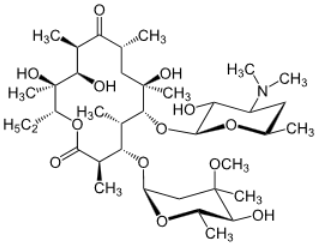


Polyketides

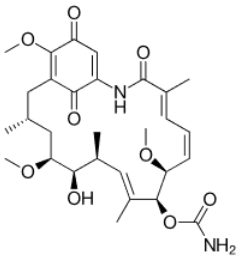
Polyketides are a broad class of natural products derived from β-polyketones, i.e., compounds containing $[-\overset{\overset{O}{\|}}{C}-CH_2-]_n$ repeat units in their backbone or their reduced forms, such as $[-\overset{\overset{OH}{|}}{C}-CH_2-]_n$, and $[-CH_2-CH_2-]_n$. Polyketides are important in the pharmaceutical industry. They are used as antimicrobial, antiparasitic, and anticancer agents; some are toxins. About 20% of the top-selling medicines are polyketides. Some examples of medically important polyketide products are shown below.



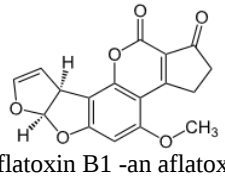
Tetracycline -an antibiotic



Erythromycin A -an antibiotic



Geldanamycin -an antitumor antibiotic



Aflatoxin B1 -an aflatoxin

Prenols

Prenols are synthesized from five-carbon isoprene precursors: isopentenyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP), as shown in Figure 6.6.1 (left). The five-carbon skeleton of isoprene units in these products can often be easily distinguished, as shown in by different color sections of some examples in Figure 6.6.1 (right). Since they contain a multiple of five C′s in their skeleton, they are also classified as monoterpens (10 C′s), sesquiterpenes (15 C′s), diterpenes (20 C′s), triterpenes ((30 C′s), and tetraterpenes (40 C′s). Terpenoids are modified terpenes that contain additional functional groups, usually oxygen functional groups as those shown in 6.6.1 (right).

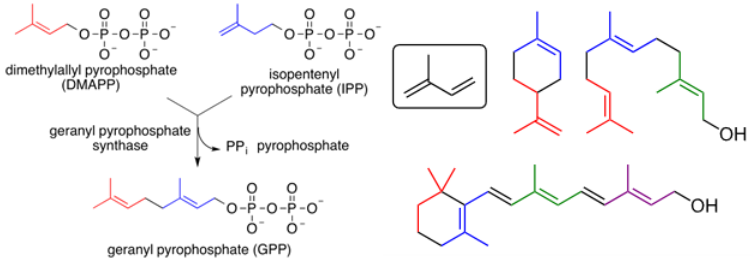


Figure 6.6.1: Dimethylallyl pyrophosphate (DMAPP) and isopentenyl pyrophosphate (IPP) condense to produce geranyl pyrophosphate, a precursor to all terpenes and terpenoids (left, Copyright; Andrew Murkin, CC0, via Wikimedia Commons) and structure of Isoprene, limonene, farnesol, and retinol, respectively, showing isoprene skeleton in these products in different color sections (right, Copyright; Calvero., Public domain, via Wikimedia Commons)

.

Steroids -other lipids described later are derived from triterpene squalene, as illustrated in Figure 6.6.2.

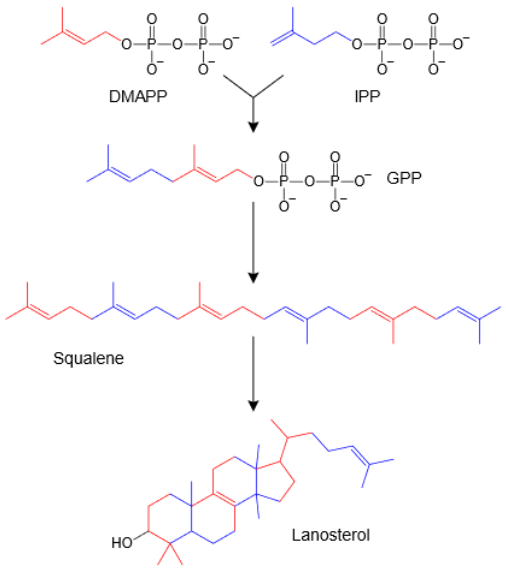


Figure 6.6.2: A simplified scheme for steroid biosynthesis via terpenoid intermediates: isopentenyl pyrophosphate (IPP), dimethylallyl pyrophosphate (DMAPP), geranyl pyrophosphate (GPP) and squalene. (Copyright; Fvasconcellos, original by Tim Vickers, Public domain, via Wikimedia Commons)

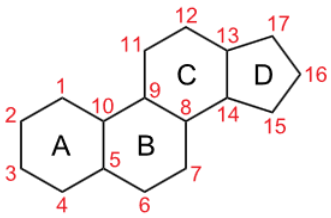
6.7: Sterols

Learning Objectives

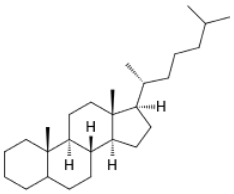
- Understand the core structure of steroids and sterols.
- Understand the structural features of cholesterol, its production, consumption, transport, and medical problems.
- Understand bile acids' structure features, their functions, and medical problems.
- Understand the structural features of steroid hormones, functions, abuses, and medical problems.

What are steroids?

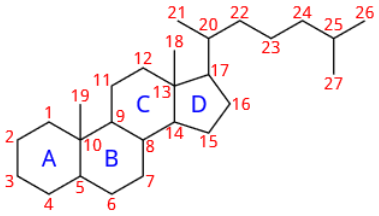
Steroids are biologically active compounds that have a core structure composed of four fused rings in a specific configuration: three six-membered rings designated A, B, and C and one five-membered ring D, as Figure 6.7.1 right. Gonane has this core structure of steroids. Other steroids have gonane skeletons with some groups attached. For example, cholestane has methyl groups at positions 10 and 13 and an aliphatic chain at position 17. Cholesterol has the structure of cholestane. It is a biomarker found in rocks and petroleum deposits.



Gonane or sterane with C′s numbered



Cholestan with stereochemistry shown

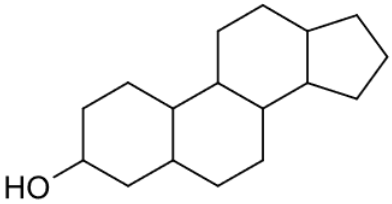


cholestan with C′s numbered

Figure 6.7.1: Steroid core structure represented by gonane with C labels shown and cholestane that has the C skeleton of cholestrol with C labels. (Copyright; Public domain)

What are sterols?

Sterols are organic compounds derived from gonane with H #3 replaced with an alcohol (−OH) group. The sterols are a sub-class of steroids. The simplest sterol is the alcohol gonane shown in the figure on the right. Other sterols have other groups attached to the gonane structure.



Cholesterol

Cholesterol is the most abundant sterol found in animals. It is a structural component of animal cell membranes, a precursor for steroid hormones, bile acids, and vitamin D. Cholesterol has the C skeleton of cholestane, with an −OH group at position #3 and a C=C bond at position 5 and 6, as shown in Figure 6.7.2.

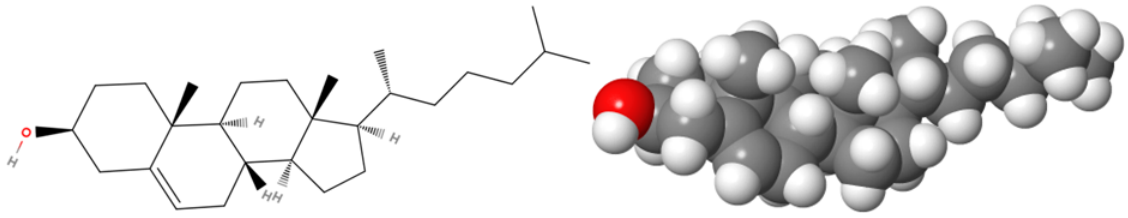


Figure 6.7.2: Structure of cholesterol (left) and space-filling model of cholesterol: hydrophilic head, i.e., −OH group (O is read and H is white) and hydrophobic tail of tetracyclic rings and an alkyl side chain (C′sare gray and H′s are white). (Copyright; Public domain)

Cholesterol sources and consumption

Cholesterol is biosynthesized from lanosterol, which, in turn, is synthesized from squalene, as described in the previous section. Cholesterol is synthesized in the liver using fats, proteins, and carbohydrates as raw materials. An average human weighing 68 kg synthesizes about 1000 mg of cholesterol daily. Food is also a source of cholesterol intake, as shown in the figure on the right.

Typically about 307 mg of cholesterol intake is through food per day in the US. The liver reduces cholesterol production to adjust for the intake from food sources.

Cholesterol is amphiphilic like phospholipids and sphingolipids, with $-OH$ as a polar head and cholestane structure as a hydrophobic tail. The cholesterol molecule is stiff due to restricted movements around $C-C$ bonds locked in the cyclic structure.

> Cholesterol is used in the cell membrane to impart rigidity to the ring -the higher the proportion of cholesterol in the cell membrane, the more rigid the membrane.

Usually, cholesterol is 30% of the cell membrane. It is involved in signaling and is found in the brain, plasma membrane, and myelin sheath of nerve cells. Cholesterol is a precursor to vitamin D and steroid hormones, including corticosteroids, such as cortisol and aldosterone, and sex hormones, such as progesterone, estrogens, and testosterone.

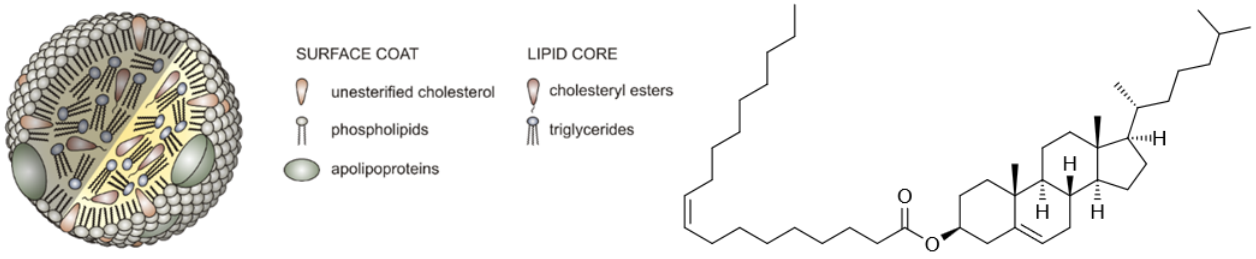
The liver recycles the excess cholesterol by converting it into bile acids stored in the gallbladder and excreted into the digestive tract. Bile acids emulsify fats in the food, which helps digest the food. About 50% of the excreted cholesterol is re-absorbed into the bloodstream by the small intestine.

Cholesterol contents of foods

Foods	Cholesterol (mg/100 g)
Beef brain	3100
Egg yolk	1085
Egg	373
Lamb kidney	337
Pork liver	301
Butter	215
Lobster	200
Shrimp	125
Light whipping cream	111
Yellow cheese	108
Beef	90
Chicken	88
ChPork	80
Fish	70
Yogurt	13
Egg whites	0
Fruits	0
Grains	0
Nuts	0

Transport of cholesterol

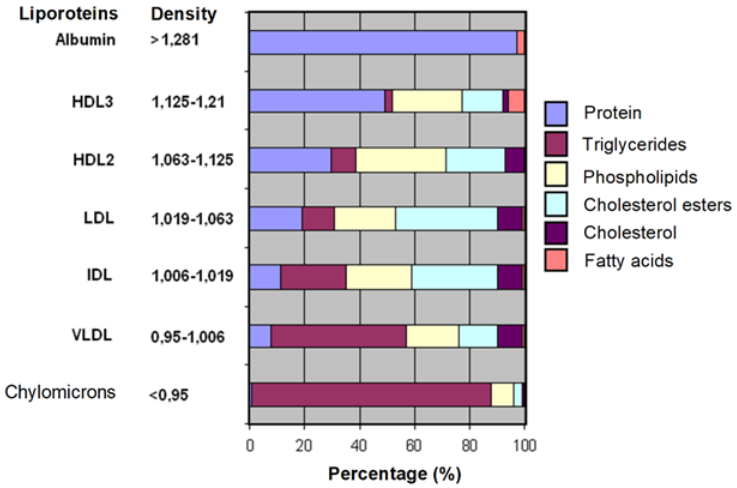
Cholesterol is insoluble in water like other lipids. It is transported by the blood, which is water rich medium. Like soap carries greases in water during washing, lipoprotein transport glycerolipids and cholesterol in the body. The lipoproteins exist as particles suspended in blood. Glycerides and cholesteryl esters are in the core, and phospholipids, proteins (apolipoprotein), and cholesterol form the outer layer having their hydrophobic parts oriented outwards in contact with water, as illustrated in Figure 6.7.3.



High-density lipoprotein (HDL) Cholesterol oleate

Figure 6.7.3: High-density lipoprotein (HDL), as an example lipoprotein particles, are composed of a lipid core containing cholesteryl esters and triglycerides, and a surface coat of phospholipids, unesterified cholesterol, and apolipoprotein (left, copyright: AntiSense, CC BY-SA 3.0, via Wikimedia Commons) and cholesterol oleate as an example of cholesteryl ester (right, Public domain)

Lipoproteins are classified based on their density, as shown in the figure on the right (*Copyright: Modified from İnfoCan at Turkish Wikipedia., FAL, via Wikimedia Commons, Public domain*). Their density is based on the proportion of proteins in them -the higher the proportion of proteins, the higher the density. They also have roles in the transportation process of cholesterol and lipids:



- **Chylomicrons** – Transport triglycerides from the small intestine to the liver.
- **Very low-density lipoproteins** (VLDL) - transport newly synthesized triglycerides from the liver to the fatty tissues.
- **Intermediate-density lipoproteins** (IDL) are between VLDL and LDL in density. They are not usually seen in the blood.

- **Low-density lipoproteins** (LDL) - carry cholesterol from the liver to other tissues. They are also sometimes referred to as "**bad cholesterol**."
- **High-density lipoproteins** (HDL) - collect cholesterol from other tissues and return it to the liver. They are sometimes referred to as "**good cholesterol**."

Atherosclerosis and heart attack

Excessive fats and cholesterol in the body are associated with diabetes; cancers of the breast, pancreas, and colon; heart attack; and stroke. Lipoproteins transport cholesterol and fats in the body. LDLs transport cholesterol from the liver to the cells where they are needed, and HDLs transport surplus cholesterol back to the liver, where it is converted to other steroids and bile acids. If the LDL level is too high, some cholesterol is deposited in the artery as plagues, as shown in Figure 6.7.4. That is why LDLs are sometimes called **bad cholesterol**. The plaque causes inflammation and death of the tissue in that area and restricts blood flow. It may lead to a heart attack if the blood flow to the heart is inadequate or a stroke if the blood flow to the brain is insufficient.

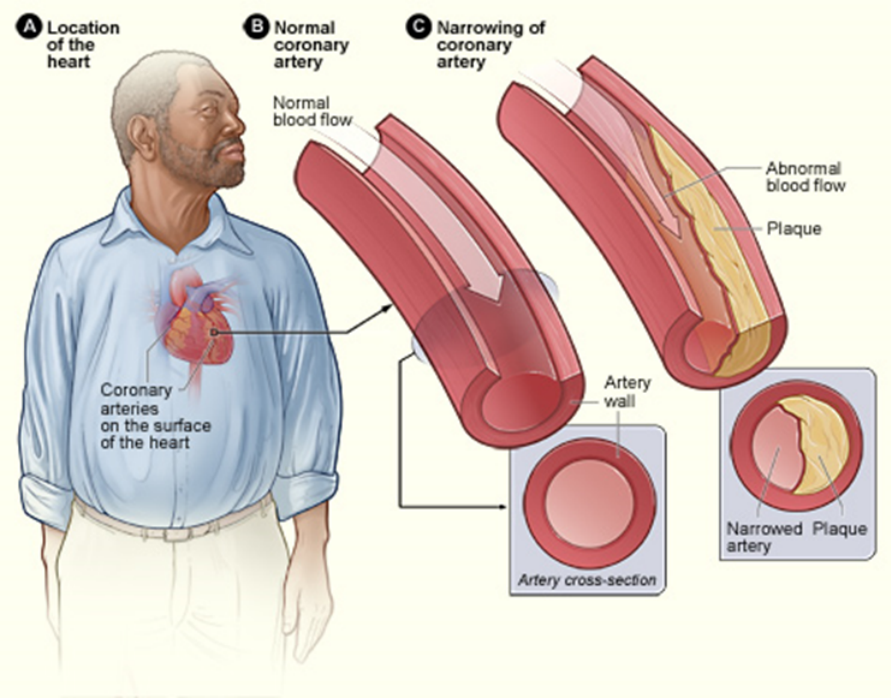


Figure 6.7.4: Figure A shows the location of the heart in the body. Figure B shows a normal coronary artery with normal blood flow. The inset image shows a cross-section of a normal coronary artery. Figure C shows a coronary artery narrowed by plaque. The buildup of plaque limits the flow of oxygen-rich blood through the artery. The inset image shows a cross-section of the plaque-narrowed artery. (Copyright; National Heart Lung and Blood Insitute (NIH), Public domain, via Wikimedia Commons)

Since HDLs transport the surplus cholesterol back to the liver for recycling and avoid plaque formation, HDLs are also called good cholesterol. The HDL levels must be sufficiently high to prevent a heart attack. A lipid panel test is prescribed to assess the risk of a heart attack. It measures the levels of total cholesterol, LDLs, HDLs, and triglycerides in the blood. The normal levels are the following.

Total cholesterol	About 150 mg/dL, less is better
LDL ("bad") cholesterol	About 100 mg/dL, less is better

HDL ("good") cholesterol	At least 40 mg/dL in men and 50 mg/dL in women; higher is better
Triglycerides	Less than 150 mg/dL, less is better

Bile salts

Bile salts are synthesized from cholesterol in the liver. First, additional $-OH$ groups are incorporated, and the alkyl side chain is shortened and oxidized to a carboxylic acid ($-COOH$) group. Then the $-COOH$ group is esterified with either glycine or taurine that imparts a negative charge group, as illustrated in Figure 6.7.5.

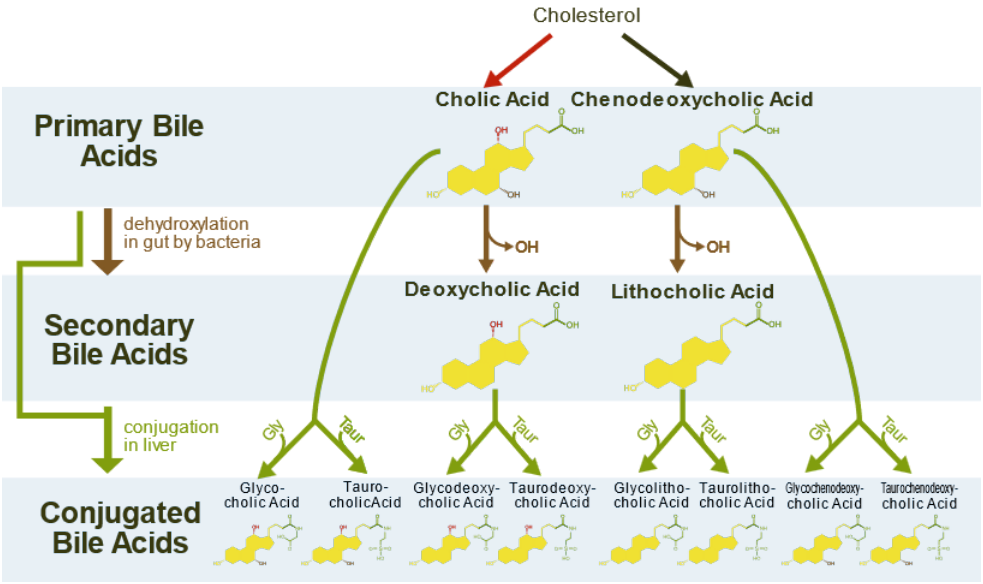


Figure 6.7.5: Synthesis of the 12 bile acids in human bile from cholesterol. (Copyright; Mcstrother, CC BY 3.0, via Wikimedia Commons)

Bile salts act like soaps with an ionic head and the remaining hydrophobic part. They are stored in the gallbladder and released into the small intestine, where they emulsify fats, i.e., convert the larger fat drops into smaller droplets, just like greases and dirt are emulsified by soaps. The smaller droplets have a larger surface area, increasing the fats' reaction rate with the pancreatic lipase enzymes, as illustrated in Figure 6.7.6.

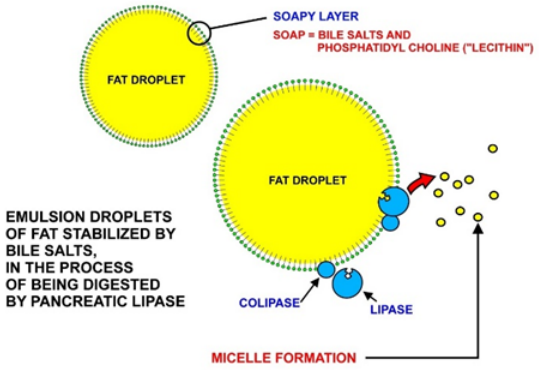


Figure 6.7.6: Dietary fats are emulsified in the duodenum by soaps in bile salts and phospholipids, which pancreatic lipase can attack. (Copyright; Cruithne9, CC BY-SA 4.0, via Wikimedia Commons)

Since bile salts are excreted with the feces, they remove cholesterol in two ways: 1) they are themselves the products of cholesterol, and 2) they dissolve cholesterol in the form of bile salt-cholesterol particles that are eliminated.

Gallstones and jaundice

When a large amount of cholesterol accumulates in the gallbladder, it turns into solid particles, i.e., gallstones. The gallstones are almost 100% cholesterol with a small amount of calcium salts, glycerophospholipids, and fatty acids. Small gallstones pass through the bile duct and enter the duodenum, i.e., the initial part of the small intestine, and are excreted. However, large gallstones may get stuck in the bile duct and cause pain, as illustrated in Figure 6.7.7. If a gallstone block the bile duct, the bile can not be excreted. Then the bile pigment, known as bilirubin, will back up into the liver and be passed via blood. It causes jaundice which gives a yellow color to the skin and whiteness to the eyes.

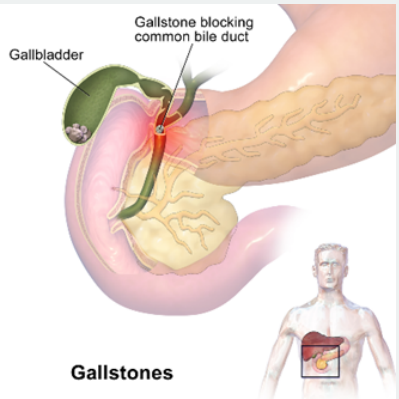


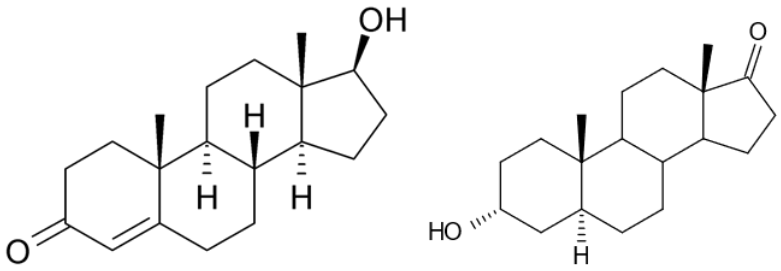
Figure 6.7.7: Illustration of gallstones blocking the bile duct that causes pain and jaundice. (Copyright; BruceBlaus, CC BY-SA 4.0, via Wikimedia Commons)

Steroid hormones

Steroid hormones have structural similarities with cholesterol and are derived from it. Two significant subclasses of steroid hormones are sex and adrenocortical hormones, described in the following sections.

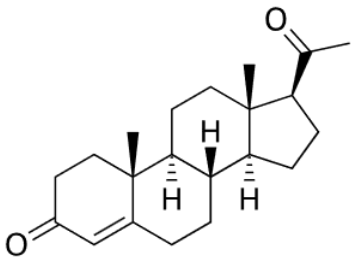
Sex hormones

The male sex hormones are called androgens which include testosterone and androsterone. The female sex hormones are estrogens, e.g., estradiol and estrone, and progestins, e.g., progesterone. These five of the most essential sex hormones are shown in the figure below. All five are present in males and females, but androgens are produced more in males, and estrogens and progestins are produced more in females.

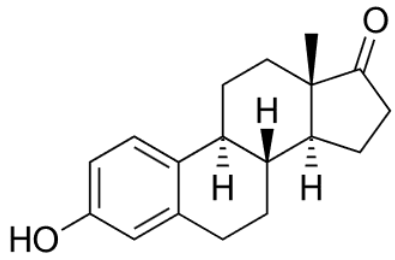


Testosterone Androsterone

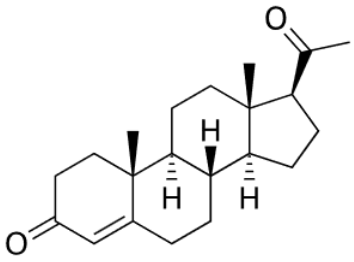
Male sex hormones (androgens)



Estradiol



Estrone



Progesterone

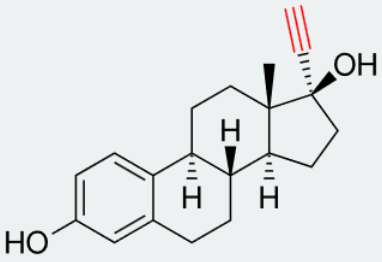
Female sex hormones

Functions of sex hormones

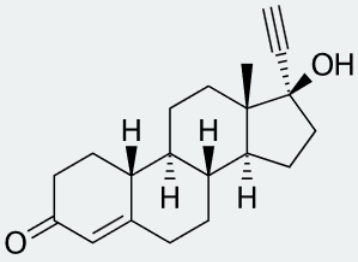
Testosterone and androsterone are the most potent male sex hormones. They control the development of male secondary sex characteristics, which include the growth of muscles, facial hair, maturation of sex organs, and sperms. Estrogens and progestins control the development of female secondary sex characteristics. Estrogens increase the size of the uterus, deposit fat in the breasts, and broaden the pelvis of females. Progestins prepare the uterus for nurturing the fertilized egg. If the egg is not fertilized, the levels of estrogens and progestins drop sharply, and menstruation begins.

Birth control

The release of estrogens and progestins during pregnancy prevents ovulation. This is mimicked by synthetic estrogens, such as ethynyl estradiol, and synthetic progestins, such as norethindrone, shown below, used in birth control formulations. As with the use of any steroid hormones, there are risks associated with synthetic hormones for birth control, including weight gain and a greater chance of forming blood clots.



Ethynyl estradiol



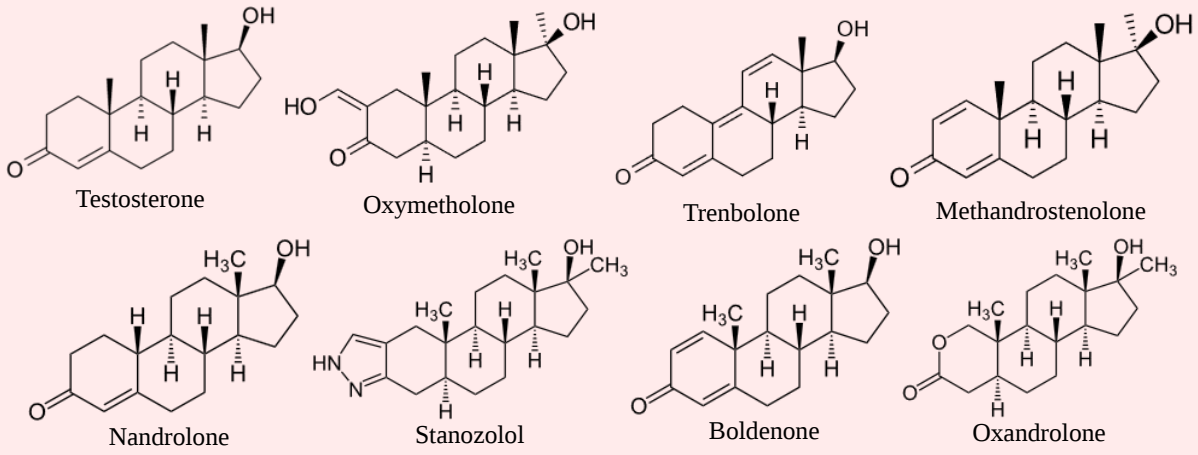
Norethindrone

Synthetic estrogens used in birth control

Anabolic steroids

One physiological function of testosterone is increasing muscle mass and decreasing body fat. Several synthetic equivalents, called anabolic steroids, have been developed for use as **appearance and performance-enhancing drugs**, as shown in the figure below. Although there are some benefits of anabolic steroids, their harmful effects are numerous and outweigh their benefits. They may not produce a euphoric high but develop substance use disorder. They can cause early heart attacks, strokes, kidney failure, liver tumors, make pimples pop up, hair fallout, grow breasts, and reduced testicles in males, grow beards in females, and cause extreme mood changes that may lead to committing suicide. Some of these harmful effects are

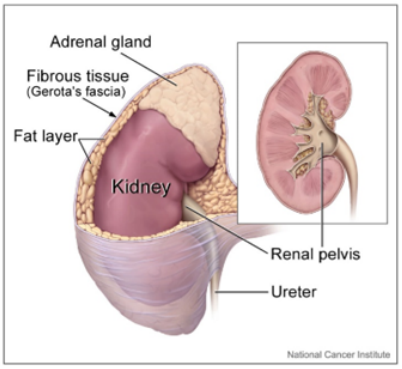
related to the changes in cholesterol levels, causing an increase in low-density lipoprotein (bad cholesterol), a decrease in high-density lipoprotein (good cholesterol), high blood pressure, and liver damage. The harmful effects are long-lasting and may be irreversible.



Anabolic steroids that are most commonly encountered by United States law enforcement

Adrenocortical hormones

Adrenocortical hormones are produced by adrenal glands on the top of the kidneys, as shown in the figure on the right (*Copyright; Alan Hoofring (Illustrator), Public domain, via Wikimedia Commons*). There are two major subclasses: **mineralocorticoids** which regulate Na^+ and K^+ ions, and **glucocorticoids** which control carbohydrate metabolism.



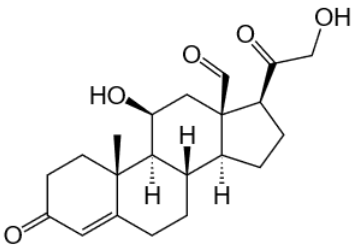
Mineralocorticoid

Aldosterone is an important mineralocorticoid that enhances the reabsorption of Na^+ and K^+ ions in the kidney tubules. The control of Na^+ and K^+ concentrations controls water absorption and swelling of tissues.

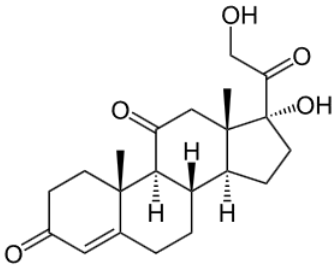
glucocorticoid

Cortisone is a glucocorticoid that increases blood glucose levels and glycogen synthesis in the liver. Cortisol is another glucocorticoid released under stress to increase blood glucose levels at the expense of the metabolism of carbohydrates, fats, and proteins.

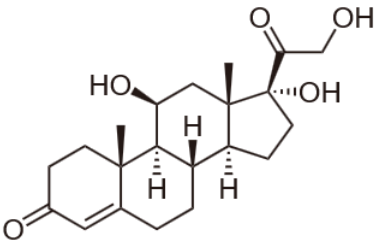
These hormones or their synthetic derivative, such as prednisone, shown below, are used to reduce inflammation, asthma, and arthritis. As with the use of steroids in general, health problems can result if these steroids are used for the long term.



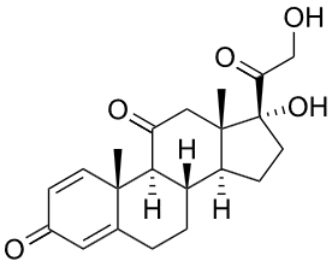
Aldosterone



Cortisone



Cortisol



Prednisone (a synthetic corticoid)

Adrenocortical hormones and their synthetic derivative

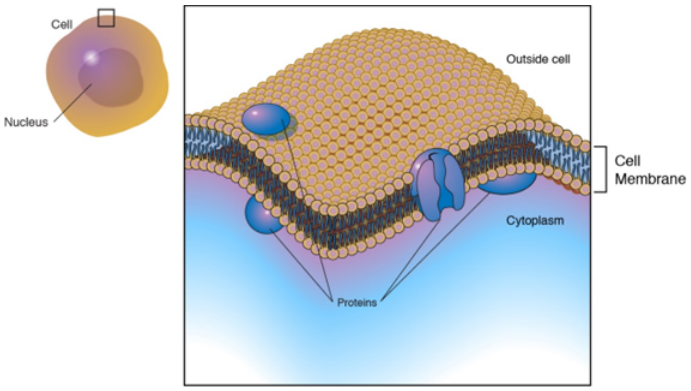
6.8: Cell membrane

Learning Objectives

- Define and understand the composition of the cell membrane and how the composition adjusts the fluidity fo the membrane.
- Understand three modes of transport through the cell membrane: diffusion, facilitated transport, and active transport.

What is a cell membrane?

The cell membrane, also known as the plasma membrane, cytoplasmic membrane, and plasmalemma, is a lipid bilayer with proteins dispersed in it that separates the cell interior from the extracellular space, as illustrated in the figure on the right (*Copyright; National Human Genome Research Institute, Public domain*).



Composition of the cell membrane

A cell membrane is a complex structure with several components, as shown in Figure 6.8.1. and described here.

- Phospholipid bilayer that has polar heads on the outside in contact with water and nonpolar tails inside the bilayer. Unsaturated fatty acids in the lipid hydrophobic tails increase the membrane fluidity -the more proportion of unsaturated fatty acids, the higher the fluidity.
- Cholesterol interspersed between phospholipids controls the rigidity of the membrane -the more the proportion of cholesterol, the more rigid the membrane.
- Two types of proteins: integral proteins that span the membrane and serve as transporters of species, and peripheral proteins that are loosely attached to the outer side of the membrane that act as enzymes to facilitate the interaction with the cell's environment.
- Glycoproteins and glycolipids have carbohydrate oligomers attached to the outer lipid layer and serve the purpose of cell-to-cell recognition.

The cell membrane controls the movement of substances in and out of the cells and organelles. It is selectively permeable to ions and organic molecules. It plays a role in cell adhesion, cell signaling, and attachment surface for the cytoskeleton to shape the cells and attach to the extracellular matrix to hold them together in tissues.

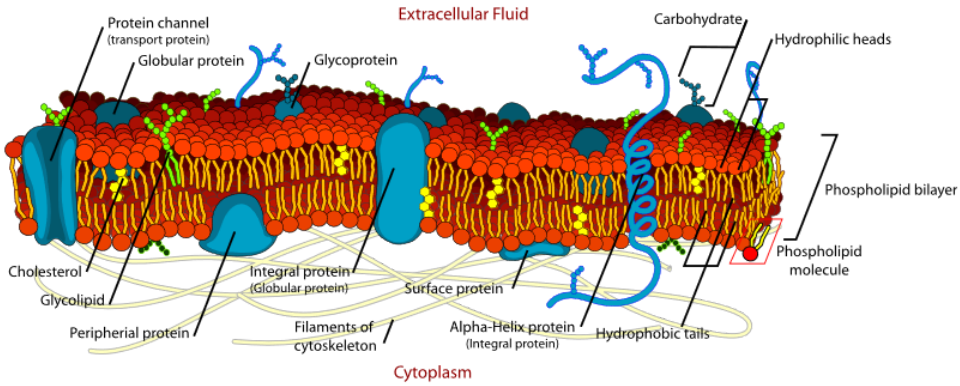


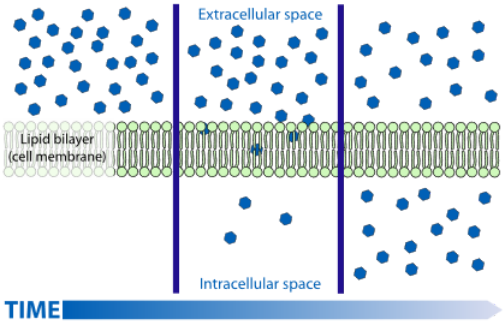
Figure 6.8.1: A detailed cell membrane structure. (Copyright; LadyofHats, Public domain, via Wikimedia Commons)

Transport through the cell membrane

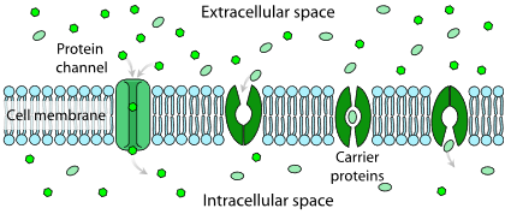
The cell membrane is a partition between intracellular and extracellular spaces. Still, some substances needed by the cell need to enter, and some products or wastes need to exit the cell. The cell membrane allows a selective movement of substances in and out of the cell in several ways.

Diffusion (passive) transport

The molecules in the lipid bilayer are vibrating due to thermal energy. Therefore, some molecules, such as O_2, CO_2, urea, water, etc., can move across the membrane from a higher concentration region to a lower concentration region through the process of diffusion, as illustrated in Figure $6.8.2$ left.



Diffusion or passive transport



Facilitated transport through protein channels or carrier protein

Figure 6.8.2: Illustration of diffusion and facilitated transport across the cell membrane. (Copyright; LadyofHats, Public domain, via Wikimedia Commons)

Facilitated transport

Integral proteins form channels through which certain substances can diffuse more rapidly than by simple diffusion. The proteins have a channel size that matches the substance's size or changes the shape to adjust to the size of the substance that needs to be selectively transported through the facilitated transport, as illustrated in Figure $6.8.2$ right. Particularly, water-soluble substances such as chloride ion (Cl^-), bicarbonate ion HCO_3^-, and glucose molecules do not move fast enough through simple diffusion and are transported by the facilitate transport process to meet the need of the cells.

Active transport

Sometimes substances need to be moved against the concentration gradient. It takes place at the expense of energy in the form of ATP, just like pumping substances at the cost of electricity in everyday life. For example, K^+ concentration is greater inside the cell, and that of Na^+ is greater outside the cell. In the conduction of nerve impulses and contraction of muscles, K^+ moves into the cell, and Na^+ moves out of the cell by active transport process, as illustrated in Figure $6.8.3$.

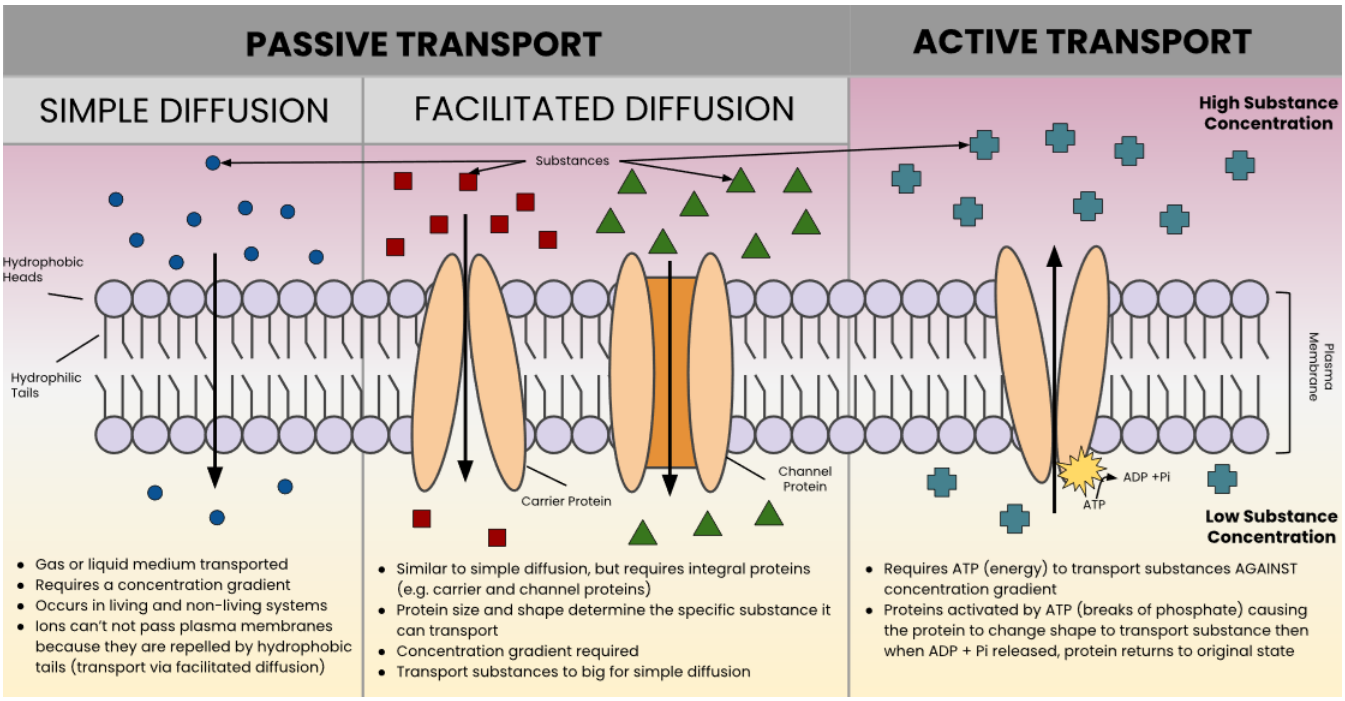


Figure 6.8.3: Comparison of membrane transport methods: passive transport, which includes simple and facilitated diffusion, and active transport, which takes place at the expense of energy. (Copyright; LSumi, CC BY-SA 4.0, via Wikimedia Commons)

CHAPTER OVERVIEW

7: Proteins

7.1: What are proteins?

Learning Objectives

- Define proteins and understand their classification and the importance of their structure in their functions.

Proteins

Proteins are bio-polymers containing one or more polymer chains composed of amino acid monomers linked together by amide bonds, i.e., proteins are polyamide biochemicals.

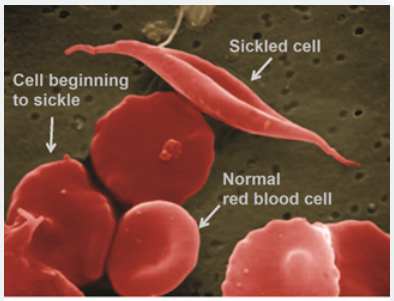
Classification of proteins based on their functions

Proteins are considered the most abundant biochemicals that perform various functions in living things. The major types of proteins, classified according to their role in biochemical systems, are the following.

- **Structural**: Structural proteins are the primary structural materials in animals. For example, collagen constitutes tendons and cartilage, and keratin comprises hair, skin, wool, and nails.
- **Contractile**: These proteins are responsible for movements; e.g., myosin and actin are proteins responsible for the contraction of muscles.
- **Transport**: These proteins move the molecules in the body. For example, hemoglobin transports oxygen, and lipoproteins carry lipids.
- **Storage**: They store nutrients or essential chemicals, e.g., casein in milk and ovalbumin in eggs store nutrients, and ferritin stores iron in the spleen and liver.
- **Hormones**: These proteins regulate metabolism and the nervous system; e.g., insulin regulates blood glucose levels, and growth hormones regulate growth.
- **Enzymes**: They catalyze biochemical reactions in the body of living things, e.g., sucrase catalyzes the hydrolysis of sucrose, and digestive enzymes like trypsin catalyze the hydrolysis of proteins.
- **Protection**: These proteins protect living things, e.g., antibodies destroy harmful foreign substances, and fibrinogen coagulates blood when needed to avoid blood loss through injuries.

Importance of protein structure

Many proteins are present in living things' bodies to perform different functions. For example, the human body contains about 100,000 different proteins. The structure of proteins, i.e., the exact sequence of amino acids and how they fold and interact with other protein molecules in the physiological environment, is critically important for their functions. For example, hemoglobin in red blood cells is needed to transport oxygen. It comprises more than 800 amino acids arranged in a specific sequence. Sickle cell anemia is caused by a change of one out of more than 800 amino acids in hemoglobin. It causes red blood cells to change from regular rounded shapes to sickle shapes, as shown in the figure to the right, which can not function properly and causes other medical problems.



Copyright; Modified from: NHLBI, Public domain, via Wikimedia Commons

To understand the proteins, their functions, and their malfunction, it is essential to understand the structure and properties of amino acids and proteins described in the next section. Proteins have levels of structures, i.e., i) primary -the sequence of amino acids, ii) secondary -the folding of sections of the protein chains, iii) tertiary -the overall shape of the protein polymer, and iv) quaternary -a combination of more than one proteins in a unit which is described in a later section.

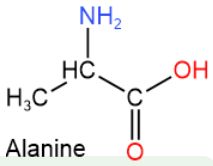
7.2: Amino acids

Learning Objectives

- Define and distinguish amino acids, α-amino acids, and proteinogenic amino acids.
- Draw Fisher projections and assign D/L or R/S stereodescriptors to proteinogenic amino acids.
- Understand the classification of proteinogenic amino acids based on the characteristics of the side chain.
- Define isoelectric point and understand the ionization states of amino acids under physiological conditions.
- Define essential amino acids and their sources.

What are amino acids?

Amino acids are the building blocks of proteins, i.e., they are the monomers of proteins. Amino acids are organic compounds that contain both an amine ($-NH_2$) and a carboxylic acid ($-COOH$) group in the same molecule. For example, alanine, shown on the right, is an amino acid. Proteins contain a subclass of amino acids called α-Amino acids.



Alanine

Alpha (α)-amino acids

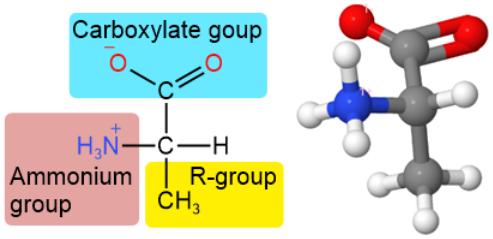
α-Amino acids have an amine ($-NH_2$) on the αC to $-COOH$ group, i.e., both the amine ($-NH_2$) group and the carboxylic acid ($-COOH$) group are attached to the same C. For example, alanine is an α-amino acid.

Proteinogenic amino acids

Proteinogenic amino acids are a subclass of α-amino acids incorporated into proteins during biosynthesis. Twenty proteinogenic amino acids are usually present in proteins, and two additional are included in exceptional cases. This chapter's word "amino acid" refers to the 20 standard proteinogenic amino acids.

Configuration of αC of an α-amino acids

the αC of α-amino acid have four different groups attached to it: amine ($-NH_2$), carboxylic acid (($-COOH$), hydrogen ($-H$), and alkyl side chain ($-R$, as shown in Figure 7.2.1. There is one exception "glycine" that has two H's at the αC



Fisher projection

Model

Figure 7.2.1: Fisher projection and model of an amino acid shown using alanine as an example. (Copyright; Public domain)

Three groups, i.e., $-NH_2$, ($-COOH$, and $-H$ are present in all α-amino acids, and the fourth group, i.e., the side chain ($-R$) varies in different α-amino acids.

The $-COOH$ is acidic (pK_a 2-5) and the $-NH_2$ is basic (pK_a ~10). The $-COOH$ loses its proton and exists as a carboxylate ion ($-COO^-$; the $-NH_2$ gains a proton and exists as ammonium ion ($-NH_3^+$ at physiological pH of 7.4. These groups ionize the same way if present in the side chain in addition to the αC.

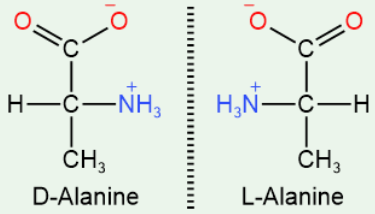
Zwitterions

Compounds that have a positive charge on one atom and a negative charge on another atom in the same molecule are called **zwitterions**. Most of the α-amino acids usually exist as zwitterions at physiological pH, as shown in Figure 7.2.1.

A C with four different groups is a chiral center. αC of α-amino acids is a chair center at αC, with one exception, "glycine" that has two H′s at the αC. α-Amino acids are presented in Fisher projections with the C-chain placed vertically and the $-NH_2$ and $-H$ bonds shown horizontally, as shown in Figure 7.2.1.

D- and L-configurations of α-amino acids

The α-amino acids that have their $-NH_2$ group placed to the left side of their Fisher projections are called D-amino acids, and those having $-NH_2$ group placed to the left side are called L-amino acids. The D- and the corresponding L-amino acids are enantiomers, as shown in the figure below for the case of alanine.



Proteinogenic amino acids are L-amino acids, with the exception of glycine, which is not chiral. In the R/S system, proteinogenic amino acids are (*S*) at the αC, with the exception of "cysteine" being (*R*) and glycine not chiral.

The α-amino acids are known by their common names, which are also abbreviated in three letters and one letter, as shown in Figure 7.2.2. For example, Alanine is abbreviated as Ala in three letters or as A in one letter form. The proteinogenic amino acids are classified based on the nature of the side chain.

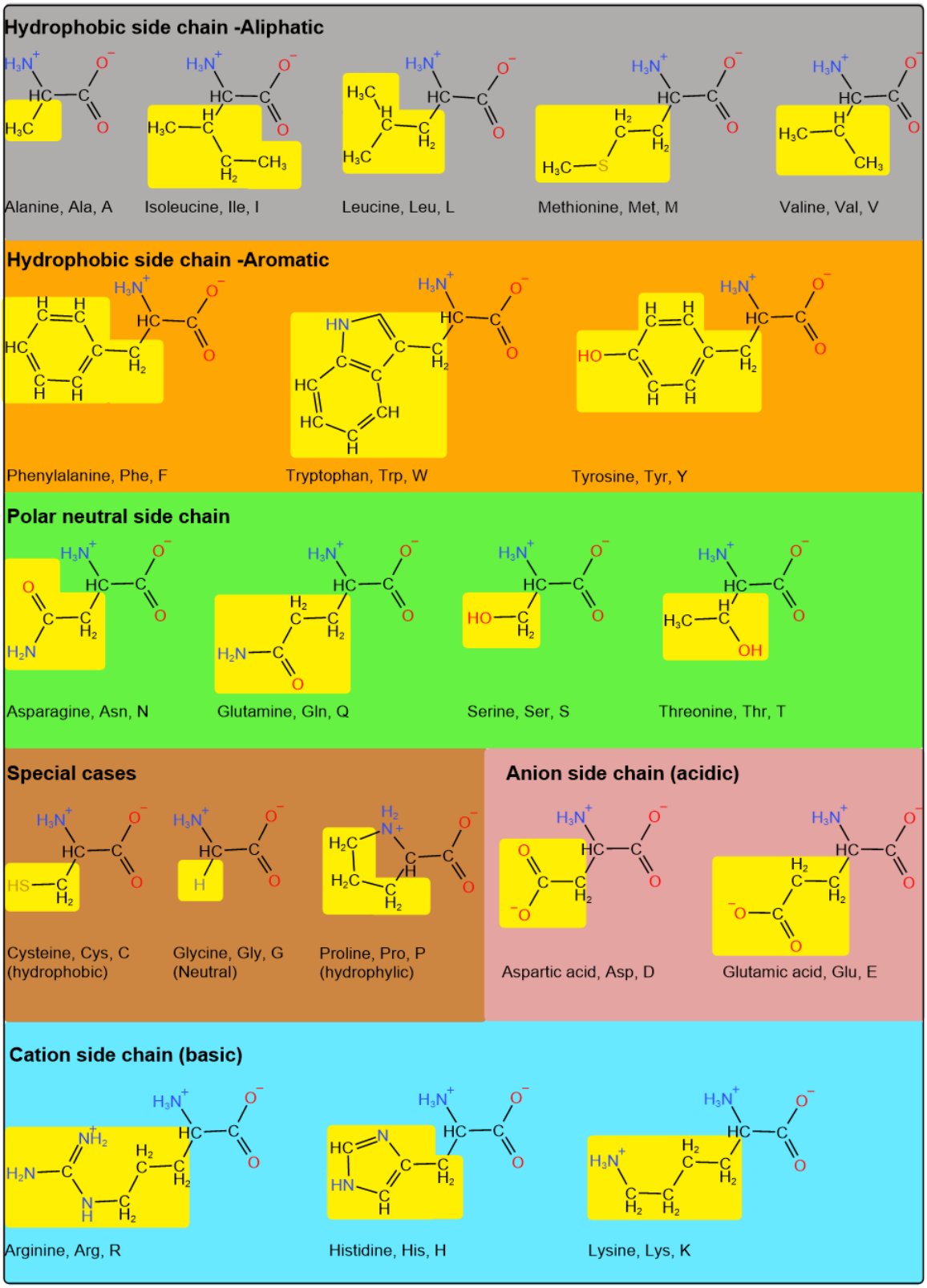


Figure 7.2.2: Structures of proteinogenic amino acids are shown in their zwitterionic forms with side chain protonation states determined at physiological pH of 7.4. Names, three-letters, and one-letter abbreviations are also shown. The alkyl side chains are highlighted in yellow color boxes. (Copyright; Public domain)

Classification of proteinogenic amino acids

α-Amino acids are classified based on the hydrophobicity of their side chain.

Hydrophobic and hydrophilic

Hydrophobic means "water fearing", i.e., molecules or entities that tend to repel water, not dissolve, or not wetted by water.
Hydrophilic means "water-loving", i.e., molecules or entities that tend to mix with, dissolve, or be wetted by water.

Hydrocarbons and other non-polar compounds are hydrophobic compounds that do not dissolve in water. Polar or ionic compounds are usually soluble in water. Based on these criteria, amino acids are classified into the following classes, shown in Figure 7.2.2.

- Nonpolar amino acids with an **aliphatic side chain** are hydrophobic. Methionine having a thioether group in the side chain is non-polar and placed in this group.
- Nonpolar amino acids with an **aromatic side chain** are hydrophobic. Tyrosine has phenol as part of its side chain, but it is hydrophobic and placed in this group.
- Polar neutral amino acids with a **polar neutral side chain** are hydrophilic. This class of side chain contains alcohol ($-\mathrm{OH}$) or amide $-\mathrm{CONH_2}$ groups in their side chain that hydrogen bond with water but do not ionize.
- Amino acids in a special class include cysteine, glycine, and proline.
  - Speciality of cysteine is that its thiol {$ce{-SH}$ group is easily oxidized forming is a disulfide ($\mathrm{S-S}$ bond, which is the only covalent bond in protein besides the amide bonds. Cystein is classified as nonpolar and hydrophobic because {$ce{-SH}$ is a nonpolar group.
  - Glycine is the only proteinogenic acid that has no chiral center. The $-\mathrm{H}$ side chain places it at a borderline between hydrophilic and hydrophobic categories, it is considered neutral.
  - Proline is the only amino acid with a secondary α-amine group. The side chain is a five member ring with N of α-amine as part of the ring. It is classified as a hydrophilic. The ring structure puts restrictions on the allowed configurations when it is incorporated in proteins. It creates a bent or kink in the protein backbone structure.
- Acidic amino acids with an **acidic side chain** are hydrophilic. These amino acids has carboxylic acid ($-\mathrm{COOH}$) group in their side chain that ionize to anion $-\mathrm{COO^-}$ under physiological conditions.
- Basic amino acids with a **basic side chain** are hydrophilic. These amino acids have basic primary or secondary amine groups in their side chain that ionize to cation $-\mathrm{NH_3^+}$ or $-\mathrm{NRH_2^+}$. Histidine has an amine group that has pKa 6.04 and is not ionized at pH 7.4, but it is placed in this group as it ionizes at pH below 6.0.

Acid-base nature of α-amino acids

The pKa is a measure of the strength of an acid, i.e., the lower the $\mathrm{pK_a}$ stronger the acid. Amino acids have $-\mathrm{COOH}$ group that is acidic with $\mathrm{pK_a}$ 2-3 and $-\mathrm{NH_2}$ on adjacent C that is basic with $\mathrm{pK_a}$ ~40. Conjugate acid of $-\mathrm{NH_2}$, i.e., $-\mathrm{NH_3^+}$ has $\mathrm{pK_a}$ ~10. Acids lose their proton when they are in a medium with a pH higher than the $\mathrm{pK_a}$ of the acid. For example $-\mathrm{COOH}$ exist as $-\mathrm{COO^-}$ in physiological medium with pH ~7.4. Bases gain protons in a medium with a pH lower than the $\mathrm{pK_a}$ of their conjugate acid. For example, $-\mathrm{NH_2}$ exist as $-\mathrm{NH_3^+}$ in physiological medium with pH ~7.4. Amino acids exist as zwitterions, i.e., have both cation $-\mathrm{NH_3^+}$ and anion $-\mathrm{COO^-}$ in the same molecule. Some amino acids have an additional acid or base group in their side chain that also ionizes depending on the pH of the medium.

The gain or loss of protons is an equilibrium process. In a strongly acidic medium, basic groups gain more protons than the protons lost by their acid groups. So, the amino acids have an overall positive charge in a strong acid medium. In a strongly basic medium basic groups gain fewer protons than those lost by their acid groups. So, the amino acids have an overall negative charge in a strongly basic medium. At a certain pH in the middle, an amino acid has an equal positive and negative charge and is neutral overall.

Isoelectric point (pI)

An amino acid's i**soelectric point (pI)** is the pH at which it has equal positive and negative charges and carries no net charge, i.e., it is neutral overall.

The pK_a value of $-COOH$ and $-NH_3^+$ on the α-C, pK_a values of acidic or basic groups in the side chain, and pI values of 20 α-amino acids found in proteins are listed in Table 1 below.

Table 1: Names, three-letters, and one-letter abbreviations, pK_a values of alpha-carboxylic acid and ammonium groups, pK_a values of acid groups in the side chain, and isoelectric point (pI) of 20 amino acids found in proteins. (Reference: D.R. Lide, *Handbook of Chemistry and Physics, 72nd Edition*, CRC Press, Boca Raton, FL, 1991)

Amino acid	Three letters abbreviations	One letter abbreviations	pK_a of $\alpha-COOH$	pK_a of $\alpha-NH_3^+$	pK_a of side chain group	Isoelectric point (pI)
Alanine	Ala	A	2.34	9.69	–	6.00
Arginine	Arg	R	2.17	9.04	12.48	10.76
Asparagine	Asn	N	2.02	8.80	–	5.41
Aspartic acid	Asp	D	1.88	9.60	3.65	2.77
Cysteine	Cys	C	1.96	10.28	8.18	5.07
Glutamic acid	Glu	E	2.19	9.67	4.25	3.22
Glutamine	Gln	Q	2.17	9.13	–	5.65
Glycine	Gly	G	2.34	9.60	–	5.97
Histidine	His	H	1.82	9.17	6.00	7.59
Isoleucine	Ile	I	2.36	9.60	–	6.02
Leucine	Leu	L	2.36	9.60	–	5.98
Lysine	Lys	K	2.18	8.95	10.53	9.74
Methionine	Met	M	2.28	9.21	–	5.74
Phenylalanine	Phe	F	1.83	9.13	–	5.48
Proline	Pro	P	1.99	10.60	–	6.30
Serine	Ser	S	2.21	9.15	–	5.68
Threonine	Thr	T	2.09	9.10	–	5.60
Tryptophan	Trp	W	2.83	9.39	–	5.89
Tyrosine	Tyr	Y	2.20	9.11	10.07	5.66
Valine	Val	V	2.32	9.62	–	5.96

Essential amino acids

Nine amino acids are essential because humans can not synthesize them fast enough to meet their demands.

> The **essential amino acids** are valine, isoleucine, leucine, methionine, phenylalanine, tryptophan, threonine, histidine, and lysine.

The essential amino acids are obtained from foods. Foods from animal sources, e.g., eggs, milk, fish, meat, etc., are complete foods with all the essential amino acids. Foods from plant sources, e.g., wheat, rice, corn, etc., are usually deficient in one or more essential amino acids. So, vegetarians have to eat various vegetarian foods to obtain all the essential amino acids.

Deficiency of essential amino acids in different foods

- Wheat, rice, and oats are deficient in lysine.
- Corn is deficient in lysine and tryptophan.
- Soy is deficient in methionine.
- Beans are deficient in methionine and tryptophan.
- Peas and peanuts are deficient in methionine
- Almonds and walnuts are deficient in lysine, tryptophan
- Foods from animal sources, e.g., milk, eggs, meat, fish, etc., have all the essential amino acids.

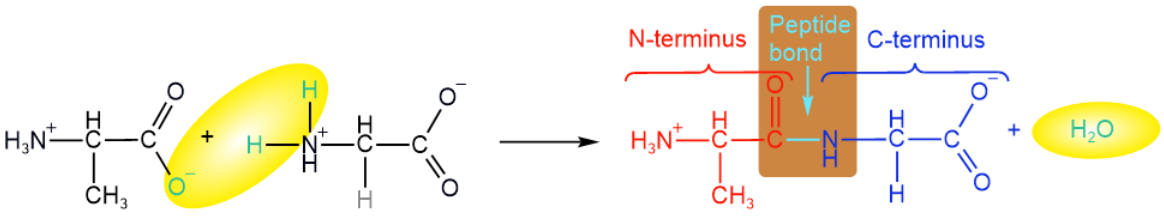
7.3: Primary structure of proteins

Learning Objectives

- Understand a peptide bond and disulfide, their nomenclature, and their characteristics.
- Define and write proteins' primary structure, importance, and related terminologies.

Peptide

An amine reacts with a carboxylic acid and makes an amide by eliminating water. Amino acids have both an amine and a carboxylic acid group. If the amine group of one amino acid reacts with the carboxylic acid of another, the two amino acids become bonded through an amide bond. For example, alanine and glycine become bonded through an amide bond, as illustrated below.

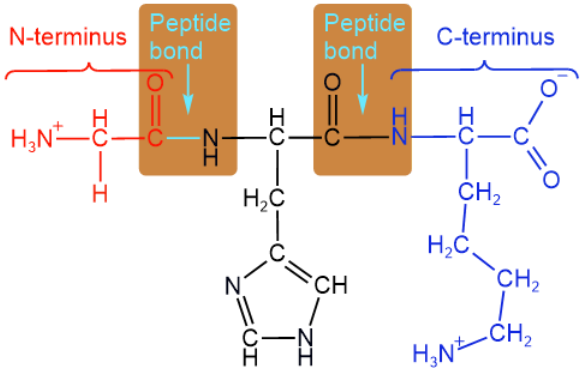


Alanine, Ala, A Glycine, Gly, G Alanylglycine, Ala-Gly, AG

Peptide bond

An amide bond that links two amino acids is called a **peptide bond** or **peptide linkage**. For example, a peptide bond that links alanine and glycine is highlighted in a brown box in the above illustration.

A single amino acid is also called a **monopeptide**, and two amino acids linked by a peptide bond is called a **dipeptide**. Two amino acids, alanine, and glycine, are monopeptides; their product, alanylglycine, shown above, is a dipeptide. One amino acid in a dipeptide has free ammonium ($-NH_3^+$) group that can make a peptide with $-COO^-$ of another amino acid. Similarly, one amino acid in a dipeptide has a free carboxylate ($-COO^-$) group that can make a peptide bond with $-NH_3^+$ group of another amino acid. Three amino acids linked by peptide bonds is called a **tripeptide**, four are called **tetrapeptide**, five are **pentapeptide**, six are **hexapeptide**, seven are **heptapeptide**, and so on. A tripeptide of glycine, histidine, and lysine is shown below.

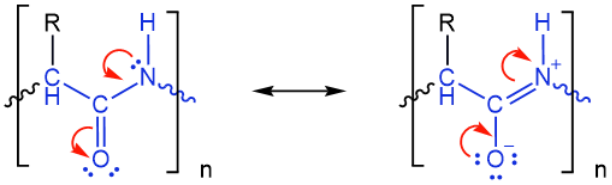


Glycylhistidyllysine, Gly-His-Lys, GHL

Shorter chains of amino acids linked by peptide bonds are called **peptides;** longer ones are called **polypeptides**. Polypeptides and **proteins** are used interchangeably, but more than 50 amino acid chains are usually called proteins. Amino acids in peptides are usually called **residues**.

Structure of a peptide backbone

The backbone of a peptide chain is $-C-C-N-$ where the middle C is the carbonly $C=O$ and $C-N$ is the peptide bond. The peptide bond has two resonance contributors, as shown below.



Due to the resonance, the peptide bond has about 40% double bond character. There is no free rotation around the peptide bond due to its double bond character. Therefore, the four groups around the peptide bond (shown in blue color in the above structure) exist in the same plane as in the case of alkenes. Free rotation is around the other σ-bonds in the peptide backbone. Therefore, the structure of the peptide backbone is like cards connected by a swivel at the opposite corners, as shown in Figure 7.3.1.

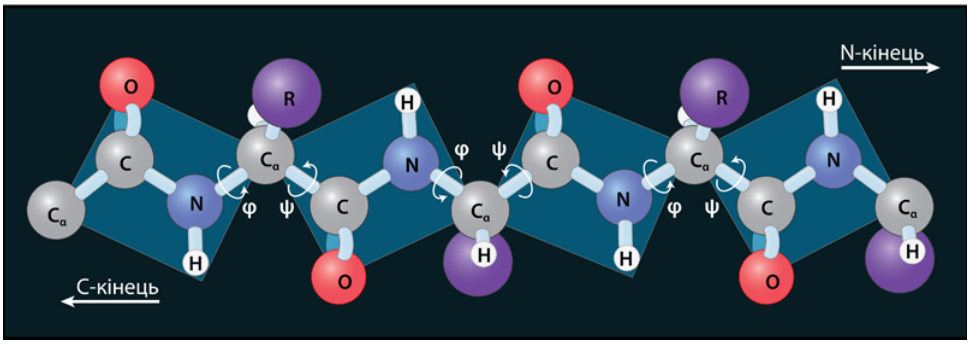
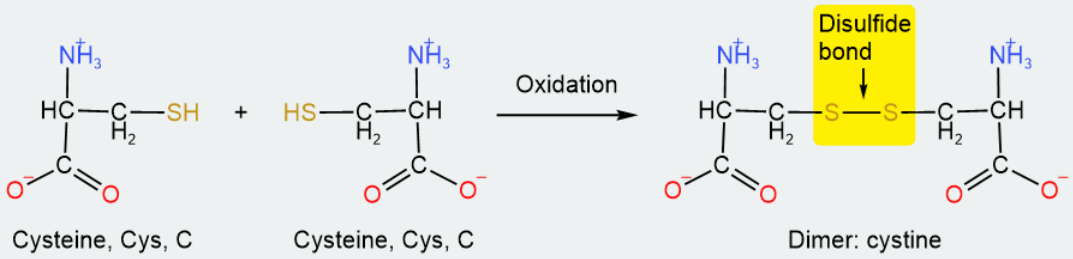


Figure 7.3.1: Illustration of planer structure around peptide bonds as card-like structures and free rotation around other σ-bonds shown by arrows. (Copyright; Zlir'a, CC0, via Wikimedia Commons)

The peptide bond's rigidity limits the peptide backbone's possible orientations, affecting its secondary and tertiary structure. The $\mathrm{N-H}$ groups can establish hydrogen bonds with the $\mathrm{C=O}$ groups within the same chain or between the neighboring chains that play an essential role in determining the secondary and tertiary structures of proteins that are described in a later section.

Disulfide bond

Amino acid cysteine has a thiol ($\mathrm{-SH}$) group that can easily be oxidized to **disulfide** ($\mathrm{-S-S-}$) bond or **disulfide linkage** linking two cysteines into a dimer called cystine, as illustrated below.



When a cysteine residue makes a disulfide bond with another cysteine residue in the same chain or another chain, it provides a covalent linkage that binds parts of the same chain or two different chains, as shown in Figure 7.3.2. Examples of both types are found in the structure of insulin, which is composed of two polypeptides joined by sulfide linkage, as shown in Figure 7.3.2.

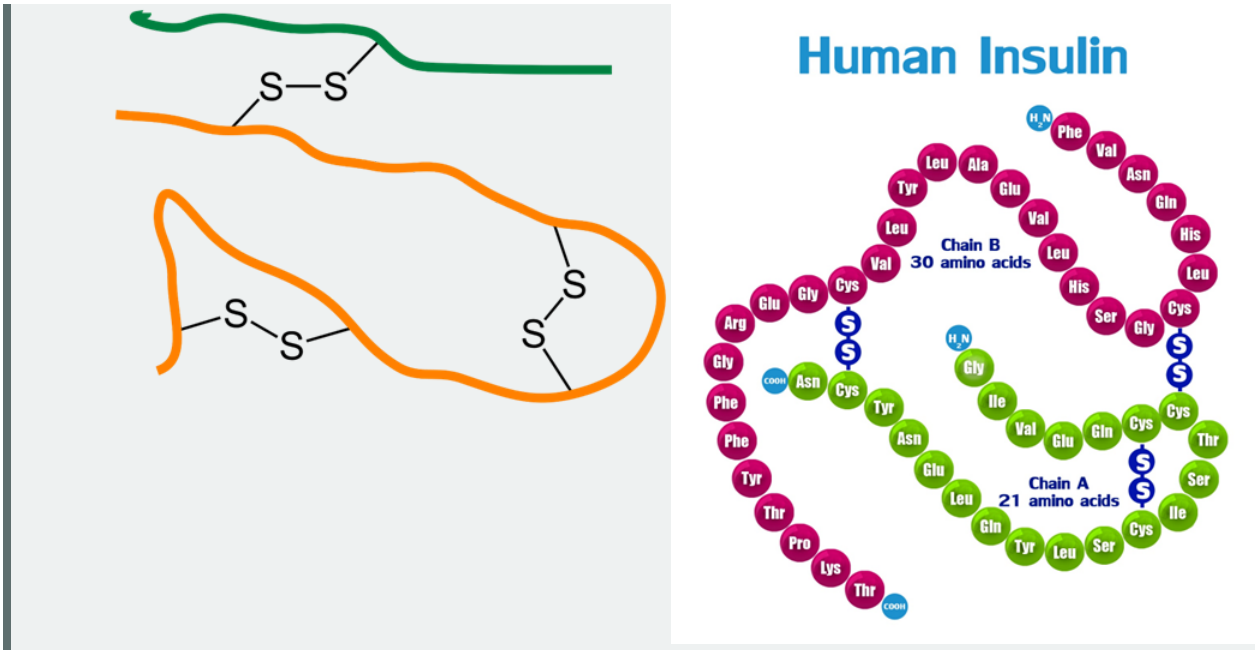


Figure 7.3.2: Illustration of disulfide linkage within the same chain and between different chains of polypeptides (left) and with the example of human insulin (right). (Copyright; Left: Jü, CC0, via Wikimedia Commons, right: Zappys Technology Solutions, CC BY 2.0, via Wikimedia Commons)

Naming peptides

When amino acids combine by peptide bonds, one amino acid on one terminal has free ammonium ($-NH_3^+$) group, an amino acid on the other terminal has free carboxylate ($-COO^-$) group, as highlighted in the structures of dipeptide and tripeptides shown above.

N-Terminus and C-terminus of a peptide

- An amino acid in a peptide that has free ammonium ($-NH_3^+$) group is called **N-terminus**. For example, glycine in the tripeptide shown above is N-terminus.
- An amino acid that has a free carboxylate ($-COO^-$) group is called **C-terminus**. For example, lysine in the tripeptide shown above is C-terminus.

Amino acids in a peptide are written horizontally from left to right, where N-terminus is the leftmost amino acid, and C-terminus is the rightmost amino acid.

A **peptide is named** by listing the names of its constituent amino acids in a sequence from N-terminus to C-terminus, with the last syllable changed to yl, except for the C-terminus. For example, the dipeptide of alanine and glycine is alan**yl**glycine, and the tripeptide of glycine, histidine, and lysine is glyc**yl**histid**yl**lysine.

Often three-letter abbreviations of the amino acids in a peptide are written in a sequence from N-terminus to C-terminus, separated by hyphens. For example, the dipeptide alan**yl**glycine can be written as Ala-Gly, and the tripeptide glyc**yl**histid**yl**lysine as Gly-His-Lys. For polypeptides, one-letter abbreviations of the amino acid residues are usually written in a sequence from N-terminus to C-terminus. For example, dipeptide Ala-Gly is AG, and tripeptide Gly-His-Lys is GHL.

What is the primary structure of proteins?

Primary structure of proteins

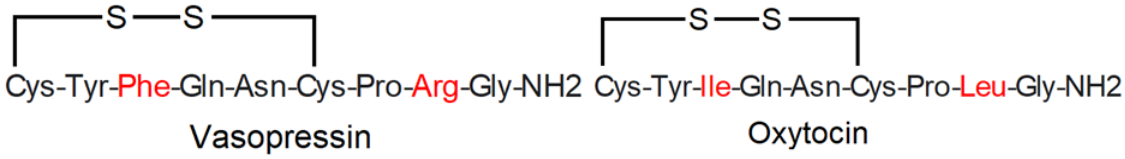
The primary structure of peptides or proteins is the sequence of amino acids linked together by peptide bonds. For example, the primary structures of the dipeptide and tripeptides shown above are Ala-Gly and Gly-His-Lys.

The primary structure of a protein is shown as a sequence of amino acids written from the N-terminus to the C-terminus. When the sequence of amino acids is known, three-letter abbreviations are separated by hyphens, e.g., Gly-His-Lys. When the sequence of amino acids in a peptide is not known, the three-letter abbreviations of the constituent amino acid are listed, separated by commas. For example, Ala, Gly could mean Ala-Gly or Gly-Ala, which are different compounds with different properties related to each other as constitutional isomers. Similarly, Gly, His, Lys could mean any one of the following six constitutional isomers: Gly-His-Lys, Gly-Lys-His, His-Lys-Gly, His-Gly-Lys, Lys-Gly-His, or Lys-His-Gly.

The number of constitutional isomers increases exponentially as the number of amino acids in the peptide increases. Constitutional isomers of a polypeptide of n amino acids chosen from 20 amino acids commonly found in proteins are given by 20^n. For example, a polypeptide containing 60 amino acids selected from 20 amino acids found in proteins may have 20^{60}, i.e., 10^{78}, which is an enormous number of possibilities. This analysis shows that the number of proteins that can be synthesized using 20 amino acids is enormously large. An analog is the entire English language composed of letters from 26 different alphabets.

When the primary structure of a polypeptide is modified, its function is affected. The extent of the effect depends on the number of amino acids replaced and their nature. For example, human insulin comprises two peptides: chain A of 21 amino acids and chain B of 30 amino acids, two chains joined by disulfide linkages, as shown in Figure 7.3.1. The amino acids at positions 8, 9, and 10 in chain A (-Thr-Ser-Ile-) and position 30 in chain B (-Thr) in human insulin are replaced with -Ala-Ser-Val- and -Ala, respectively, in bovine insulin, but the two perform the same function. Humans can use bovine insulin, though it is less effective in humans and sometimes causes an allergic reaction.

As shown below, vasopressin and oxytocin are two nonapeptides that differ in two amino acid residues at positions 3 and 8. Their cysteine residues form a disulfide bond, and the carboxylate group ($-\mathrm{COO}^-$) on C-terminus is converted to a primary amide (($-\mathrm{CONH}_2$).

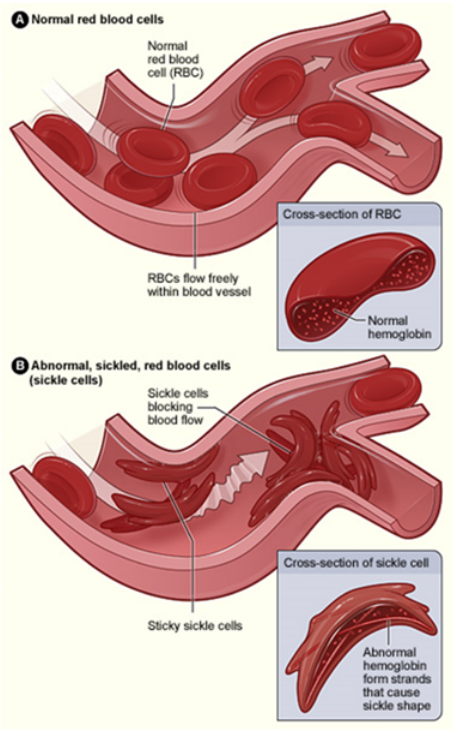


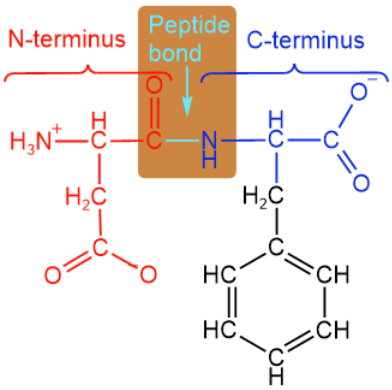
Both vasopressin and oxytocin are hormones the pituitary gland produces but have different functions. Vasopressin is an antidiuretic hormone that regulates blood pressure by adjusting the water reabsorbed by the kidneys. Oxytocin stimulates uterine contractions in labor.

Another example where a slight change in the primary structure of a protein alters its function significantly is sickle cell anemia. Hemoglobin in the red blood cell is responsible for carrying oxygen. Hemoglobin comprises four polypeptides, two alpha chains, and two beta chains containing 574 amino acid residues. A change of glutamic acid (a hydrophilic amino acid) with valine (a hydrophobic amino acid) in the sixth position of the two beta-chains changes the structure of hemoglobin so much that it causes the red blood cells to change from a rounded shape to a sickle body. Sickled cells do not function properly and block the blood flow -a medical condition called sickle cell anemia, illustrated in The Figure on the right.

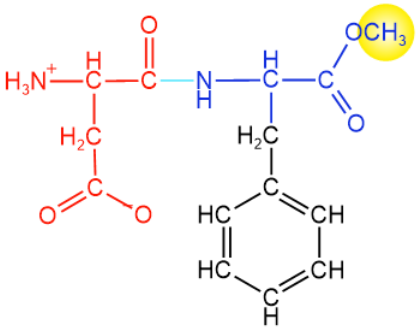
Some examples and uses of peptides

Peptides have different functions in the body of living things. Some of them are commercially important also. For example, dipeptide Ala-Gly shown is a dietary supplement. Aspartame, 200 times sweeter than sucrose and used as a sugar substitute, is a methyl ester of a dipeptide Asp-Phe, shown below.





Aspartylphenyalanine, Asp-Phe, DF



Asparame

The tripeptide Gly-His-Lys shown before is a human copper-binding peptide with wound healing and skin remodeling activity. Met-enkephalin (Tyr-Gly-Gly-Phe-Met) is a pentapeptide related to pain signals in the body.

Vasopressin and oxytocin are two nonapeptide hormones produced by the pituitary gland. Vasopressin is an antidiuretic hormone that regulates blood pressure by adjusting the water reabsorbed by the kidneys. Oxytocin stimulates uterine contractions in labor.

Substance P (SP) is an undecapeptide (Arg-Pro-Lys-Pro-Gln-Gln-Phe-Phe-Gly-Leu-Met) that is a neuropeptide acting as a neurotransmitter and as a neuromodulation. Insulin is a combination of two polypeptides linked by disulfide linkage that regulates glucose in the blood. Malfunctioning of insulin causes diabetes.

7.4: Secondary structure of proteins

Learning Objectives

- Define the secondary structure of proteins and understand the structural features of major secondary structures, including α-helix, β-pleated sheet, random coil, and triple helix structures.

The polypeptide backbone is composed of repeated $-\mathrm{C}-\mathrm{C}-\mathrm{N}-$ units where the side chain ($-\mathrm{R}$) is hanging on the first C. The second carbon is the carbonyl ($\mathrm{C}=\mathrm{O}$) group and the nitrogen is the amine ($-\mathrm{NH}-$) group in amide. The polymer chain acquires confirmation to achieve:

1. maximize hydrogen bonding between the $\mathrm{C}=\mathrm{O}$ and $-\mathrm{NH}-$ of the same chain (intramolecular) or among the neighboring chains (intermolecular), and
2. minimize the steric-strain due to the $-\mathrm{R}$ groups.

This way, the polymer backbone acquires repetitive patterns in portions of the backbone.

Secondary structure of proteins

The secondary structure of proteins comprises organized regions of polypeptide backbone stabilized by hydrogen bonds between atoms.

The two common secondary structures encountered in proteins are (α-helix and β-pleated sheet. The other portions of the polymer backbone that are regular but not repetitive are called random coils. Triple-helix is another common secondary structure found in collagen proteins in connective tissues. These secondary structures are illustrated in Figure 7.4.1. below and described in the following sections.

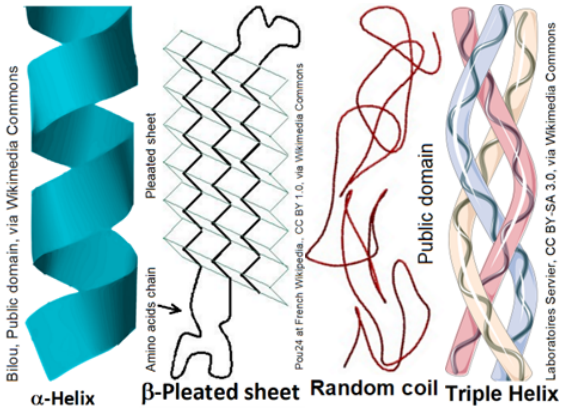


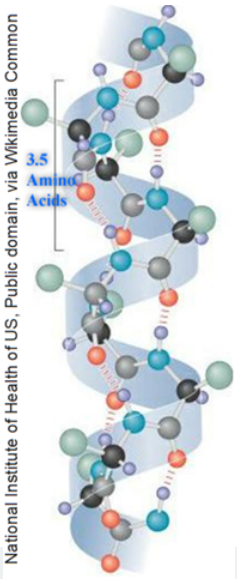
Figure 7.4.1: Basic secondary structures of proteins illustrated.

α-Helix

In the α-helix portion, the polypeptide chain acquires a coil shape spiraling clockwise from N-terminus to C-terminus, held by hydrogen bonds between the carbonly groups and amine groups in the backbone and side chains protruding outwards, as illustrated in the model shown on the left

Each $\mathrm{N}-\mathrm{H}$ is hydrogen bonded with a $\mathrm{C}=\mathrm{O}$ group four units away, shown by hashed lines in the model. Each side chain ($-\mathrm{R}$) group on an amino acid reside is protruding outwards from the helix.

The α-helix is represented by a right-handed spiral shape ribbon, as shown by the red-color ribbon in the structure of glucagon shown in Figure 7.4.2. The side chain on each amino acid can be seen protruding out from the helix. The random coil portion of glucagon is shown by a blue-color tube. Since the side chains hang out in free space, short and long-side chain amino acids can easily be accommodated in an α-helix without steric strain. Keratin, a fibrous protein of hair, fingernails, horns, and wool, is composed of a major portion of α-helix.



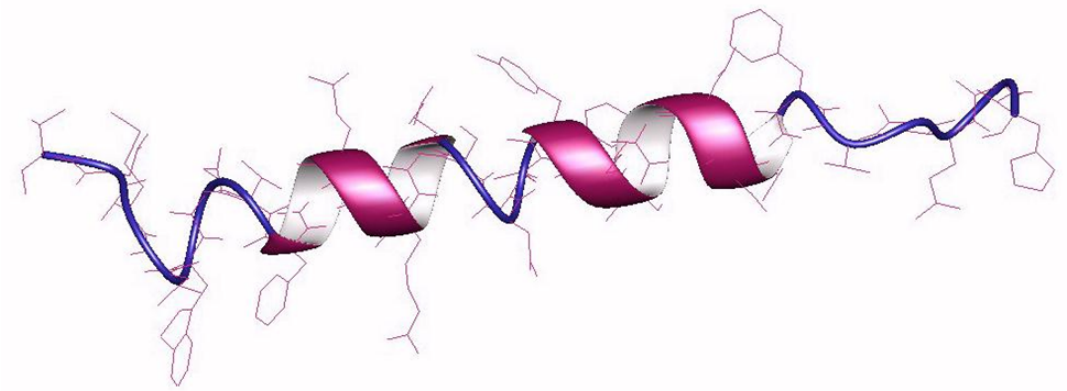
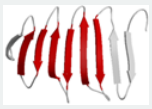
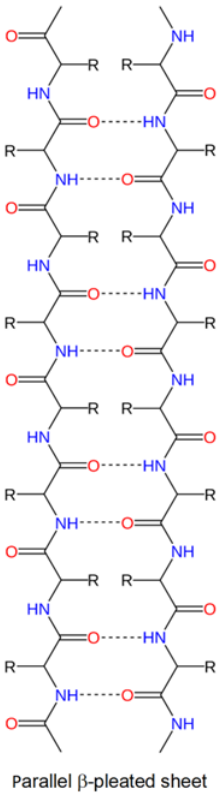


Figure 7.4.2: Secondary structure of Glucagon illustrated: α-helix shown as red ribbon and random coil as blue tubes with polypeptide backbone and side chains shown as stick model of the molecule. (Copyright; Deposition authors: Blundell, T.L., Sasaki, K., Dockerill, S., Tickle, I.J.; visualization author: User:Astrojan, CC BY 3.0, via Wikimedia Commons)

β-Pleated sheet



The β-Pleated sheet is a portion of polypeptide chains in which sections of polypeptide chains are aligned parallel or antiparallel and held together by hydrogen bonds between the carbonyl groups and amine groups of the neighboring polypeptide chains. The β-Pleated sheet is usually represented as a ribbon with an arrowhead pointing toward the N-terminus, as shown on the right.



Since the polypeptide backbone is in a zig-zag shape, multiple chains running parallel to each other in a β-pleated sheet is a pleated shape as shown in Figure 7.4.3. The two neighboring chains are **parallel** if both have N-terminus on the same side, as shown on the right, and **antiparallel** if the two adjacent chains have N-terminus on the opposite sides, as shown on the left. The neighboring chains may be different polypeptides or the same polypeptide chain with hairclip-like bents at the ends.

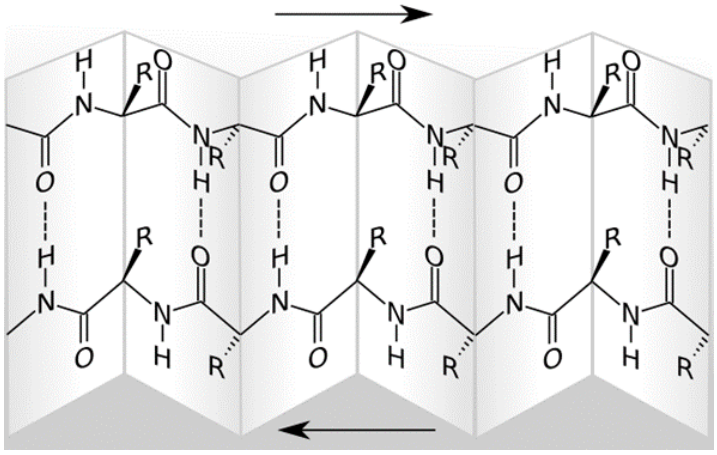
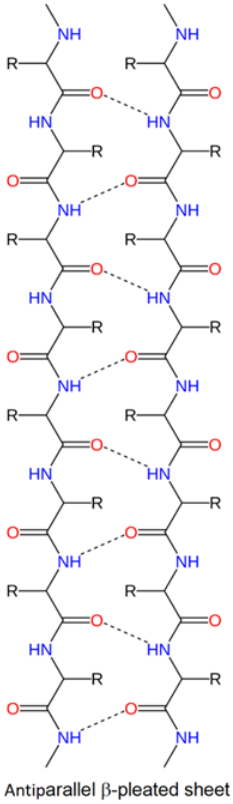


Figure 7.4.3: A β-pleated sheet structure portion of a protein illustrated. (Copyright; Roland.chem, CC0, via Wikimedia Commons)



The alternate C=O and N−H on the backbone of neighboring chains point towards each other and hydrogen bond with each other, as shown in Figure 7.4.3. The other set of alternate C=O and N−H point to the other side and establish hydrogen bonds with the chains on the other sides. These hydrogen bonds give mechanical strength to the proteins. Silk is made of a fibrous protein fibroin with a significant portion as β-pleated sheets.

The alternate side chains ($-\mathrm{R}$) point above and below in a β-pleated sheet. The amino acids in β-pleated sheets usually have short side chains, such as glycine, valine, alanine, and serine, as the long chains cause more steric strain.

Random coil

Most of the proteins have a mix of α-helix and β-pleated sheets and organized but not repeated structures between the two that are called random coils as illustrated in Figure 7.4.4. The random coins are represented by tubes.

Figure 7.4.4: Rotating 3d cartoon model of the human insulin monomer, showing α-helix portions as red ribbons, β-plated sheet portion as yellow ribbon, and random coil portions as white tubes. Ayacop, Public domain, via Wikimedia Commons. (Copyright; Ayacop, Public domain, via Wikimedia Commons.)

Triple helix

The triple helix is a secondary structure found in collagen. Collagen is a structural protein that is strong and elastic, present in connective tissues of the tendon, cartilage, blood vessels, skin, and bone. It is the most abundant protein of vertebrates making up 30% to 35% of their proteins by weight.

The triple helix is made of three collagen peptides which are left-handed helices that are placed together on a common axis and displaced (translated) along the axis, forming a right-handed triple helix, as shown in Figure 7.4.1 and Figure 7.4.5.

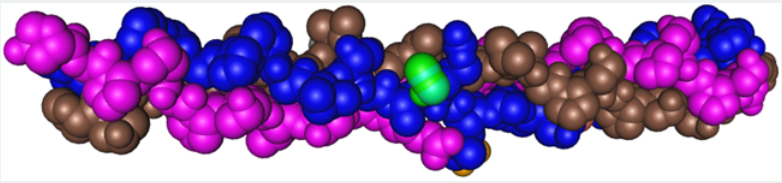


Figure 7.4.5: Triple helix of collagen illustrated. (Copyright; Nevit Dilmen, CC BY-SA 3.0, via Wikimedia Commons)

The collagen peptide is composed of repeat units of Gly-X-Y, where X is usually proline, and Y is generally hydroxyproline - a proline with a hydroxyl ($-\mathrm{OH}$ group installed on the side chain. Proline has a secondary α-amine group that causes bent in the peptide chain, which is needed for the helical structure of collagen peptides. The extra ($-\mathrm{OH}$) in the side chain of hydroxyproline allows more hydrogen bonding in the collagen peptide, which makes connective tissues stronger.

7.5: Tertiary structure of proteins

Learning Objectives

- Define the tertiary structure of proteins and understand the interactions that form and maintain the tertiary structure.

What is the tertiary structure of proteins?

The secondary structure of proteins represents the folding of portions of the polypeptide held primarily by hydrogen bonding between $\mathrm{C{=}O}$ and $\mathrm{N{-}H}$ groups of the polymer backbone. The overall protein chain also folds in a configuration that causes more stabilizing interaction primarily between the groups on the side chains ($\mathrm{-R}$) and the between side chain groups and the groups on the polymer backbone.

Tertiary structure

The three-dimensional arrangement of all the atoms of a single polypeptide chain in space, held together by stabilizing interactions between the side chain and the backbone groups, is called the **tertiary structure of proteins**.

The interactions stabilizing the tertiary structure include disulfide linkage, salt bridge, coordinate bonds with metal ions, hydrogen bonding, and hydrophobic interaction, as shown in Figure 7.5.1. and explained below.

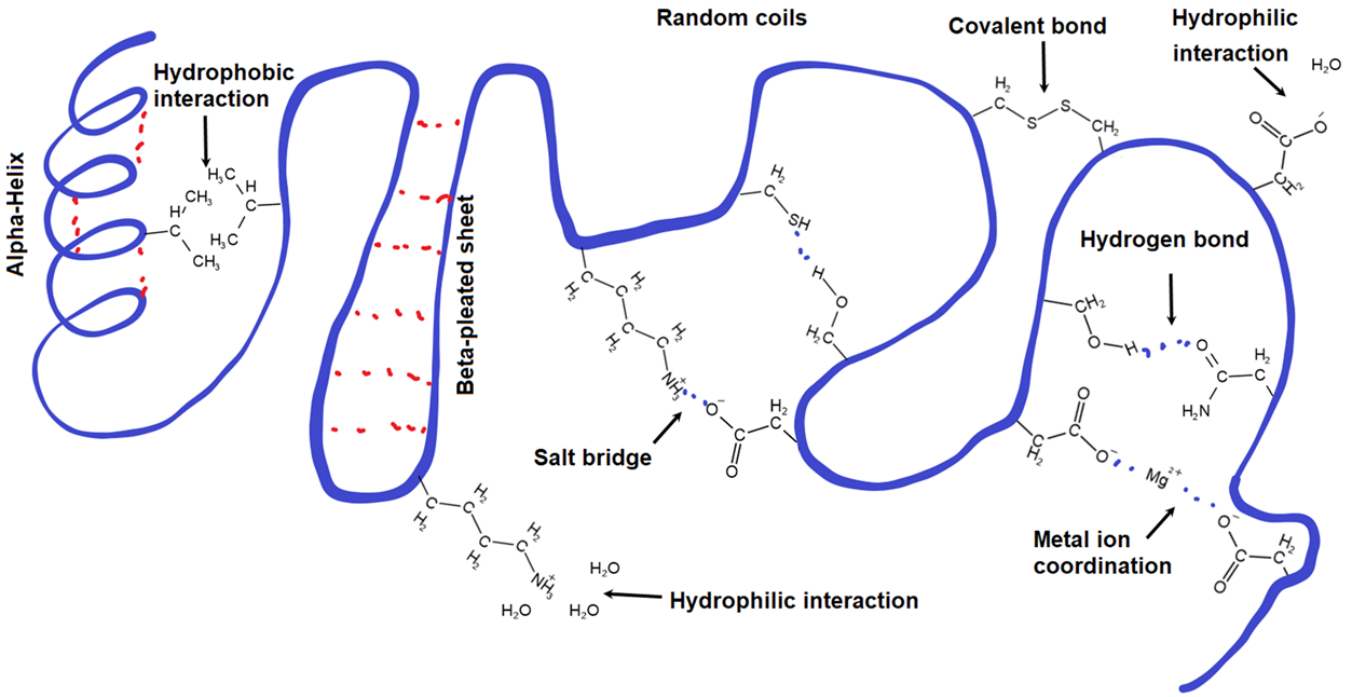


Figure 7.5.1: Illustration of disulfide linkage, salt bridge, coordinate bonds with metal ions, hydrogen bonding, and hydrophobic interaction that stabilize the tertiary structure of proteins. (Copyright; Public domain)

Disulfide linkage

When thiol ($\mathrm{-SH}$) groups of two cysteine residues come close to each other in a folded protein chain, they can make disulfide linkage (($\mathrm{-S{-}S-}$), i.e., the only covalent bond, besides peptide bond, holding amino acid residues together in proteins.

$$\mathrm{R{-}SH + HS{-}R' \xrightarrow{Oxidation} R{-}S{-}S{-}R + 2\,H^{+} + 2\,e^{-}}$$

Salt bridge

The side chain of acidic amino acids, e.g., Asp, loses protons and becomes an anion, and that of basic amino acids, e.g., Lys, gains a proton and becomes a cation. When the opposite ions come close to each other, they make an ionic bond, like that in salts, called

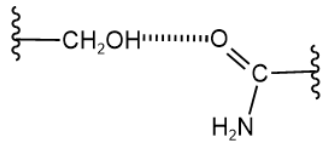
salt-bridge, e.g., $\cdots -CH_2-COO^{-}{}^{+}NH_3-(CH_2)_4-\cdots$) can form between Asp and Lys residues. This is an ionic bond holding opposite charges together.

Coordinate bond

Some metals can act as Lewis acids and make a coordinate covalent bond by accepting a lone pair of electrons from Lewis bases. Oxygen and nitrogen atoms in proteins have lone pairs of electrons and can act as Lewis bases. For example, oxygen atoms of two carboxylate $-COO^-$ groups can make coordinate covalent bonds with Mg^{2+}, resulting in a coordinate covalent linkage ($-COO^- \longrightarrow Mg^{2+} \longleftarrow OOC-$) between the amino acid residues.

Hydrogen bonding

Hydrogen atoms bonded to oxygen or nitrogen atoms, e.g., in $-OH$ and $-NH_2$, can make hydrogen bonds with any oxygen or nitrogen atom. These bonds could be between side chains, backbone groups, or side chains and backbone groups. For example, $-OH$ in the side chain of serine can make hydrogen with the carbonyl oxygen of the amide group ($-CONH_2$) of asparagine residue, as shown below.



Hydrophobic interactions

As shown below, amino acids with aliphatic or aromatic hydrophobic side chains, e.g., valine and phenylalanine, establish hydrophobic interaction.

Hydrophobic interaction between Val and Phe side cians shown in blue color.

Hydrophobic interaction is London dispersion forces between nonpolar groups. The physiological medium is predominantly water. As a nonpolar liquid like oil, when mixed with a polar liquid like water, the oil drops merge, expelling water. Similarly, hydrophobic groups in proteins come close to each other due to the hydrophobic interaction and tend to stay in the interior of the protein, expelling water away from those regions. Ionic and polar compounds dissolve in water. Similarly, ionic and polar groups on the side chains stay on the proteins' outer surfaces, interacting more with water.

The tertiary structure of proteins may be predominantly α-helix, β-pleated sheets, or random coil, but often it is a mix of these, for example, Carboxypeptidase A from bovine pancreas, shown in Figure 7.5.2.

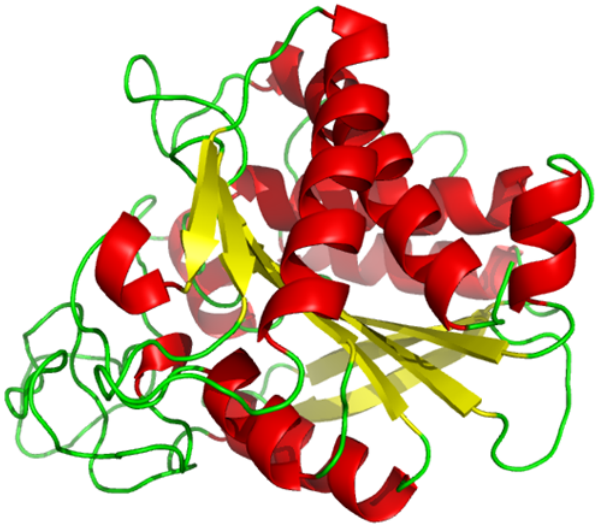


Figure 7.5.2: Tertiary structure of Carboxypeptidase A, from bovine pancreas, showing α-helix (red), β-pleated sheets (yellow) and or random coil (green). (Copyright; Lijealso, Public domain, via Wikimedia Commons)

7.6: Quaternary structure of proteins

Learning Objectives

- Define and understand the quaternary structure of proteins and its need.

Several proteins function as a single polypeptide with primary, secondary, and tertiary structures. Some proteins require more than one polypeptide with their primary, secondary, and tertiary structures combined to function.

Quaternary structure of proteins

The association of more than one protein into a closely packed arrangement is called the **quaternary structure** of proteins.

The individual proteins in the assembly are called monomers or subunits. The submit are held together by the same interactions that hold the tertiary structure of individual subunits, i.e., hydrogen bonds, hydrophobic interactions, salt bridges, etc.

An association of two proteins is called a dimer; three proteins is a trimer; four is a tetramer;five is a pentamer. For example, hemoglobin, shown in Figure 7.6.1. is a tetramer of two α subunits of 141 amino acid residues each and two β subunits of 146 amino acid residues each. These subunits are structurally similar, about the same size, and work cooperatively in the hemoglobin tetramer.

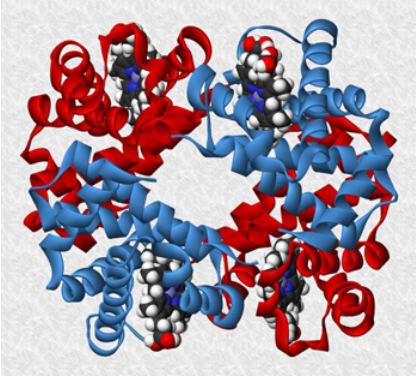


Figure 7.6.1: Illustration of a hemoglobin composed of four subunits: two α and two β subunits. Each subunit has a heme molecule inserted into cavities inside the submit. (Copyright; Benjah-bmm27. Modificado por Alejandro Porto., CC0, via Wikimedia Commons)

7.7: Summary of protein structure levels

Learning Objectives

- Understand the structure levels of proteins.

The summary of protein structural levels is illustrated in Figure 7.7.1 and described in a video. The protein structure in summary form is the following.

1. The primary structure is the sequence of amino acids in a polypeptide chain.
2. The secondary structure is the organized sections of the polypeptide chain, e.g., α helix and β pleated sheets.
3. The tertiary structure is the overall folded 3D structure of the entire polypeptide chain, held by disulfide linkages, salt bridges, hydrogen bonding, and hydrophobic interactions.
4. Quaternary structure is the association of more than one polypeptide chain (subunit) held together by the same type of interactions holding a subunit's tertiary structure.

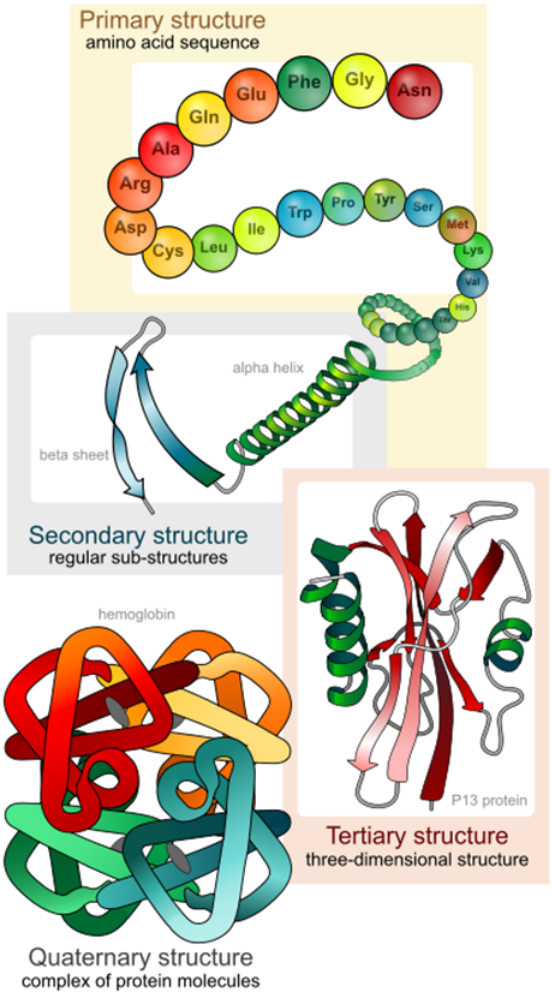


Figure 7.7.1: Illustration of primary, secondary, tertiary, and quaternary structure levels of proteins (left) and the video on What is a protein from RCSBProteinDataBank (right). (Copyright; Left: LadyofHats, Public domain, via Wikimedia Commons, right:)

7.8: Protein misfolding and denaturation

Learning Objectives

- Understand misfolding of proteins and the medical problems associated with it.
- Understand the denaturation of proteins and what causes the denaturation.

Protein misfolding and associated diseases

Newly synthesized proteins fold in a specific way, i.e., into secondary, tertiary, and quaternary structures, to be able to perform their function, as illustrated in Figure 7.8.1. Some proteins can fold in only one way, but others can fold in multiple ways. There are proteins in the cell, called chaperones, that help newly formed proteins to fold in the way needed for their function.

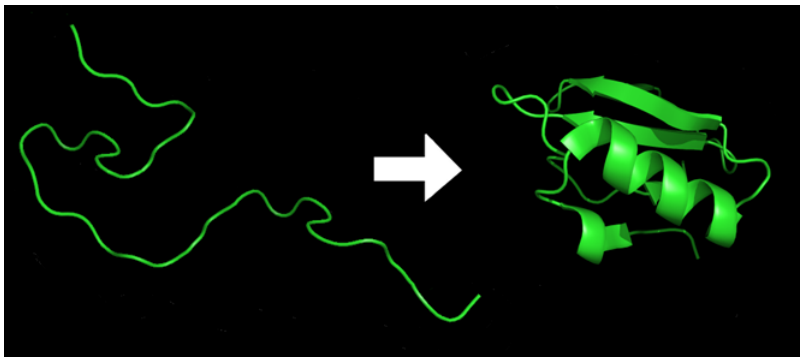


Figure 7.8.1: Illustration of the protein folding process. Chymotrypsin inhibitor 2 from pdb file 1LW6. (Copyright; DrKjaergaard, Public domain, via Wikimedia Commons)

Sometimes normal proteins misfold and become pathological. Often, these proteins are in soluble α helix forms that re-assembles into β-pleated sheet forms that are sticky and aggregate into plaques or amyloid structures, as illustrated in Figure 7.8.2, with the example of plaque formation in Alzheimer's affected brain.

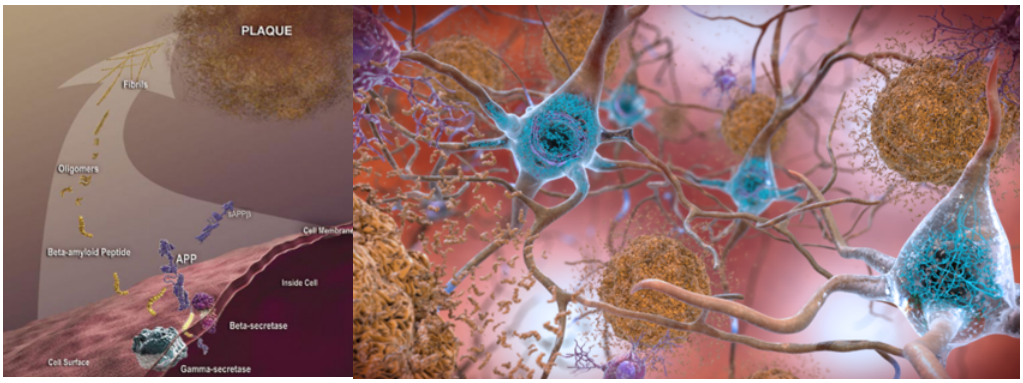


Figure 7.8.2: Illustration of plaque formation when β-amyloid changes confirmation from α-helix to β-pleated sheets that aggregate to form the plaque (left) and abnormal levels of β-amyloid proteins aggregated to form plaques (shown in brown color) that collect between neurons. (Copyright; Left: National Institute on Aging, Public domain, via Wikimedia Commons, Right: NIH Image Gallery from Bethesda, Maryland, USA, Public domain, via Wikimedia Commons)

Prions are small proteins found in nerve tissue. Their exact functions are unknown, but when they misfold, they can cause more normal proteins to misfold. This protein misfolding is related to diseases such as mad cow disease in cows, Creutzfeldt–Jakob disease in humans, Alzheimer's disease, and familial amyloid cardiomyopathy or polyneuropathy, as well as intracellular aggregation diseases such as Huntington's and Parkinson's disease.

Denaturation of proteins

Proteins usually keep their primary, secondary, tertiary, and quaternary structures under physiological conditions. However, some physical processes and chemical agents can beak the interactions, i.e., hydrogen bonding, disulfide linkages, salt bridges, and hydrophobic interactions.

Denaturation

Loss of the secondary, tertiary, and quaternary structures of proteins by a physical process or a chemical agent while maintaining the primary structure almost intact is called **denaturation of proteins**.

Proteins unfold and become almost linear polypeptide chains upon denaturation. Denatured proteins can not perform their functions. Denaturation can be caused by heat, acids or bases, organic compounds and solvents, heavy metal ions, and agitation, as explained below.

- **Heating** above 50 ºC disrupts hydrogen bonding and hydrophobic interactions, causing protein denaturation. For example, cooking food and sterilizing surgical instruments by autoclave treatment. In laser surgery, laser, i.e., the light of a single wavelength, is focused on a spot, causing heat that denatures proteins. Heating by laser cauterizes incisions, i.e., burns the site or the wound. It helps prevent blood loss.
- **Acids and bases** disrupt hydrogen bonding and salt bridges, e.g., by neutralizing some ions involved in the salt bridges.
- **Organic compounds** like urea disrupt hydrogen bonding by replacing them with stronger hydrogen bonding with the compound, and **organic solvents** disrupt hydrogen bonding and hydrophobic interactions. For example, 75% alcohol sterilizes skin by denaturing and coagulating proteins.
- **Heavy metals ions** like Pb^{2+} and Hg^{2+} attack $-SH$ groups making salt bridges, like $-S^{-}Pb^{2+}{}^{-}S-$. Egg white or milk is an antidote to heavy metal poisoning as they precipitate the metal ions in the stomach, based on the salt bridge forming reaction. Vomiting is induced to remove the metal before the precipitate is dissolved, releasing the metal in the later parts of the digestive system.
- **Agitation** physically disrupts hydrogen bonding and hydrophobic interactions. Whipped cream and whipped egg white are prepared based on the agitation process.

7.9: Enzymes

Learning Objectives

- Understand enzyme-related terminology, nomenclature, and classification of enzymes.
- Understand the mode of action of enzymes, the factors that affect them, and the inhibitors that retard or damage the enzyme activity.

Enzymes and the related terminology

Three conditions are necessary for a chemical reaction to happen:

1. Reactants must collide -the more frequent the collisions faster the reaction,
2. the reactant must have proper orientation at the time of the collision -the higher the probability of proper orientation at the time of the collision, the faster the reaction, and
3. there must be enough energy at collision to surpass the energy barrier, i.e., the energy of activation for the reaction -the lower the activation energy, the faster the reaction.

Often chemical reactions are possible but so slow that they are practically useless. For example, a reaction between hydrogen H_2 and nitrogen N_2 producing ammonia NH_3 is possible, but to make it practically useful, high-pressure, high-temperature, and catalysts are needed.

$$N_2(g) + H_2(g) \overset{heat, pressure, catalyst}{\rightleftharpoons} 2\,NH_3(g) \quad -92 \cdot 4\,kJmol^{-1}$$

Catalyst

A catalyst is a reagent that increases the rate of a chemical reaction without itself being altered in the process.

The catalysts usually increase the rate of chemical reactions by improving the last two factors, i.e., increasing the probability of proper orientation and providing an alternate route for the reaction with lower activation energy. There are more constraints for chemical reactions in living things, e.g., the reaction has to occur under physiological conditions of pH ~7.4 and body temperature ~37 ^{o}C. Special catalysts called enzymes are used to regulate chemical reactions in living things.

Enzyme

An **enzyme** is a substance that regulates the rate of chemical reaction in living things without itself being altered in the process. In other words, enzymes are biological catalysts.

Like other catalysts, the enzymes usually increase the rate of chemical reactions by improving the last two factors, i.e., increasing the probability of proper orientation and providing an alternate route for the reaction with lower activation energy. The enzymes achieve the appropriate orientation of the reactants by binding them in a specific region within the enzyme, which has the geometry of its interacting groups right for securing the reactant in a particular orientation.

Substrate

A reactant in an organic or biochemical reaction is a substrate.

Active site

The enzyme's active site is the region within an enzyme where the substrate binds for the reaction.

Enzymes are usually proteins having primary, secondary, and tertiary structures. The active site is usually a small region (10% to 20%) within the enzyme. A few side chains of amino acid residues within the active site participate in the catalytic action, as shown in Figure 7.9.1. The rest of the amino acid residues define and hold the secondary and tertiary structures.

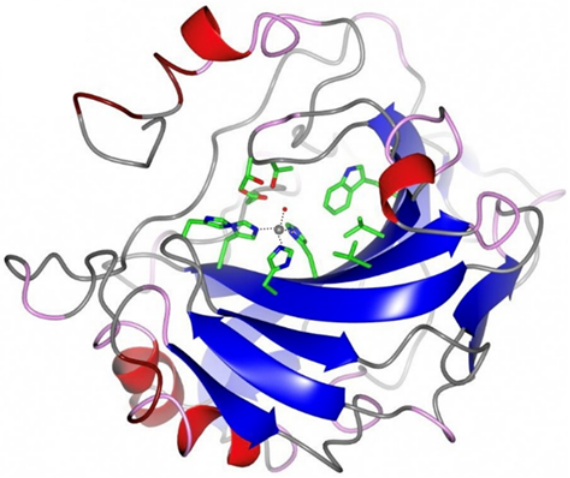


Figure 7.9.1: Human carbonic anhydrase II, CAII enzyme (PDB code: 1CA2) showing the secondary structures and the active site. The active site has a hydrophilic pocket comprising of three histidine residues (green ring with two blue nitrogen per ring) bonded to Zn^{2+} (gray dot), a water molecule (red dot) also connected to zinc and two threonine and one glutamic acid residues (green with red tips) above water. A hydrophobic pocket on the right side in the active site comprises two valine (gree Y-shapes) and a tryptophan (two fused rings, green). (Copyright; Eriksson, A.E., Jones, T.A., Liljas, A.(1988) Proteins 4: 274-282PubMed: 3151019 Search on PubMedDOI: 10.1002/prot.340040406, CC0, via Wikimedia Commons)

Some enzymes need a non-protein part to combine with them for their function.

Apoenzyme, cofactors, and coenzyme

The enzymes that need a non-protein portion to combine with them for their function are called **apoenzymes**. The non-protein portion of the enzymes is called the **cofactor**, as illustrated in Figure 7.9.2. The cofactor could be a metallic ion, .e.g., Zn^{2+} or Mg^{2+}, or an organic compound. Organic cofactor is called **coenzyme**. The apoenzyme and cofactors together are called holoenzymes, as illustrated in Figure 7.9.2.

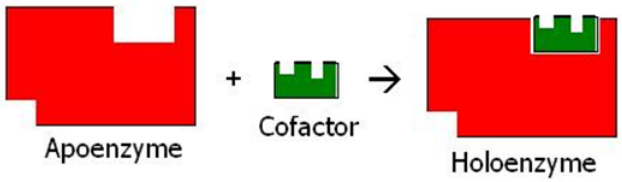
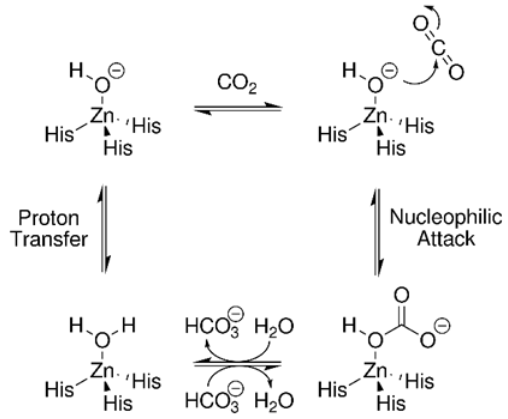


Figure 7.9.2: Apoenzyme, cofactor, and their combination, i.e., the holoenzyme illustrated. (Copyright; Moniquepena, Public domain, via Wikimedia Commons)

Enzymes usually make the reaction happen millions of times faster. For example, one molecule of carbonic anhydrase enzyme, shown in Figure 7.9.1, can catalyze about one million molecules in one second in human blood by the following reaction.



Copyright: Synpath, CC0, via Wikimedia Commons.

Overall reaction: $CO_2 + H_2O \rightleftharpoons HCO_3^- + H^+$

Selective

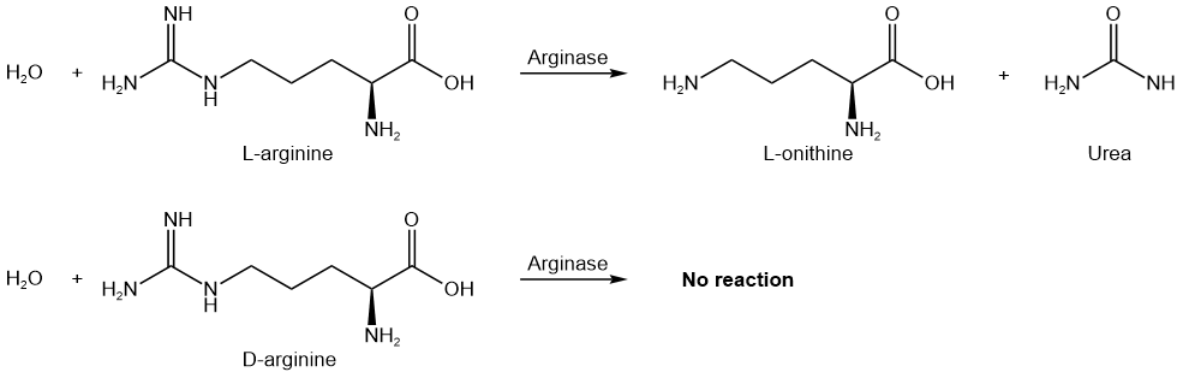
A catalyst or a reagent is selective if it produces one product preferentially or exclusively when more than one product is possibly formed.

Enzymes are not only selective; they usually react with one compound or stereoisomer.

Stereospecific

Stereospecific catalysts or reagents are those that, when reacting with one stereoisomer, selectively produce a stereoisomer and either do not react with the isomers of the reactant or selectively produce other stereoisomers from them. Enzymes are usually stereospecific, i.e., they are selective among reactants and selective among products.

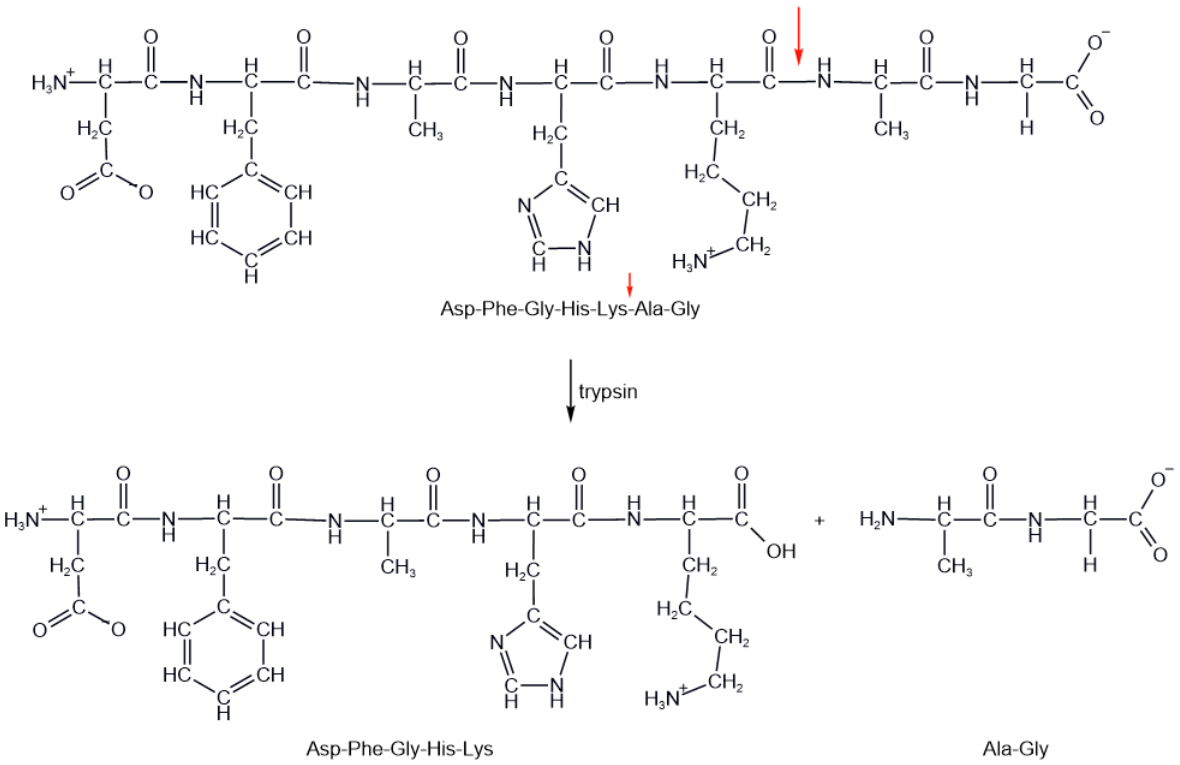
For example, the enzyme arginase is stereospecific, hydrolysis amino acid L-arginine to L-ornithine and urea, but does not react with D-arginine.



Usually, enzymes react with one compound or a specific bond of one compound. For example, enzyme ureas catalysis hydrolysis of urea ($(NH_2)_2C{=}O$) and does not hydrolyze any other amide.

$$(NH_2)_2C{=}O + H_2O \xrightarrow{urease} 2\,NH_3 + CO_2$$

Trypsin is a digestive enzyme that cleaves peptide bonds of proteins but not every peptide bond, only those on the C-side of lysine and arginine residue.



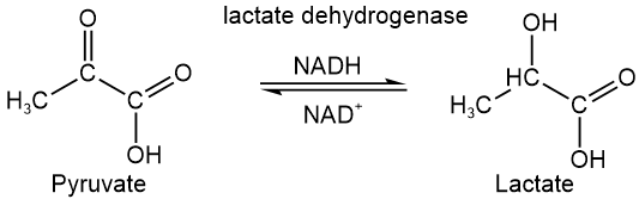
Some enzymes are specific for a class of compounds, e.g., lipases catalyzing triglyceride hydrolysis, but do not react with carbohydrates or proteins.

Names and classification of enzymes

Names of enzymes are derived by replacing the end of the name of a reactant or reaction with the suffix *ase*. For example, sucrase hydrolysis sucrose, lipase hydrolyzes lipids, oxidase catalyzes oxidation reactions, dehydrogenase removes hydrogen atoms, etc. Some old names of enzymes have the suffix *in*, e.g., digestive enzymes pepsin and trypsin. In the recent classification of enzymes, the name or the class name indicates the type of reaction it catalyzes. There are six major classes of enzymes, as described in the following.

Oxidoreductases

Oxidoreductases do oxidation-reduction reactions, i.e., **oxidases** oxidize a substance, **reductases** reduce a substance, and **dehydrogenases** remove hydrogen. For example, lactate dehydrogenase reduces pyruvate and oxidizes lactate, as shown below.

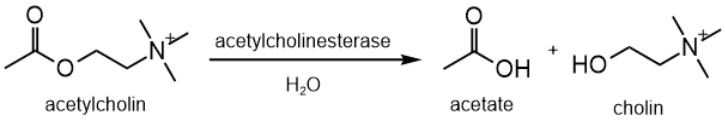


Transferases

Transferases transfer a group between two compounds. For example, kinases transfer phosphate groups, and transansaminases transfer amino groups, as shown below.

Hydrolases

Hydrolases do hydrolysis reactions: lipases hydrolyze lipids, carbohydrates hydrolyze carbohydrates, proteases hydrolyze proteins, phosphatases hydrolyze phosphate esters, and nucleases hydrolyze nucleic acids. For example, acetylcholinesterase hydrolysis acetylcholine, as shown below.



Lyases

Lyases add two groups to double bond or remove two groups from adjacent atoms to create a double bond or a ring structure by means other than hydrolysis and oxidation, e.g., carboxylases add or remove CO^2, and deaminases add or remove NH^3. For example, actonitase, shown in Figure 7.9.3, adds or removes water. double bond or add a new ring structure.

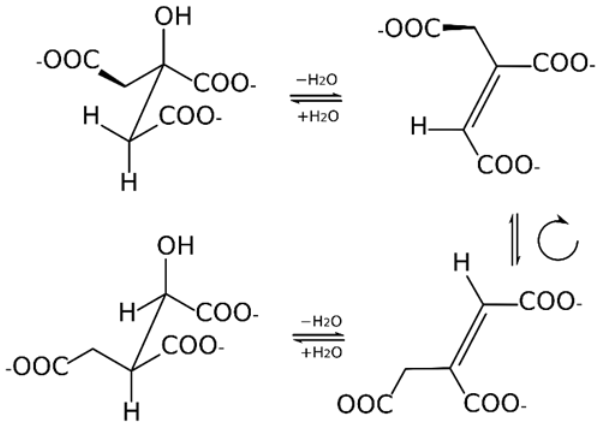
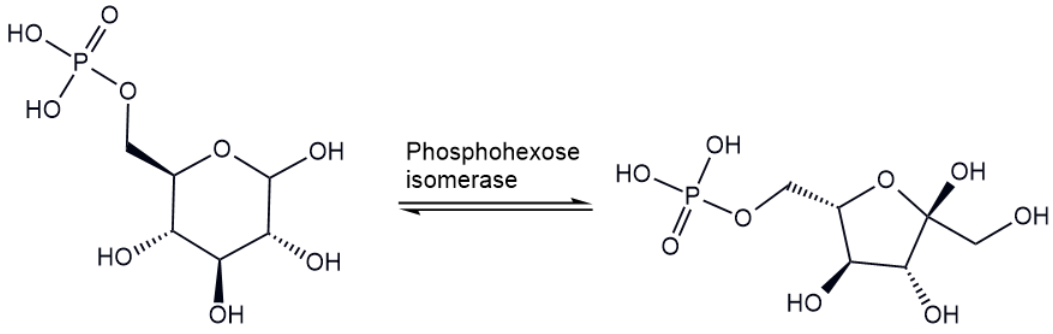


Figure 7.9.3: Reaction steps in the reaction catalyzed by aconitase, stereospecific view. Top left: Citrate. Bottom left: Isocitrate. Right: Aconitate. The aconitate has to flip by 180 degrees between the reactions. (Copyright; Ayacop, Public domain, via Wikimedia Commons

Isomerases

Isomerases rearrange atoms in a molecule: isomerases convert cis to trans or trans to cis isomer, and epimerases convert D to L or L to D streoisomers. Phosphohexose isomerizes shown below isomerizes glucose-6-phosphate to fructose-6-phosphate

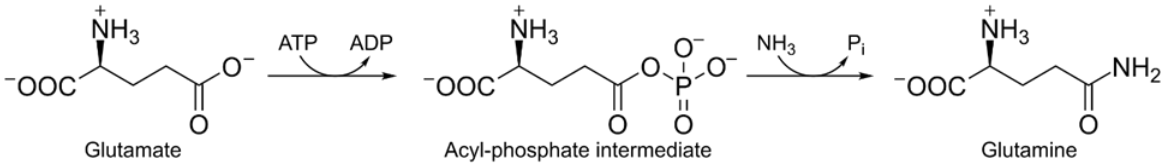


Glucose-6-phosphate

Fructose-6-phosphate

Ligases

Ligases or synthetases catalyze the joining of two molecules, e.g., glutamate to glutamine conversion by glutamine synthetase shown below.



(Copyright; Hbf878, CC0, via Wikimedia Commons)

How do enzymes catalyze the reactions?

The chemical reactions that enzymes catalyze can occur without enzymes, but the reactions' rates are usually prolonged due to high activation energies. Enzymes provide alternate routes to the reactions with lower energy barriers (activation energies) and proper orientations of substrates that result in fast reactions, as illustrated in Figure 7.9.3

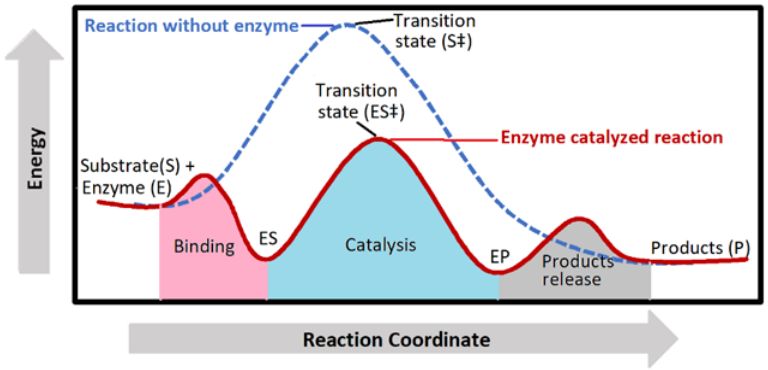


Figure 7.9.3: Illustration of reaction coordinates of a biochemical reaction with and without enzyme. (Copyright; Public domain)

In a generalized enzyme-catalyzed reaction, enzyme (E) and substrate (S) bind to make an enzyme-substrate complex (ES). The intermolecular interactions between the enzyme and the substrate usually loosen the bonds of the substrate that need to be broken, resulting in a lower energy barrier for the catalyzed reaction that leads to the product (P) at a faster rate than otherwise. The product leaves the enzyme so it can bind with another substrate.

$$\mathrm{E + S \rightleftharpoons ES \longrightarrow E + P}$$

For example, carbonic anhydrase catalyzes the reaction of water ($\mathrm{H_2O}$) with carbon dioxide ($\mathrm{CO_2}$) in blood as illustrated in Figure 7.9.4. The carbonic anhydrase enzyme is bound with cofactor $\mathrm{Zn^{2+}}$ through three histidine residues. The $\mathrm{Zn^{2+}}$ is also bound with $\mathrm{H_2O}$ that bond makes water ready for the proton transfer. After the proton transfer, a strong electrophile $-\bar{O}H$ is generated that attacks $\mathrm{CO_2}$, which is placed in the proper location in the hydrophobic pocket nearby. The product is displaced with another water molecule that repeats the process. One molecule of carbonic anhydrase can convert about million $\mathrm{CO_2}$ molecules per second in this reaction.

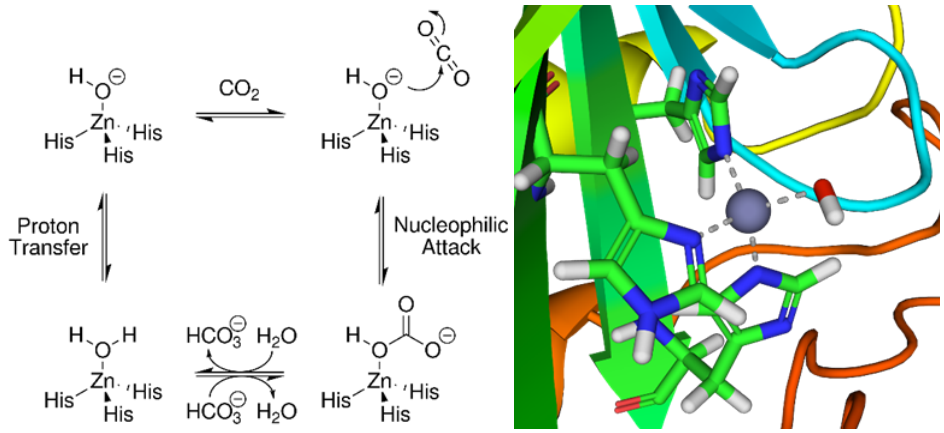


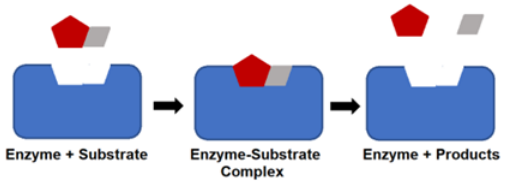
Figure 7.9.4: Mechanism of carbonic anhydrase catalyzed reaction of water with carbon dioxide in the blood (left) and a closeup of the active site of the enzyme showing Zn^{2+} (gray sphere) bonded with three histidine residue and $-\bar{O}H$ (red and white bar). There is a hydrophobic pocket near $-\bar{O}H$ that is shown in the full model in Figure 7.9.1. (Copyright; left: Own work, Public domain, via Wikimedia Commons, right: Own work, Public domain, via Wikimedia Commons)

Models of enzyme action

Enzymes are usually specific for a particular substrate or a class of reaction. This specificity was explained first by models of enzyme action explained below.

Lock-and-key model

The lock and key model assumes enzymes have a fixed shape with active sites similar to a lock in which a particular substrate with a fixed shape similar to a key can fit, as shown in the figure on the right.



Lock-and-key model of enzyme action illustrated.

This model explains the specificity of enzymes but ignores the dynamic nature of molecules. Further, according to this model, the products would be expected to be snug-fit in the active site and not easy to release. In reality, all the bonds in a molecule are vibrating and stretching, and the whole molecule is jiggling due to thermal energy, as illustrated in Figure 7.9.5.

Figure 7.9.5: Illustration of thermal motion of a segment of protein alpha helix. These motions have various modes of vibrations, stretching, and rotations. In addition to these internal movements, molecules' external portions still move— like the jiggling of a water balloon. (Copyright; en:User:Greg L, CC BY-SA 3.0, via Wikimedia Commons)

Induced-fit model

The induced-fit model of enzyme action accounts for the flexibility of enzyme and substrate molecules. According to this model, the flexibility of the enzyme molecule allows the active site to adapt to the shape of the substrate and, at the same time, the substrate adopts the shape of the enzyme to acquire the best possible orientation for the reaction to occur, as illustrated in Figure 7.9.6. Then the active site shape re-adjusts to let the products be released to allow the next cycle of the enzyme action. The observation from experiments during the actual catalysis reaction supports the view that not only does the active site of the enzyme change shape, the backbone and the side chains of the enzyme molecule remain in constant motion during the enzyme action.

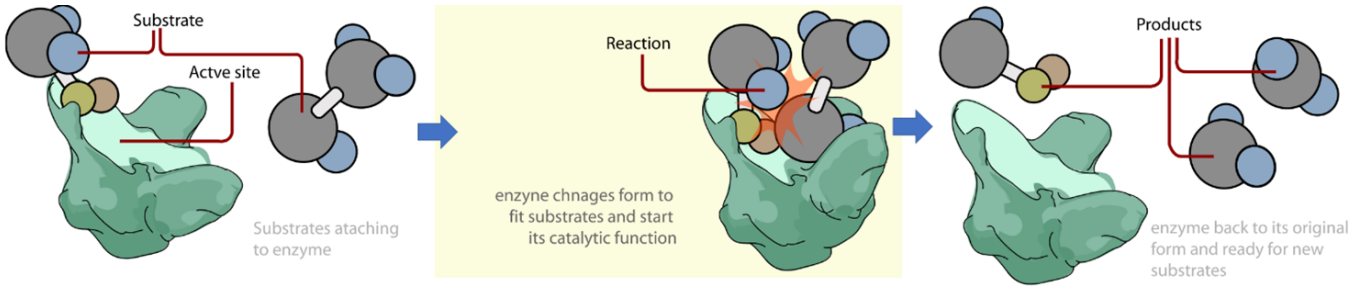


Figure 7.9.6: diagram showing the induced fit model of enzymes. (Copyright; Modified from LadyofHats, Public domain, via Wikimedia Commons)

Experimental data shows that the active site is usually a tiny portion (10% to 20%) of the enzyme. Within the active site, two or a few side chains of amino acid residues usually catalyze the reaction. Usually, the catalytically active residue is one of the following: His, Cys, Asp, Arg, and Glu. These amino acids have acidic or basic or thiol functional groups, which are not only capable of hydrogen bonding but also capable of acid, base, electrophile, or nucleophile catalysis. For example, the pepsin enzyme breaks peptide bonds by using histidine and cystine, as illustrated in Figure 7.9.7.

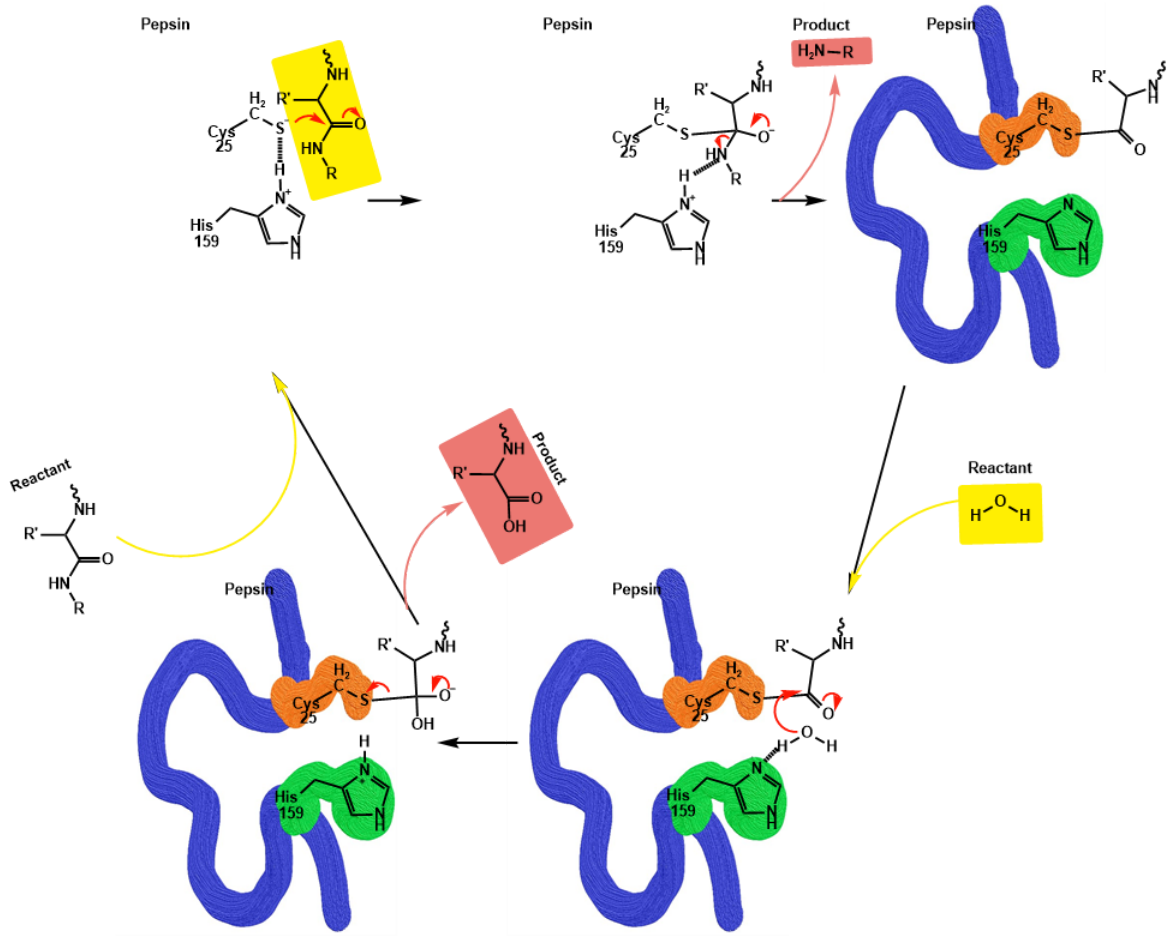
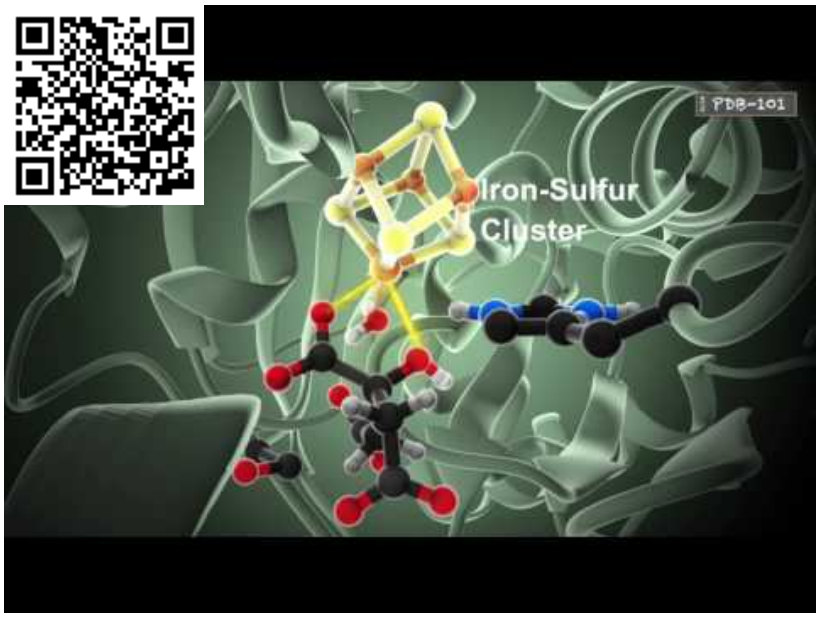


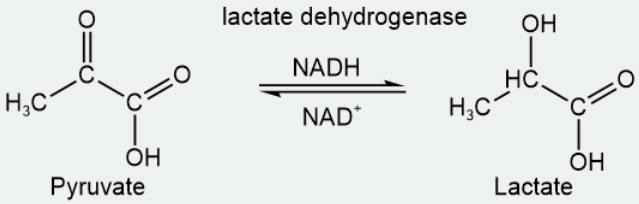
Figure 7.9.7: Illustration of the mechanism of peptide bond cleavage by pepsin enzyme: histidine removes a proton from $-SH$ group of cystine making it a strong nucleophile $-S^-$. The $-S^-$ attacks a $C{=}O$ group of the peptide and breaks the peptide bond by nucleophilic acyl substitution mechanism. Then histidine brings in a H_2O at a suitable position for a second nucleophilic acyl substitution that breaks the $S-C$ bond. Products leave, and the enzyme repeats the cycle. (Copyright; Public domain)

The following video from RCSBProteinDataBank explains how enzymes work.



Isoenzymes as a medical diagnostic tool

Isoenzymes are different enzymes that perform the same function in different body parts. They usually have a tertiary structure and differ in only a few amino acid residues. For example, lactate dehydrogenase (LDA) catalyzes the conversion between pyruvate and lactate, as shown below.



LDA is a tetramer made of two sub-units, the H-form, and the M-form, in five combinations found in different tissues, as shown in Table 1. Similarly, creatine phosphokinase (CPK) catalyzes the interconversion of phosphocreatine to creatine. CPK is a dimer in 3 isoenzyme forms shown in Table 2.

These isoenzymes usually function within cells. However, when some disease damages a tissue, the cells die, and the isoenzymes are released into the blood. Analysis of blood serum is used to diagnose the location of the damage. For example, an elevated level of LDH5 indicates liver damage and myocardial infarction (heart damage) is characterized by a high level of LDH1 isoenzyme. The heart damage will also elevate CK_2.

Table 1: Different forms of lactate dehydrogenase (LDA), their subunit composition the location in the body.

Type	Subunits	Illustration	Location
LDH_1	HHHH		Heart and Erythrocyte
LDH_2	HHHM		Heart and Erythrocyte

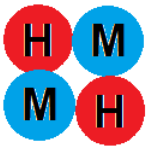
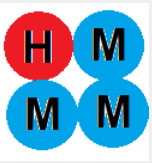
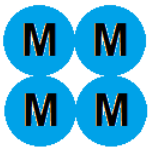
Type	Subunits	Illustration	Location
LDH_3	HHMM		Brain and Kidney
LDH_4	HMMM		Skeletal Muscle and Liver
LDH_5	MMMM		Skeletal Muscle and Liver

Table 2: Isoenzymes of creatine phosphokinase (CPK) dimer, composition, and location in the body.

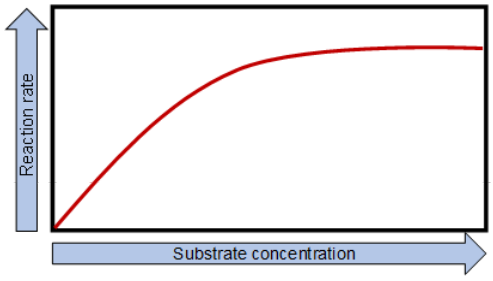
Isoenzyme	Subunit	Illustration	Tissue of Origin
CPK_1	BB		Brain
CPK_2	MB		Heart
CPK_3	MM		Skeletal muscle

Factors that affect enzyme activity

The enzyme activity is related to how much it increases the reaction rate compared to the same reaction without the enzyme. The factors that affect enzyme activity include concentration, temperature, pH, and the presence of inhibitors.

Concentration

Usually, the enzyme is in very low concentration relative to the concentration of the substrate. Therefore, an increase in the enzyme concentration increases the reaction rate linearly, i.e., doubling the enzyme concentration doubles the reaction rate, and tripling it triples the rate. An increase in the substrate concentration increases the reaction rate but not linearly; it follows a curvilinear curve. The rate increase reaches a saturation level and does not increase after that with a further increase in the substrate concentration, as illustrated in the figure on the left. This is because all the enzyme's active

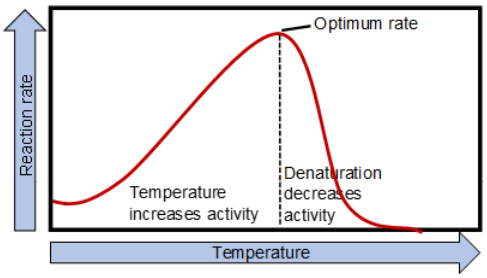


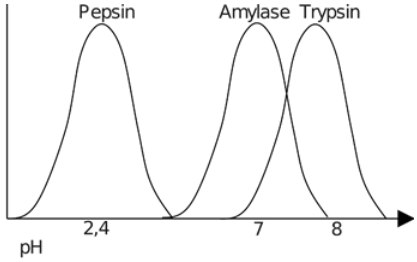
Copyright; English wikipedia, Public domain, via Wikimedia Commons

sites become occupied with the substrate at the saturation point, and the reaction proceeds at its maximum rate, as illustrated in the figure on the right.

Temperature

Generally, an increase in temperature increases the rate of a chemical reaction. This is because more molecules have energy than activation energy at higher temperatures. The same applies to enzyme-catalyzed reactions at lower temperature ranges. Still, the rate becomes optimum at around body temperature and then starts to fall off, as illustrated in the figure on the right. The enzymes have secondary, tertiary, and quaternary structures optimized to perform best at the body temperature of about 37 ^{o}C. Denaturation inactivates the enzymes. It is reversible if the temperature is slightly above body temperature, but enzymes are denatured beyond repair at a much higher temperature.





Copyright; Lanzi, GFDL, via Wikimedia Commons

pH

The enzymes work best at the pH of the tissue or organ where they work. This is because the structure of enzymes is optimized for the pH at which they usually function. An increase or a decrease in pH from the optimum value disrupts the secondary, tertiary, and quaternary structures and causes a reduction in enzyme activity. The effect of a slight change in pH is reversible, but a significant difference in pH denatures the enzymes permanently.

The optimum pH for most enzymes is around the physiological pH of 7.4. Still, some enzymes have different optimum pH, depending on the pH in their natural environment, as illustrated in the figure on the left. For example, the optimum pH of the starch-splitting amylase is pH 7-7.5 in the mouth. Pepsin breaks down proteins at around pH 2 in the stomach. Trypsin breaks down proteins at pH 8 in the intestine.

Enzyme inhibition

Inhibitors are substances that make enzyme lose their catalytic activity. Inhibitors prevent the substrate from binding with the active site of the substrate. The inhibition may be reversible or irreversible. There are two subclasses of reversible inhibitors; competitive inhibitors and noncompetitive inhibitors.

Competitive inhibitor

The competitive inhibitor has a shape similar to the enzyme's natural substrate. So, they compete with the substrate for the active site but do not react, as illustrated in Figure 7.9.8. Since there is competition between the inhibitor and the substrate for the active site, increasing the concentration of the substrate wins by outnumbering the inhibitor in the completion, and the enzyme regains its activity.

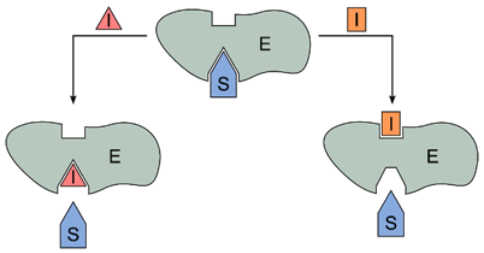


Figure 7.9.8: Illustration of an enzyme-substrate complex (top), competitive inhibitor inhibiting the enzyme activity by occupying the active site (bottom left), and a noncompetitive inhibitor binding to the enzyme outside the active area making the active site shape change such that the substrate can not fit in it properly (bottom right) (**E** enzyme, **I** inhibitor, **S** natural substrate). (Copyright; Sponk, Public domain, via Wikimedia Commons)

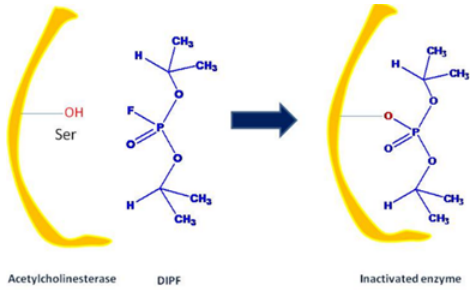
Noncompetitive inhibitor

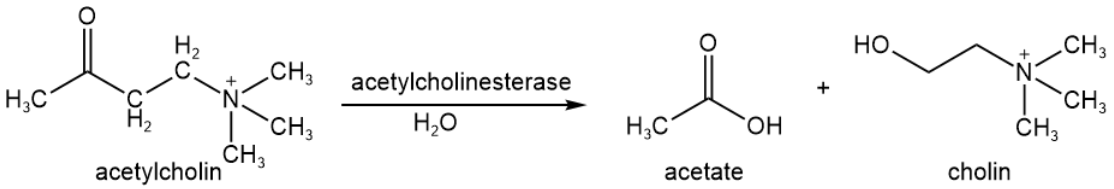
A noncompetitive inhibitor does not have a shape similar to the substrate and does not bind to the active site. It binds with the enzyme outside the active area but changes the folding patterns of the protein such that the active site can not acquire the proper shape. So, the substrate can not fit into the active site properly, as illustrated in Figure 7.9.8.

Examples of noncompetitive inhibitors are toxic metals, like Pb^{2+}, Hg^{2+}, and Ag^{+}. Since the inhibitor does not compete with the substrate, increasing the substrate concentration does not recover the enzyme activity. However, some chemical agents can remove the inhibitor from the enzyme, and the catalytic activity can be recovered.

Irreversible inhibitor

Irreversible inhibitors bind with enzymes and destroy their catalytic activity permanently. They usually bind with enzymes by covalent bonds that are not easily broken. Irreversible inhibitors include antibiotics, insecticides, and never gases. For example, penicillin is an irreversible inhibitor of an enzyme needed to form cell walls in bacteria. The bacteria without a complete cell wall can not survive. It does not affect the cell membranes of humans. Similarly, a diisopropyl fluorophosphate (DIPF) which is an organophosphate insecticide, binds with the $-OH$ of a serine residue in the enzyme acetylcholinesterase, as shown in the figure on the right. The enzyme-catalyzed reaction of acetylcholine, shown below, is needed for nerve impulse transmission. Inhibition of acetylcholinesterase blocks nerve impulses, causing paralysis.





CHAPTER OVERVIEW

8: Nucleic acids

8.1: Nucleotides -the building blocks of nucleic acids

Learning Objectives

- Understand the function of nucleic acids.
- Learn the structures and names of nucleotides -the building blocks of nucleic acids, and their constituents, including ribose, deoxyribose, nitrogen bases, and phosphate groups.

Nucleic acids are biopolymers that carry the codes for synthesizing proteins and pass them on from generation to generation, i.e., they are genetic materials. In other words, nucleic acids are the instruction manual for biochemical reactions in living things.

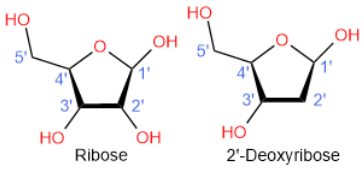
Nucleotides are the building blocks, i.e., the repeat units or monomers of nucleic acids.

Nucleotides are composed of three sub-units:

1. a 5-carbon carbohydrate,
2. a base that is an aromatic compound containing nitrogen, and
3. an anion of phosphoric acid, i.e., phosphate (PO_4^{3-}).

5-Carbon carbohydrate

Two 5-carbon carbohydrates, i.e., ribose and deoxyribose, are found in nucleic acids, as shown in the figure below. Both are D-sugars in five-membered cyclic form with hydroxyl ($-OH$) group at anomeric (C#1') pointing towards O in the ring, i.e., in a β configuration. The number labels have prime symbols, i.e., 1', 2', 3', etc., to distinguish them from regular numbers 1, 2, 3, etc., used for the nitrogen bases of the nucleic acids. The only difference between ribose and deoxyribose is that hydroxyl ($-OH$) group at C#2' is missing in the latter, which gives the deoxy- prefix to its name.

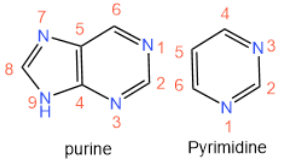


Types of nucleic acids

There are two types of nucleic acids: ribose in its nucleotides, called **ribonucleic acid** (**RNA**), and deoxyribose in its nucleotides, called **deoxyribonucleic acid** (**DNA**).

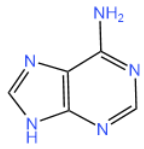
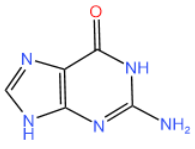
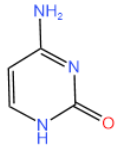
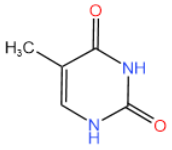
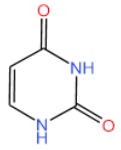
Nitrogen bases

Nitrogen bases in nucleic acids are derivatives of two aromatic compounds, purine and pyrimidine, shown below.



Cyclic compounds that contain atoms other than C′s in the cycle are called heterocyclic compounds. Purine and pyrimidine are heterocyclic aromatic compounds because they contain N′s in the cycles. Purine is bicyclic, containing a six-membered ring with two N′s fused with a five-membered ring with two N′s. Pyrimidine is a six-membered cycle with two N′s. Both purine and pyrimidine are planer molecules like benzene.

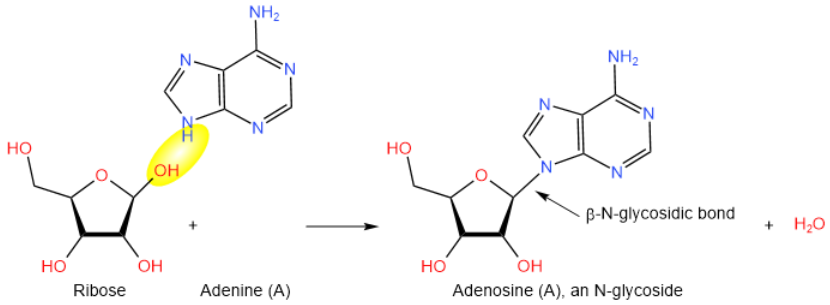
There are five nitrogen bases found in nucleic acids: two are purines, i.e., adenine (A) and guanine (G), and three are pyrimidines, i.e., cytosine (C), thymine (T), and uracil (U), as shown below. To remember what are pyrimidines, remember "cut pyramid" where 'c' is cytosine, 'u' for uracil, 't' for thymine, and 'pyramid' for pyrimidines

Purines		Pyrimidines		
 adenine	 guanine	 cytosine	 thymine	 uracil

DNA contains four nitrogen bases, i.e., adenine, guanine, cytosine, and thymine. RNA also has four nitrogen bases, i.e., adenine, guanine, cytosine, and uracil. Note that the first three, i.e., adenine, guanine, and cytosine, are common in DNA and RNA, but the fourth, i.e., ti.e.ine in DNA, is replaced with uracil in RNA.

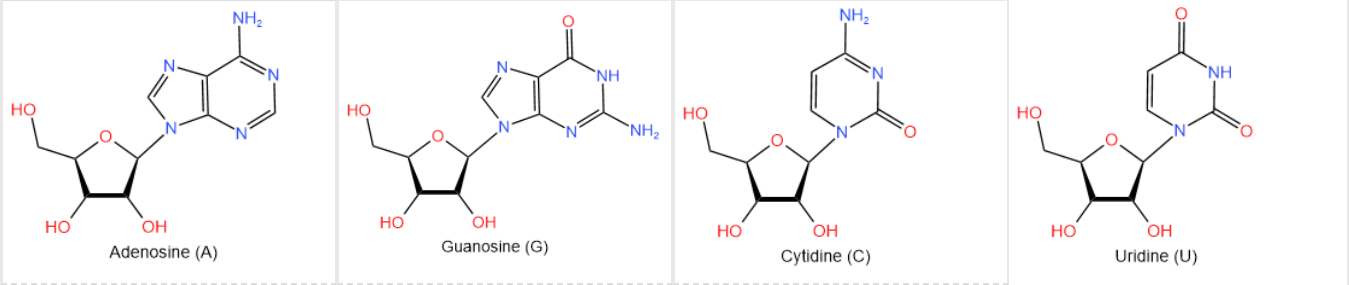
Nucleosides

When a monomer sugar like ribose or deoxyribose reacts with an amine, the $-OH$ group at C#1' is replaced with a N of the amine. The product is called N-glycoside, and the $C-N$ bond in the N-glycoside is called an N-glycosidic bond, as shown in the following example.

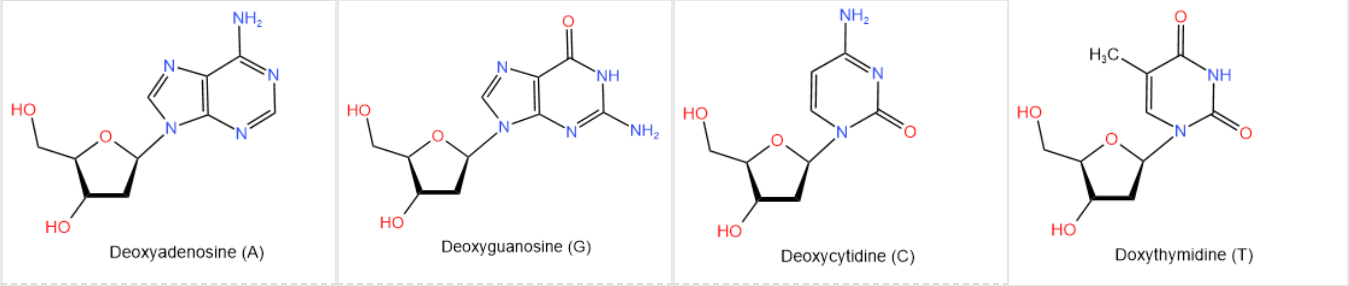


Purines connect through N#9 and pyrimidine through N#1 with anomeric carbon (C#1') of ribose or deoxyribose to form the N-glycoside.

The N-glycosides are named by using the name of purine but ending with -osine e.g., adenine becomes adenosine and guanine becomes gunosine, or by using the name of pyrimidine ending with -idine, e.g., cytosine becomes cytidine and uracil become uridine. One letter abbreviation of the nitrogen bases remains the same for the corresponding nucleoside, e.g., adenine (A) and adenosine (A). The four nucleosides found in RNA are shown below with their names and one-letter abbreviations.

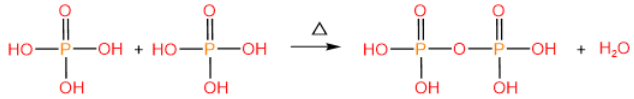


In DNA there is deoxyribose in place of ribose. So the names of nucleosides found in DNA begin with the deoxy- prefix. The four nucleosides found in DNA are shown below with their names and one-letter abbreviations.



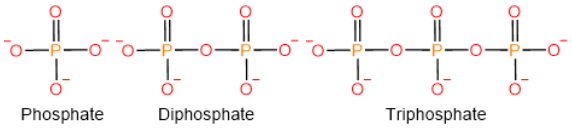
Phosphate

Phosphoric acid (H_3PO_4) is an oxyacid with three acidic protons, i.e., it is triprotic acid. Just as two carboxylic acids condense with each other and form a carboxylic acid anhydride with the elimination of water molecule, two phosphoric acids condense and form diphosphoric acid, also called pyrophosphoric acid, with the elimination of water.

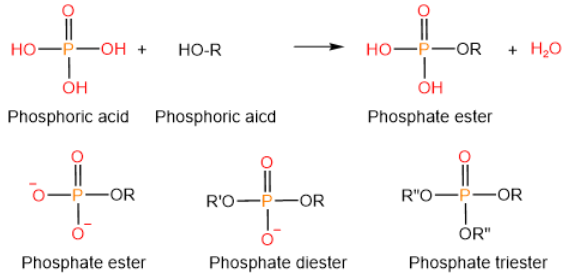


Similarly, three phosphoric acids can condense to form triphosphoric acid. Phosphoric acid, diphosphoric acid, and triphosphoric acid are shown below.

Under physiological conditions at pH ~7.4, the phosphoric acids lose proton and exist as phosphate anions, as shown below.

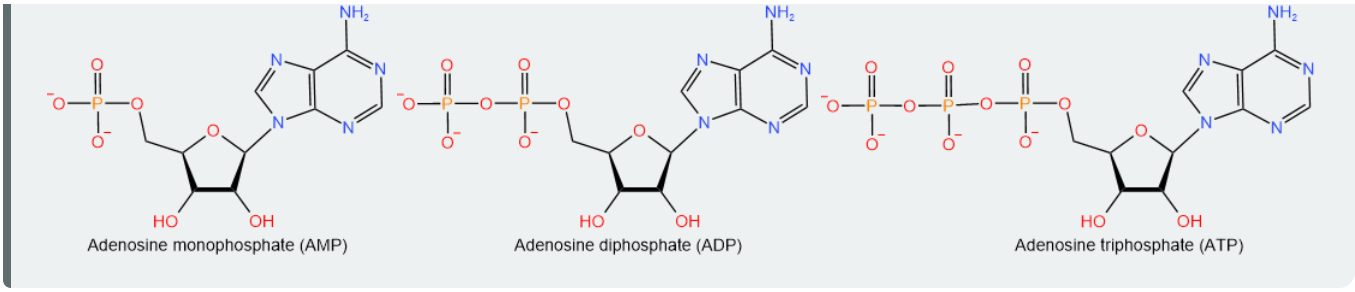


Phosphoric acids also condense with alcohols (R−OH) and form mono-, di-, or triesters, as illustrated below.



Nucleotide

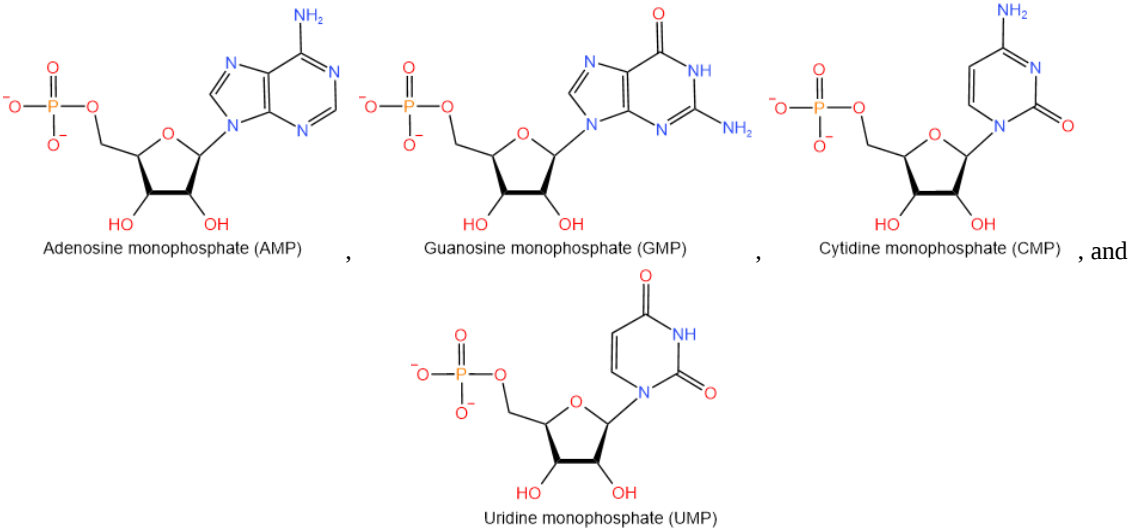
Phosphate esters of nucleosides are called nucleotides. The nucleotides found in nucleic acids are formed when −OH group at C#5' of ribose or deoxyribose makes ester with a monophosphate, diphosphate, or triphosphate, as shown below.



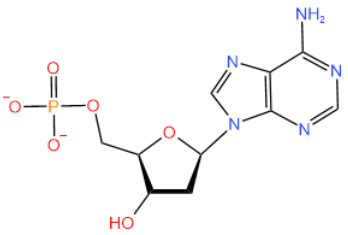
The names of the nucleotides are derived by using the name of the corresponding nucleoside with monophosphate, diphosphate, or triphosphate added at the end. Abbreviation of the nucleotide starts with the one-letter abbreviation of the nitrogen base followed by MP, DP, or TP depending on whether the phosphate group is monophosphate, diphosphate, or triphosphate, as shown in the example structures above.

The nucleotides like ADP and ATP are high-energy molecules used as energy currency in biochemical systems. For example, when ATP converts to ADP by releasing a phosphate, energy is released to do work or synthesize other compounds. Synthesis of nucleic acids begins with triphosphate esters that convert into monophosphate esters when incorporated in the nucleic acid polymer. The energy released in converting triphosphate to monophosphate drives the reaction forward.

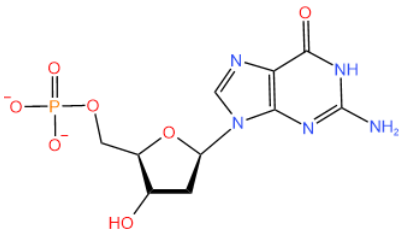
The four nucleotides found in RNA are shown below with their names and abbreviations.



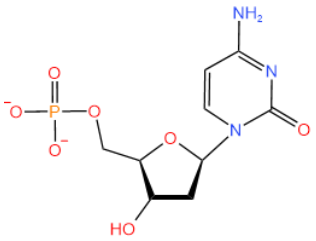
The nucleotides found in DNA are named similarly to those of RNA: start with the name of the corresponding nucleoside and end with monophosphate, diphosphate, or triphosphate. Remember that the names of DNA nucleosides begin with deoxy. The abbreviations of the nucleotides found in DNA start with a small alphabet *d* representing the deoxy- prefix of the nucleotides. The four nucleotides found in DNA are shown below with their names and abbreviations.



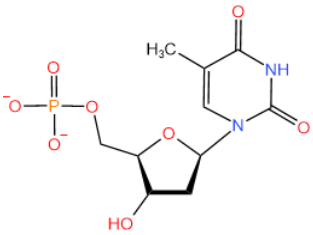
Deoxyadenosine monophosphate (dAMP) ,



Deoxyguanosine monophosphate (dGMP) ,



Deoxycytidine monophosphate (dCMP),



Doxythymidine monophosphate (dTMP)

8.2: Primary structure of deoxyribonucleic acid (DNA)

Learning Objectives

- Define nucleic acids and understand the mechanism of nucleic acid synthesis and the related terminologies.
- Define and be able to write the primary structure of nucleic acids.

Nucleic acid

Nucleic acids are long unbranched strands (polymers) of nucleotide subunits (monomers) that primarily store and express genomic information in living things.

Nucleic acids are long unbranched strands (polymers) of nucleotide subunits (monomers). A nucleotide triphosphate is the starting material for nucleic acid synthesis. The synthesis of DNA is controlled by an enzyme called DNA polymerase. A nucleoside triphosphate containing adenine (A), guanine (G), cytosine (C), or thymine (T) uses $-OH$ group at 3' position of deoxyribose to attack α-phosphorous of triphosphate group at 5' position of the next nucleoside in an SN_2 mechanism. It results in a **phosphodiester linkage** between the 3' position of the first and 5' position of the second nucleotide, and a pyrophosphate (diphosphate) is eliminated as shown in Figure 8.2.1. The pyrophosphate breaks down into two phosphate units which make the reaction irreversible.

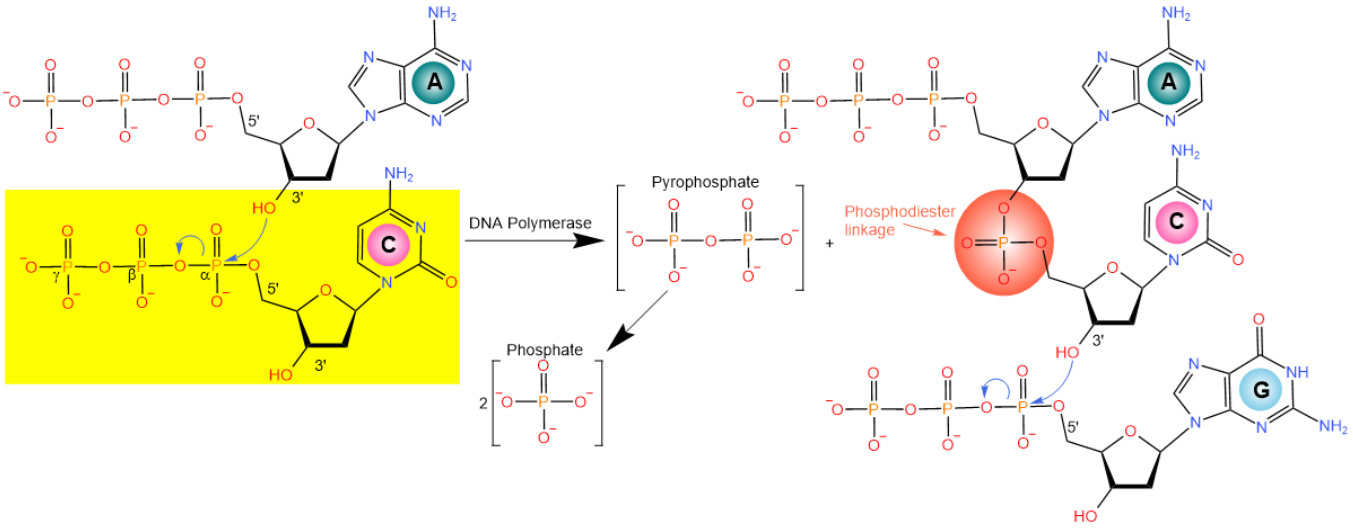


Figure 8.2.1: Chemical reaction mechanism for forming a phosphodiester linkage between nucleotides during DNA synthesis. (Copyright; Public domain)

Next, nucleotide triphosphate repeats the process and links at the 3' end of the growing strand; the cyclic process keeps repeating this way, as shown in Figure 8.2.1. The first nucleotide has unlinked triphosphate at the 5' position, called the **free 5' end**. The last nucleotide has an unlinked $-OH$ group at the 3' position, and it is called the **free 3' end**. DNA synthesis occurs from 5' to 3' as shown in Figure 8.2.2. Deoxyribose and phosphodiester linkage alternate in the polymer's backbone, and the nitrogen bases hang on the side (1' position) of the deoxyribose units. Two nucleotides linked by a phosphodiester linkage are called a **dinucleotide**, 3 to 10 linked nucleotides are called an **oligonucleotide**, and many nucleotides linked are called a **polynucleotide**.

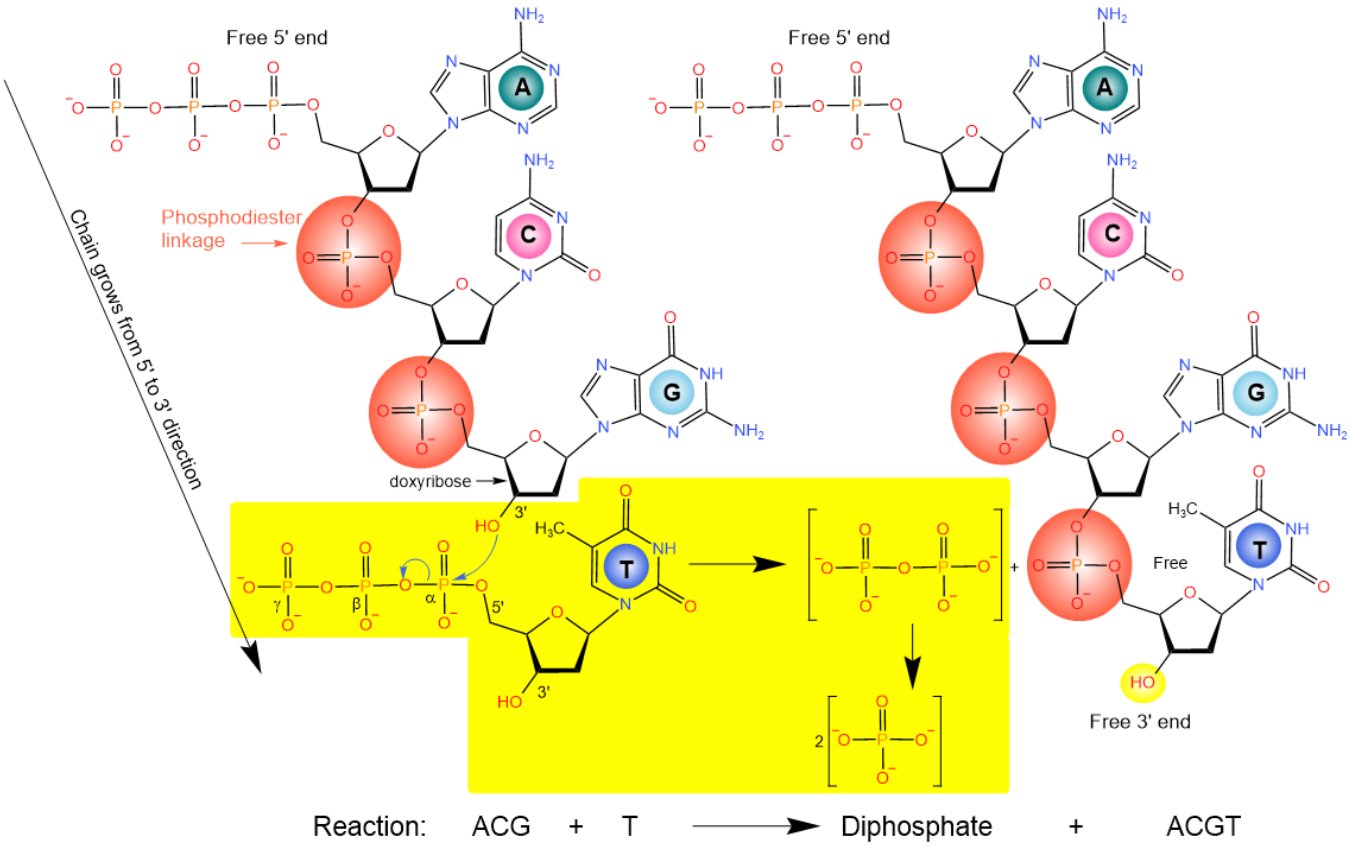


Figure 8.2.2: Synthesis of DNA illustrated. (Copyright; Public domain)

The synthesis of RNA takes place the same way as that of DNA, except for the following differences:

1. the enzyme carrying out the synthesis is RNA polymerase,
2. the sugar (5-carbon carbohydrate) is ribose, and
3. uracil (U) is used in place of thymine (T).

Primary structure of nucleic acid

The sequence of nucleotides in the nucleic acid is called the primary structure of nucleic acid.

The primary structure is written from the 5' to 3' direction, where the 5'-end is on the left end, and the one-letter abbreviation of the nitrogen base represents the nucleotides. For example, the primary structure of the oligonucleotide product shown in Figure 8.2.2 is ACGT. It is the sequence of nucleotides that carries the genetic information.

8.3: Secondary structure and replication of DNA

Learning Objectives

- Understand the double helix model and the replication process of DNA.

Secondary structure of DNA

Double helix model of DNA

It was observed from experimental data that although the composition of DNA varies from specie to specie, the amount of adenine (A) is always equal to the amount of thymine (T), and the amount of guanine (G) is always equal to the amount of cytosine (C). In 1953, Watson and Crick proposed a model of DNA with the help of Franklin's X-ray analysis. The Model, illustrated in Figure 8.3.1, is:

1. a DNA molecule is made up of two strands wound around each other like a spiral stair ladder,
2. the two strands run antiparallel, i.e., one from a 5' to 3' and the other from a 3' to 5' direction,
3. each strand has a backbone of alternating deoxyribose and phosphate, groups like the rails of the stair ladder, and
4. the two strands are connected at each nucleotide through hydrogen bonds between nitrogen bases: adenine associated with thymine of the second strand, and guanine connected with cytosine of the second strand, or vice versa, like a closed zip.

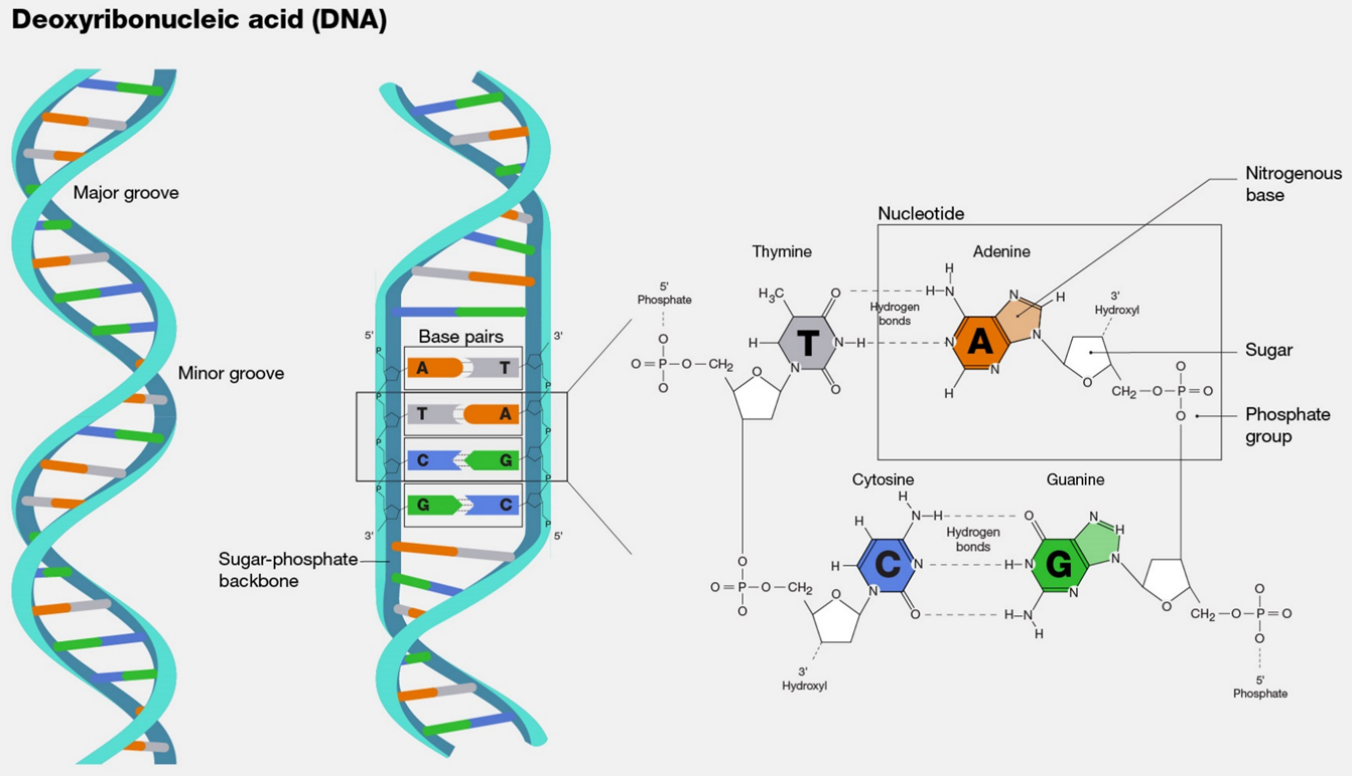


Figure 8.3.1: Double helix model of DNA (left), a section showing an overlay of alternating deoxyribose and phosphate backbone making the railing; adenine (A) paired with thymine (T) or guanine (G) paired with cytosine (C) making the steps of the double helix (middle); and a subsection enlarged to show the skeletal structure with T bonded with A by two hydrogen bonds and C connected with G by three hydrogen bonds (right). (Copyright; National Human Genome Research Institute, Public domain)

The logic of complementary base pairing

Two strands of DNA are complementary to each other as adenine (A) always faces thymine (T), and guanine (G) meets cytosine (C) in the other strand. The AT and GC are called complementary base pairs. The logic of this complementing is the following:

1. Purines, i.e., adenine (A) and guanine (G), are larger, and pyrimidines, i.e., cytosine (C) and thymine (T), are smaller. So, a purine in one strand pairs with a pyrimidine in the other strand to keep a nearly uniform gap between the two strands.

2. Adenine (A) makes two hydrogen bonds with thymine (T), and guanine (G) makes three hydrogen bonds with cytosine (C) in the other strand. The other possible combinations of purine and pyrimidine are unstable: A can have no hydrogen bond with C, and G can have only one with T.

The base pairing, i.e., AT and GC, complements the two strands. The multiple hydrogen bonding between each nucleotide of DNA holds the two strands together like a closed zip.

Replication of DNA

When a cell divides into two, each daughter cell needs a complete set of genetic information from the parent cell. The genetic information is in the form of the sequence of nucleotides in the DNA.

> DNA replication is the process by which the DNA is copied in a cell at the time of cell division.

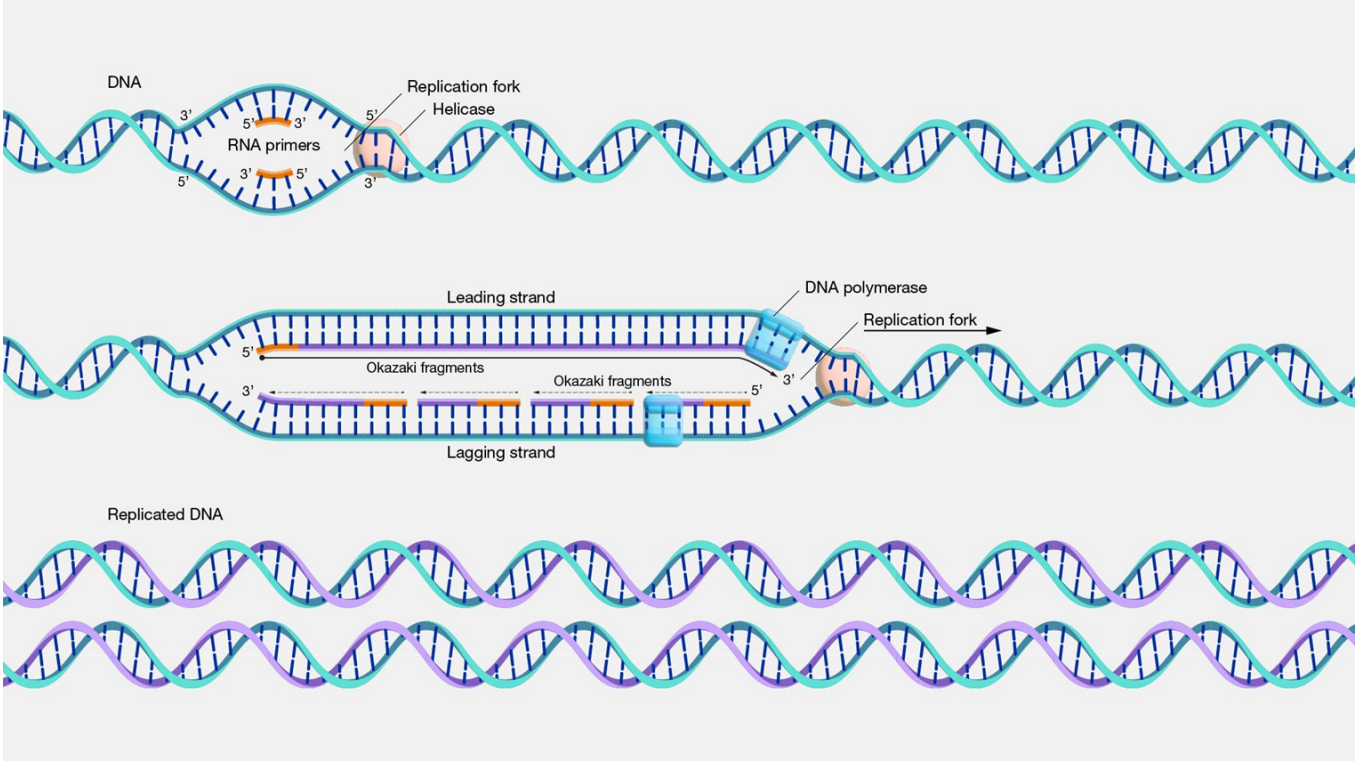


Figure 8.3.2: Illustration of DNA replication: Unzipping the double helix by enzyme helicase at the replication fork and initiation of the synthesis by RNA primers which are removed and replaced with DNA at the end (top), Continuous DNA strand synthesis in the leading strand and synthesis in fragments in the lagging strand from 5' to 3' directions (middle), and replicated DNA carrying one strand (blue) from the parent DNA and its complementary strand (purple) synthesized during the replication process. (Copyright; National Human Genome Research Institute, Public domain)

The process begins when an enzyme helicase catalyzes breaking hydrogen bonds, causing the DNA double helix to unzip, as illustrated by simulation in Figure 8.3.2 and explained in a video from yourgenome. Each strand acts as a temple for the synthesis of a new strand. As the complementary nucleotides come together, DNA polymerase joins the nucleotide by forming phosphodiester linkages from a 5' to 3' direction, as simulated in Figure 8.3.3. The leading strand is one strand of the initial DNA that unzips in a 3' to 5' direction. The leading strand's complementary strand is continually synthesized from the 5' to 3' direction. The other stand of the initial DNA that unzips in a 5' to 3' direction is called the lagging strand. The complementary strand of the lagging strand is synthesized in portions from 5' to 3' direction, called Okazaki fragments. An enzyme, called DNA ligase joins the fragments by forming phosphodiester linkages between them. This way, almost perfect copies of the entire DNA are synthesized. At this time, the cell can divide, and each daughter cell receives a replica of the parent DNA, which in many cases is identical to the DNA of the parent cell. The complementary base pairing in DNA ensures the correct placement of the nucleotides in the new DNA strands.

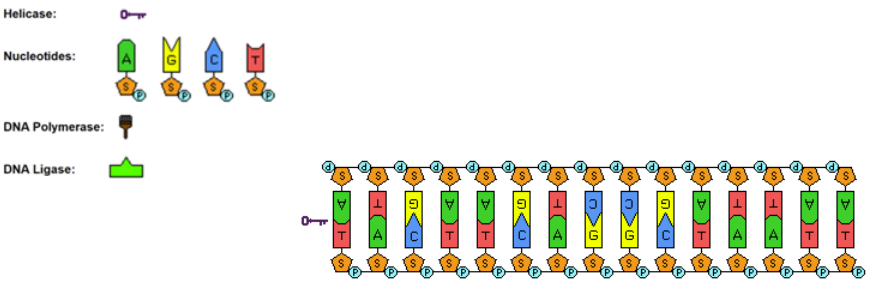
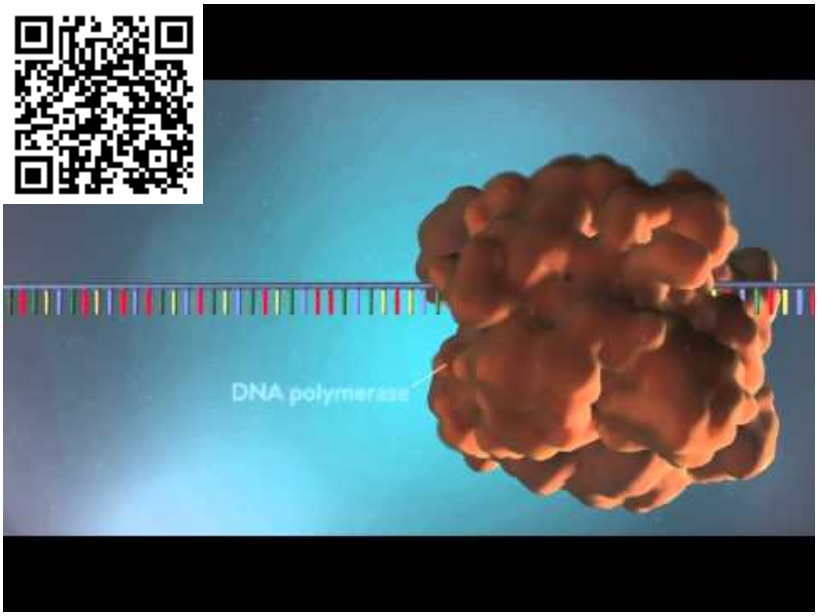


Figure 8.3.3: Simulation illustrating the DNA replication process. (Copyright: Steven Kuensting, CC BY-SA 4.0, via Wikimedia Commons)



8.4: Ribonucleic acid (RNA) and transcription

Learning Objectives

- Understand the basic structural features and functions of RNAs involved in protein synthesis, including mRNA, tRNA, and rRNA.
- Understand the transcription process, i.e., synthesis of mRNA from the DNA template, codons on DNA and mRNA, and anticodons on tRNA that specify amino acid to be incorporated in the protein.

Ribonucleic acid (RNA)

Ribonucleic acid (RNA) is a type of nucleic acid present in all cell types. It is structurally similar to DNA, as shown in Figure 8.4.1, but differs from it concerning the following.

1. RNA is often single-stranded.
2. The backbone of RNA is made of ribose units (rather than deoxyribose in DNA) connected by phosphodiester linkages.
3. Three nitrogen bases in RNA, i.e., adenine (A), Guanine (G), and cytosine (C), are the same as in DNA, but the fourth one is uracil (U) in RNA in the place of thymine (T) in DNA.

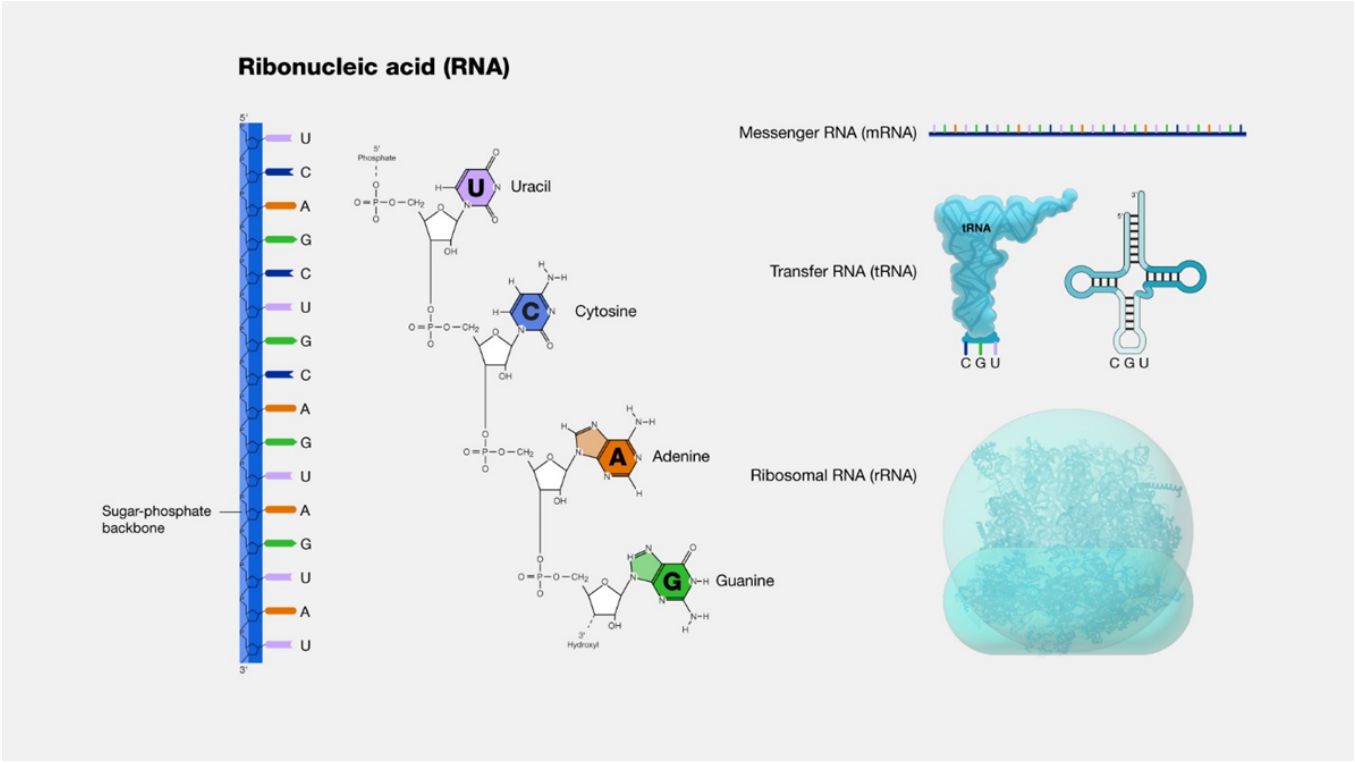


Figure 8.4.1: Illustration of a single-stranded ribonucleic acid (RNA) showing alternating phosphate-ribose backbone with nitrogen bases hanging out from ribose (left), the skeletal structure of a portion of an RNA (middle), and three types of RNA's involved in protein synthesis, i.e., messenger RNA (mRNA), transfer RNA (tRNA), and ribosomal RNA (rRNA) (right). (Copyright; National Human Genome Research Institute, Public domain)

Types of RNAs Involved in protein synthesis

Three types of RNA are involved in protein synthesis, i.e., messenger RNA (mRNA), transfer RNA (tRNA), and ribosomal RNA (rRNA), as illustrated in Figure 8.4.1. Some other RNAs have functions other than protein synthesis. Some viruses use RNA as genetic material rather than DNA.

Messenger RNA (mRNA)

Messenger RNA (mRNA) is a single-stranded RNAs as illustrated in Figure 8.4.1. It is synthesized as a complementary strand of a section of one of the two strands of DNA that carry information for synthesizing a protein cell needs. The message RNA is

synthesized in the nucleus and travels to the cytoplasm (watery interior of the cell outside the nucleus), where protein-synthesis machinery, i.e., ribosomes, reads the information and synthesizes the protein.

Transfer RNA (tRNA)

The transfer RNA (tRNA) is a small RNA comprising 70 to 90 nucleotides. Its job is to bring an amino acid to the protein synthesis location and incorporate it at a specific place in the protein, dictated by the codons on the mRNA. Hydrogen bonding in some complimentary bases in the tRNA makes it fold, producing loops and double-stranded portions, as illustrated in Figure 8.4.2. The 3D shape of the tRNA is twisted, appearing like L-shape, but it is usually depicted as a cloverleaf-like shape in 2D presentation. The loops include a D loop, a T loop, an anticodon loop, a variable loop that is a slight bulge between the T loop and the anticodon loop, and a receptor stem that contains the 3' end and 5' end of the tRNA molecule. The anticodon makes the tRNA specific for one amino acid. The anticodon is complementary to the codon on the mRNA.

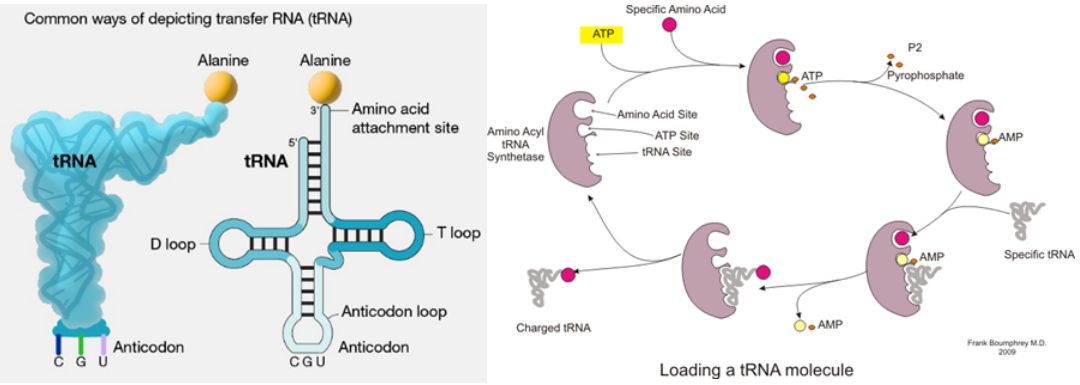
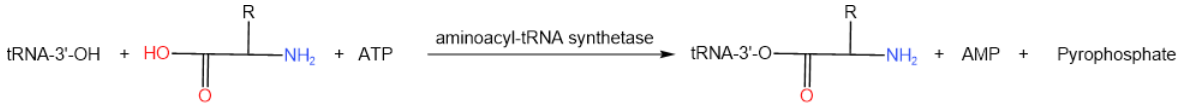


Figure 8.4.2: A 3D structure of a tRNA that is specific for amino acid alanine, its 2D depiction (left), and attachment of an amino acid to the free 3' −OH group of tRNA through an ester linkage using the energy released by ATP and catalyzed by enzyme aminoacyl tRNA synthetase (right). (Copyright; left: National Human Genome Research Institute, Public domain, right: Boumphreyfr, CC BY-SA 3.0, via Wikimedia Commons)

Activation of tRNA

Attaching an amino acid to tRNA is called activation of tRNA. The 3' end of tRNA always has adenosine (A) followed by two cytosines (C), i.e., ACC sequence. The free 3' −OH group of adenosine nucleotide makes an ester bond with an amino acid through the following reaction.



Anticodon

There is a different amino acyl-tRNA synthetase for each of the tRNA. Each tRNA is specific for one of the 20 amino acids. The particular shape of tRNA dictates this specificity. The tRNA and the amino acid corresponding to it fit in the active site of the corresponding amino acyl-tRNA synthetase, as illustrated in Figure 8.4.2. Each tRNA has a separate anticodon, a sequence of three nucleotides complementary to the corresponding codon on the mRNA, i.e., A pairs with U, G pairs with C, and vice versa. For example, mRNA codon GCA is complementary to the anticodon UGC on a tRNA shown in Figure 8.4.2 that specifies amino acid alanine in the protein.

Ribosomal RNA (rRNA)

Ribosomal RNA (rRNA) is the most abundant type of RNA that combines with proteins and makes ribosomes. Ribosome has two subunits, a large subunit, and a small subunit, as illustrated in Figure 8.4.1. Ribosome is the site for protein synthesis. There are many ribosomes in the cells.

Transcription

Gene is a section of DNA that is a basic unit of heredity passed from parent to child. Gene is a sequence of nucleotides in the DNA that encodes information for making specific proteins that lead to the expression of a particular physical character or

trait, such as hair color or eye color, or some other specific function in the cell. Humans have about 20,000 protein-coding genes. Some genes encode the synthesis of RNAs that have functions other than protein synthesis.

Transcription is the process of mRNA synthesis using a protein synthesis gene portion of a DNA strand as a template, as illustrated in Figure 8.4.3.

The process begins when the gene portion of the DNA unwinds, and RNA polymerase catalyzes mRNA synthesis from a 3' to 5' direction, using a strand of the unwound DNA going from a 5' to 3' direction as a template. The unwound gene portion of the DNA is called the **transcription bubble**. The process stops when the RNA polymerase reaches the sequence of nucleotides on the template DNA that is a stop single. The mRNA is released from the DNA, and the unwound portion of the DNA returns to its regular double helix structure. The mRNA is processed further in the nucleus. Then it leaves from the nucleus to the cytoplasm (watery interior of the cell outside the nucleus), where protein-synthesis machinery, i.e., ribosomes, reads the information and synthesizes the protein in a process called **translation**.

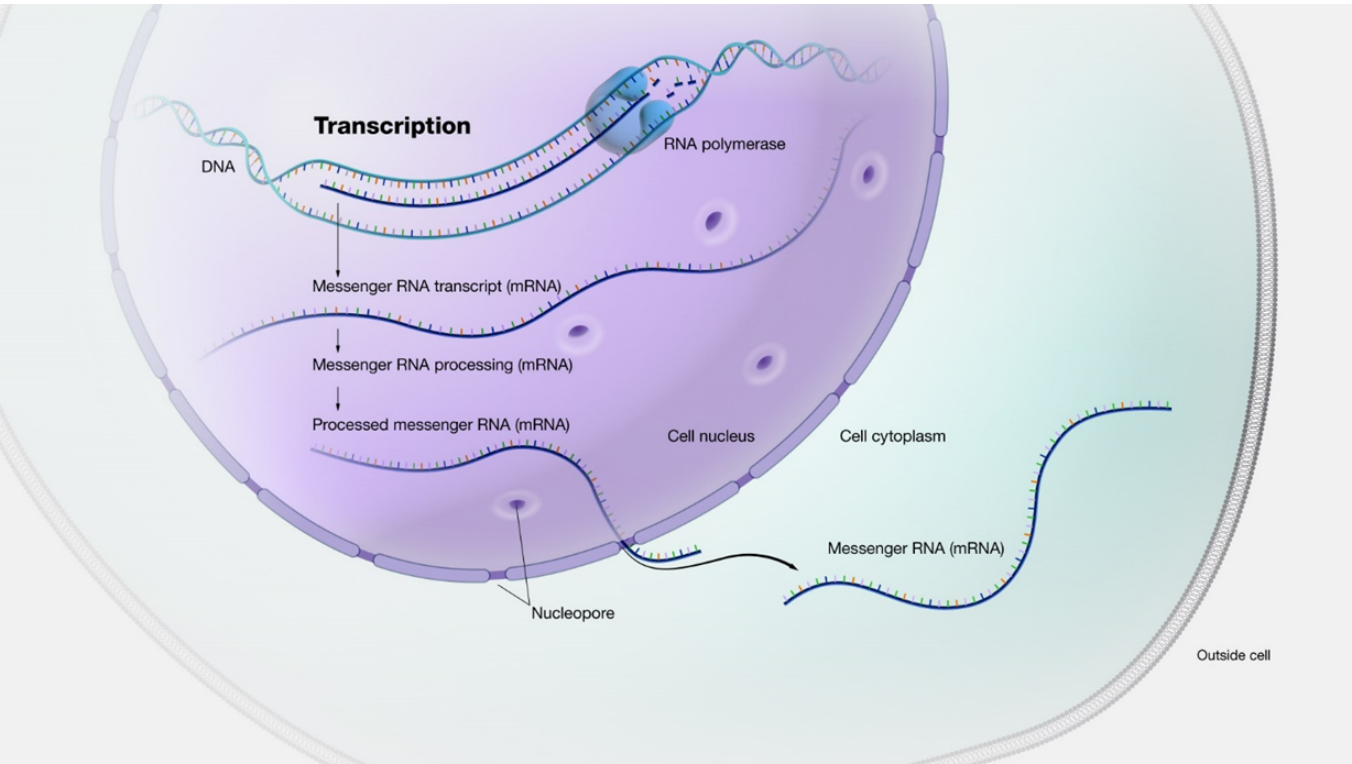


Figure 8.4.3: Illustration of the transcription process. (Copyright; National Human Genome Research Institute, Public domain)

The DNA strand used as a temple for mRNA synthesis s called the **template strand,** and the other is called the **information strand**. The mRNA complements the template strand, i.e., A in the template couples with U and G with C in the mRNA. The sequence of nucleotides in the mRNA is complementary to that of the template strand and the same as in the information strand, except that T in the information strand is replaced with U in the mRNA. Therefore, the primary structure of the information strand is shown as the primary structure of the DNA.

Codons

The gene encodes information about synthesizing a protein with the correct primary structure.

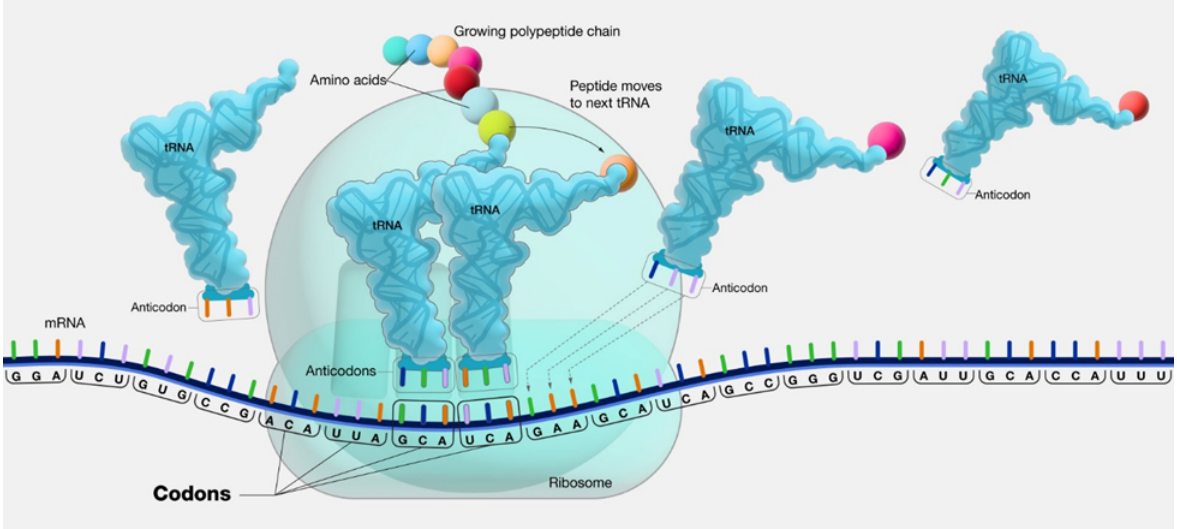


Figure 8.4.4: Illustration of codons on mRNA at the corresponding anticodons on tRNA for the synthesis of proteins. (Copyright; National Human Genome Research Institute, Public domain)

- A code called a **codon** is a sequence of three nucleotides (a trinucleotide) read from a 5' to 3' direction from mRNA, as illustrated in Figure 8.4.4.
- The codons:
  - start from the first appearance of AUG, i.e., adenine-uracil-guanine, reading from 5' to 3' direction on a mRNA,
  - are successive sets of frames of three nucleotides,
  - non-overlapping,
- Since there are four different nucleotides, there are 64 ways to spell out three-letter codons, as shown in Figure 8.4.5.
- Three codons, i.e., UGA, UAA, and UAG, are stop codons signaling the termination of protein synthesis. The remaining 61 specify amino acids.
- Since there are only 20 amino acids, amino acids have more than one codon, except tryptophan and methionine have one codon each. For example, four codons: GGU, GGC, GGA, and GGG, specify glycine.
- Codon AUG has two roles: i) in the middle of a series of codons, it specifies methionine, and ii) at the beginning of an mRNA, it signals the start of protein synthesis by bringing in methionine as the first amino acid, but the first methionine is usually removed later from the beginning (N-terminus) of the protein.

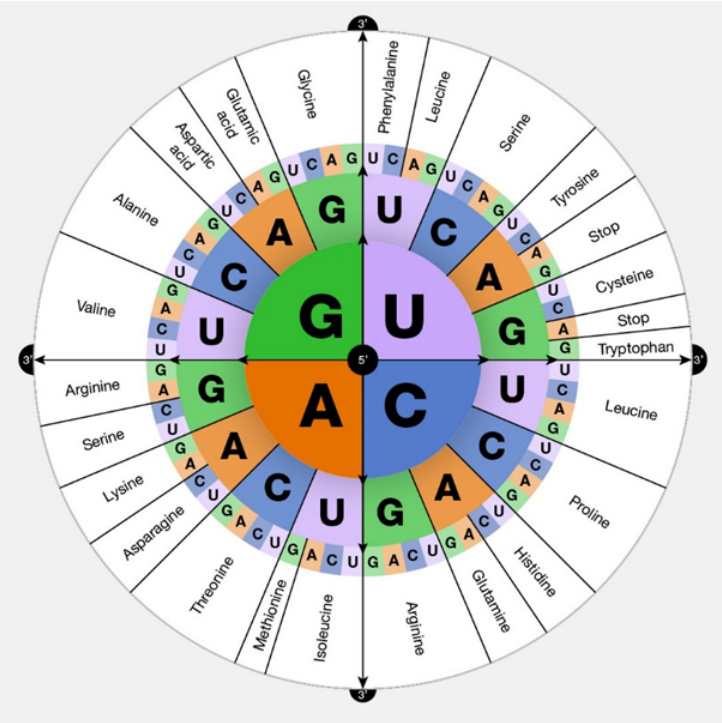


Figure 8.4.5: The 64 codons comprising three consecutive nucleotides in an mRNA read from 5' to 3' direction. The codon is read along the radius from the center to the periphery; e.g., UGG is a codon specifying tryptophan. (Copyright; National Human Genome Research Institute, Public domain)

8.5: Translation-Protein Synthesis

Learning Objectives

- Define and understand the mechanism of biological translation.

Translation

Translation in biology is the process of protein synthesis using the information encoded in mRNA.

Translation process

The process starts when the ribosome binds to mRNA in the cytoplasm, as illustrated in Figure 8.5.1. The ribosome moves along the mRNA from 5' to 3" direction until it reaches the stats codon AUG. The AUG is also the codon for methionine. So, the methionine-tRNA (methionine-loaded aminoacyl-tRNA) with the complementary anticodon UAC arrives and aligns opposite the codon AUG. The aminoacyl-tRNA with a complementary anticodon to the codon next to AUG on the mRNA arrives and aligns opposite the codon.

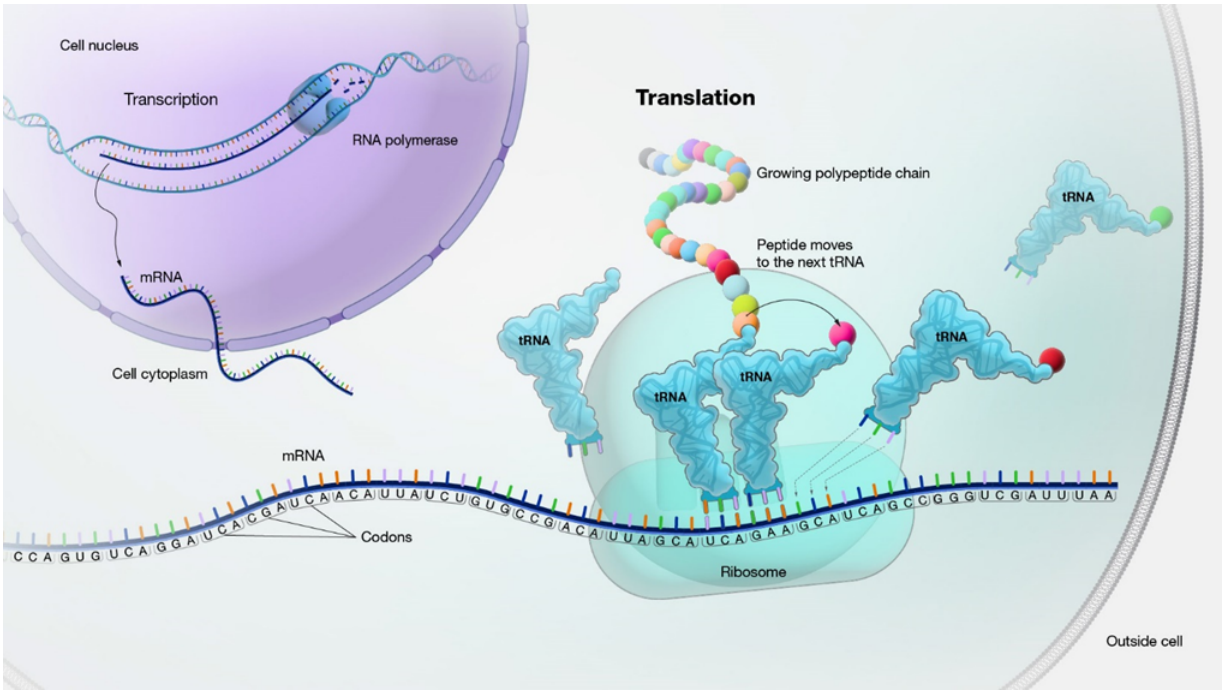


Figure 8.5.1: An illustration of the translation process. (Copyright; National Human Genome Research Institute, Public domain)

Ribosome catalyzes a peptide bond formation between $-NH_2$ group of the amino acid on the second tRNA with the carboxylate group of the first by nucleophilic acyl substitution mechanism. The first tRNA becomes empty, and the second tRNA becomes peptidyl-tRNA. The ribosome moves on to the next codon -a process called **translocation**. The empty-tRNA leaves and a new aminoacyl-tRNA with a complementary anticodon to the third codon align on the mRNA and repeat the above process, as illustrated in Figure 8.5.2. The cycle repeats, and the peptide keeps elongating until the ribosome reaches a stop codon. The translation process stops at this point, and the newly formed peptide is released. The first methionine is usually removed from the peptide. The empty tRNA are re-loaded with the amino acid later on, as illustrated in Fig. 8.4.2. The hydrogen bonding and the other intramolecular interaction, like salt bridges, disulfide bonds, etc., make peptides acquire the secondary, tertiary, and sometimes quaternary structure. This is how functional proteins with the appropriate forms are synthesized.

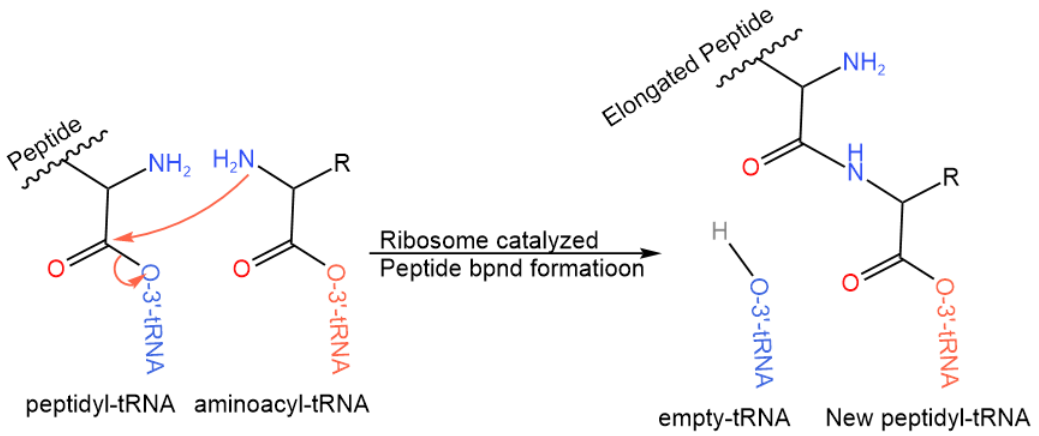


Figure 8.5.2: Illustration of the process of peptide elongation. (Copyright; author via source)

Figure 8.5.3 presents a simulation that summarizes the process of transcription and translation and the trinucleotide codons in DNA, mRNA, anticodons on tRNA, and the corresponding amino acids in the polypeptide that are shown in the simulation. Watch the DNA and RNA video summary by clicking this link or the video below.

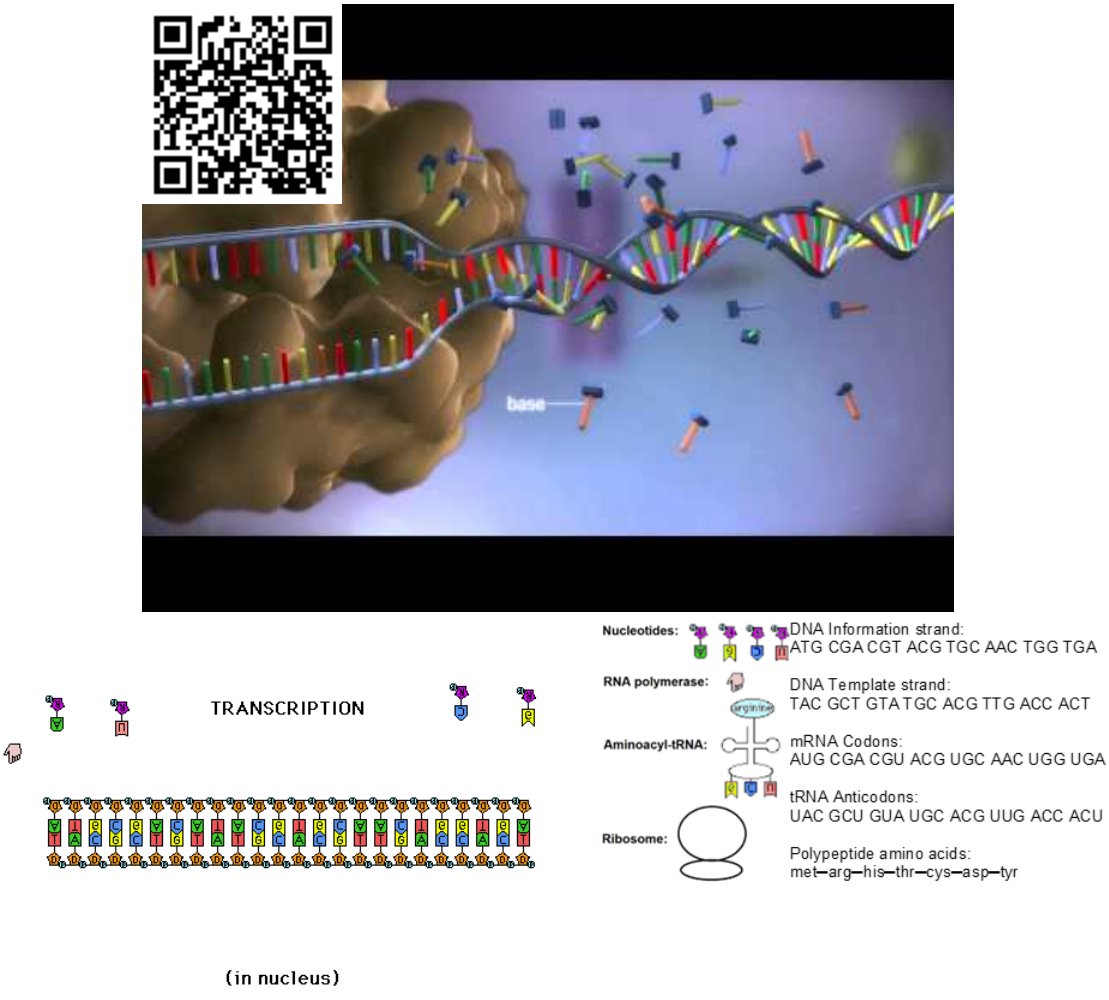


Figure 8.5.3: Simulation of the processes of transcription and translation (left), the corresponding ligands (middle), the corresponding codons, anticodon, and amino acids in DNA, mRNA, tRNA, and the polypeptide synthesized (rights). (Copyright; Steven Kuensting, CC BY-SA 4.0, via Wikimedia Commons)

Antibiotics that interrupt protein synthesis in bacteria

Antibiotics that interrupt protein synthesis in bacteria but not in humans are clinically useful. Some examples and their actions are shown in Table 1.

Table 1: Some antibiotics act by interrupting protein synthesis in bacteria but not in humans.

Antibiotic	Effect on protein synthesis in bacteria
Tetracycline	Prevents the aminoacyl-tRNA from binding to the ribosome
Erythromycin	Prevent the translocation of the ribosome along the mRNA
Streptomycin	Inhibits the initiation of protein synthesis
Chloramphenicol	Prevents the new peptide bond from being formed

8.6: Mutations and Genetic Diseases

Learning Objectives

- Understand mutation, its causes, types, and effects, including genetic diseases and cancer.

What is mutation?

A **mutation** is any change in an organism's DNA nucleotide sequence. It may occur due to an error during DNA replication, exposure to mutagens, or a viral infection. A **mutagen** is a physical or chemical agent that can cause a permanent change in an organism's DNA. Physical mutagens include exposure to radioactivity, X-rays, UV light, etc. Chemical mutagens are chemical agents that react with and change the DNA, e.g., polycyclic aromatic hydrocarbons (PAH) found in smoke and barbecued foods. Some mutagens that may lead to cancer are called **carcinogens**. Mutations may be harmful, beneficial, or may have no effect. **Somatic mutations** in body cells affect daughter cells but are not passed on to the offspring. **Germline mutations** occur in eggs or sperm and are passed to the offspring. When a mutation alters proteins or enzymes severely, the cells may not survive, and the person may have a genetic disease.

Types of mutation

A mutation may be due to i) replacement, ii) deletion, or iii) insertion of one or more nucleotides, as illustrated in Figure 8.6.1. There are many types of mutations. One subclass of mutation, i.e., related with change in one nucleotide is described here to introduce the basic terminologies related with mutaions.

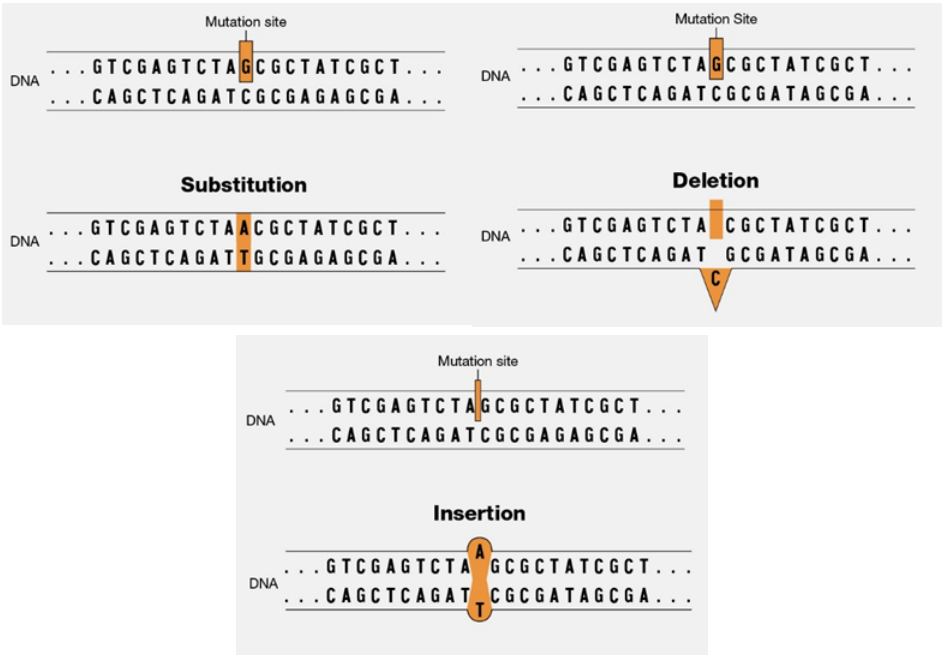


Figure 8.6.1: Illustration of three types of mutations, replacement or point mutation (left), deletion (middle), and insertion (right). (Copyright; National Human Genome Research Institute, Public domain)

Point mutation

Point mutation is the addition, deletion, or change (substitution) of one nucleotide with another. It changes the codon in which the nucleotide is located.

Substitution mutation

A replacement of one nucleotide with another in the DNA is substitution mutation. There are three sub-type of substitution mutations.

1. If the initial codon and the new codon represent the same amino acid, the mutation is called a **silent mutation**.
2. If a point mutation changes the codon to a different amino acid, it is called a **missense mutation**. If missense mutation replaces an amino acid with significantly different properties, it may cause disease. For example, sickle cell disease is caused by a point mutation of codon GAG for glutamic acid (acidic) to GTG for valine (nonpolar) in the hemoglobin gene. On the other hand, if the new amino acid is similar to the previous one, there may be little or no change rendered in the proteins. For example, a change from AAA for lysine to AGA for arginine may not affect the protein as both amino acids are basic amino acids.
3. If a point mutation changes the codon for an amino acid to a stop codon, it is called a **nonsense mutation**. It results in premature termination of protein synthesis rendering nonfunctional protein. For example, β-thalassemia is caused by a change from CAG for glutamine to UAG, i.e., stop signal. The missense and nonsense mutations are illustrated in Figure 8.6.2

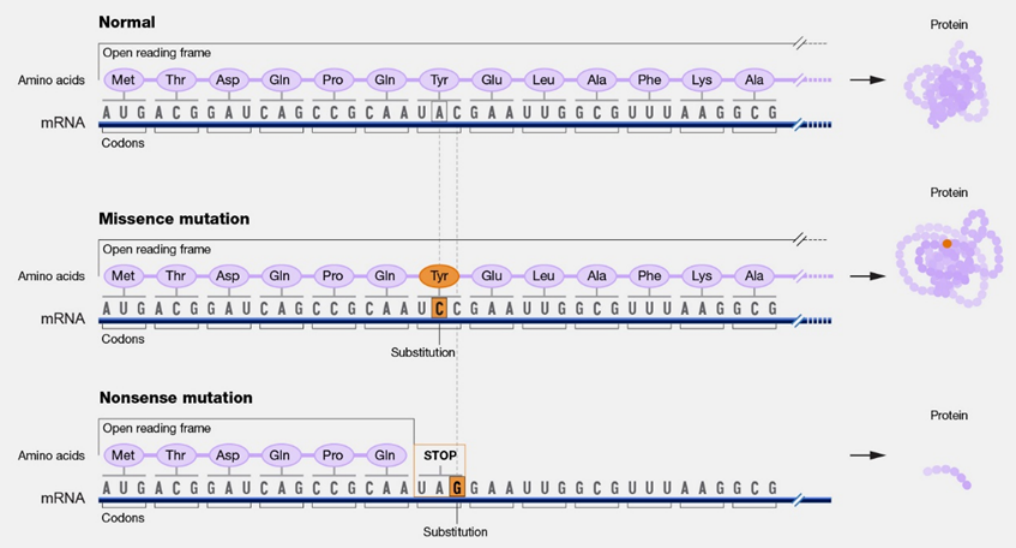


Figure 8.6.2: Misence and nonsense mutation illustrated. (Copyright; National Human Genome Research Institute, Public domain)

Deletion mutation

Deletion mutation is the loss of one or more nucleotides from a segment of DNA. Since the codon is three consecutive nucleotides, the loss of one nucleotide will not only change the codon from which the nucleotide is lost, but all the codons following it will be changed. Deletion mutation is related to a significant number of genetic diseases, e.g., two-thirds of cystic fibrosis cases are due to the loss of three nucleotides that results in the loss of amino acid phenylalanine in the protein involved, and cat cry syndrome is due to a partial chromosome deletion.

Insertion mutation

In insertion mutation, one or more nucleotides are inserted into the typical sequence of nucleotides in the DNA. So, the codon where insertion happens and all the codons following it are changed.

Effects of mutations

Some mutations do not cause a significant change in protein structures, allowing the protein to perform its function. Others may change an amino acid vital to the structure of the proteins with an amino acid of significantly different properties. It results in the proteins not being able to function correctly. For example, sickle cell disease is caused by replacing glutamic acid with a nonpolar valine that prevents the hemoglobin protein from working correctly. If the protein is an enzyme, it may no longer catalyze the reaction. In this situation, the reactants may accumulate and become poisonous, or the product may be vital for survival and not synthesized. For example, an enzyme required to metabolize galactose-1-phosphate is absent in galactosemia. It results in the accumulation of galactose-1-phosphate, which may cause cataracts and mental retardation.

Genetic disorder

Genetic disorders are health conditions caused by mutations in the genetic material. For example, phenylketonuria (PKU) results when DNA can not carry the correct codes for the synthesis of the enzyme phenylalanine hydroxylase that hydroxylates the phenyl ring of phenylalanine to convert it to tyrosine. When phenylalanine cumulates, other enzymes convert it to phenylpyruvate—

cumulation of phenylalanine and phenylpyruvate cause mental retardation. If diagnosed early on, a diet can be prescribed to the child that avoids foods containing phenylalanine. It can prevent the buildup of phenylpyruvate.

Similarly, if the enzyme that converts tyrosine to melanin is not functioning, melanin is not produced, causing albinism. Melanin is a pigment that gives color to skin and hair. People with albinism do not have skin, eye, or hair pigments. This disease also happens in animals. When a genetic disorder is inherited from one or both parents, it is called a **hereditary disease**. A few more genetic diseases are listed below.

- **Cystic fibrosis** results from a defective gene. It affects the lungs and digestive system. People with this disease can not digest food properly and have repeated lung infections.
- **Down syndrome** is a genetic disease in which peoples have cognitive impairment that may be mild or severe.
- **Hemochromatosis** is an inherited condition where the body absorbs and stores so much iron that it can lead to organ damage.
- **Haemophilia** is a genetic disorder in which blood does not clot properly. It makes bleeding challenging to control.
- **Huntington's disease** is a genetic disorder that affects the nervous system, and the effect worsens over time.
- **Tay-Sachs disease** is a genetic disorder that causes brain damage.
- **Duchenne muscular dystrophy** is a condition that causes gradual loss of muscle function.
- **Thalassemia** is a genetic disorder that causes less hemoglobin to be produced, making the blood cells small and pale.
- **Tourette syndrome** is a genetic disorder related to neurological problems. Peoples with this disorder make involuntary vocal sounds and movements. Relaxation and exercise may reduce the symptoms.

Cancer

Cancer is a disease in which some of the body's cells grow and multiply uncontrolled. Mutation in a cell may cause this uncontrolled division of cells. The uncontrolled division and growth of cells appear as a tumor, as illustrated in Figure 8.6.3. It is benign if the tumor remains limited in growth without harming the neighboring tissue. However, the tumor is cancer if it invades other issues and interferes with their normal functions.

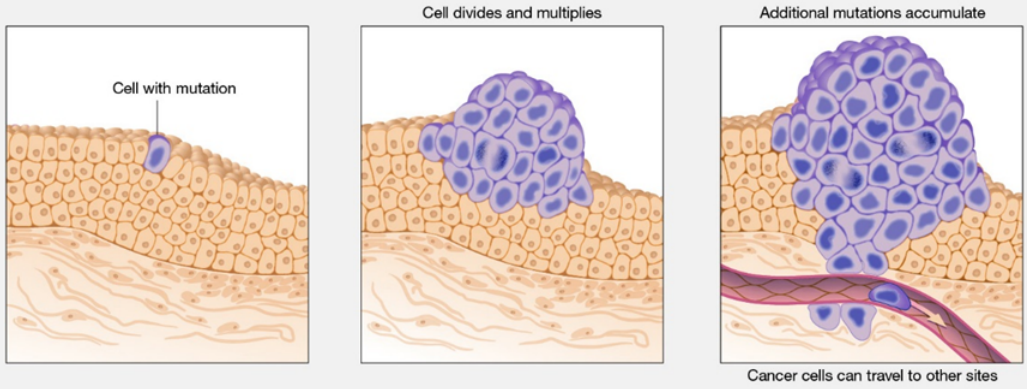


Figure 8.6.3: Illustration of tumor and cancer. (Copyright; National Human Genome Research Institute, Public domain)

The most common types of cancer are lung cancer, prostate cancer, breast cancer, colorectal cancer, stomach cancer, cervical cancer, and skin cancer. Melanoma and skin cancers account for about 40% of cancer cases. Cancer types common in children include acute lymphoblastic leukemia, brain tumors, and non-Hodgkin lymphoma.

The mutations causing cancers can result from errors during DNA replication or exposure to **carcinogens**, i.e., the substances or radiations that cause cancer. The **carcinogenic substances** include some compounds in tobacco smoke, automobile exhaust fumes, processed meat; asbestos; benzene; toxic metals like nickel, arsenic, beryllium, chromium, and cadmium and their compounds; nitrosamine; ethylene oxide, etc. About 22% of cancer-related deaths occur due to smoking. **Cancer-causing radiations** include radioactivity resulting in ionizing radiations, X-rays, and UV-radiations. About 15% of cancers are due to **cancer-causing viral infections** like HIV, hepatitis B, hepatitis C, Helicobacter pylori, Epstein-Bari virus, papillomavirus, etc. Sometimes inherited defects in genes are the cause of cancer. These factors cause or are at least partially involved in gene mutations that result in cancer.

8.7: Viruses

Learning Objectives

- Define viruses and understand how they infect and multiply, including reverse transcription.
- Understand the basic structural features, infection ways, and treatment of two viral infections, including covid-19 and HAIDS.

What is a virus?

A **virus** is a tiny infectious microbe that consists of a genetic material (DNA or RNA) surrounded by a protein coat called the **capsid** that covers and protects the genetic material.

A single virus outside the cell that has genetic material and capsid surrounding it is also called a **virus particle** or a virion**.** As Figure 8.7.1 illustrates, viruses have various shapes.

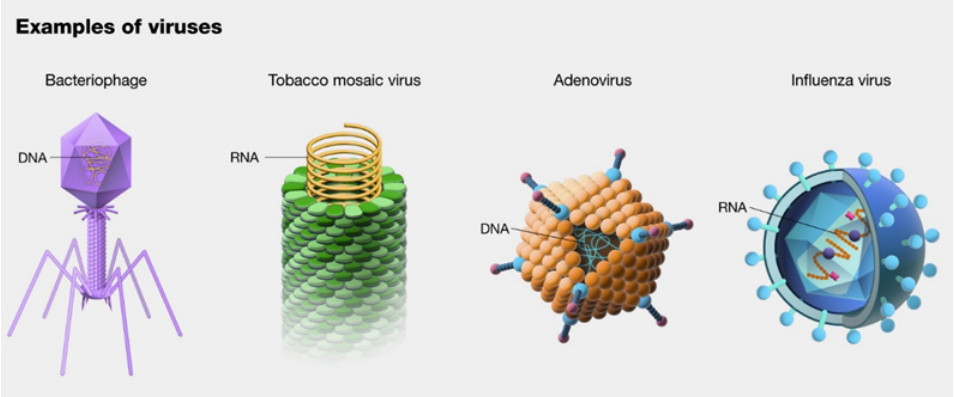


Figure 8.7.1: Some examples of viruses having genetic material (DNA or RNA) surrounded by protein coats in various shapes. (Copyright; National Human Genome Research Institute, Public domain)

A virus does not have the machinery to replicate and synthesize proteins, so it injects its genetic material into a host cell and hijacks its nucleic acid and protein synthesis machinery to reproduce. Viruses infect nearly all life forms, including humans, animals, plants, bacteria, and fungi. Figure 8.7.2. illustrates some of the human viral infections.

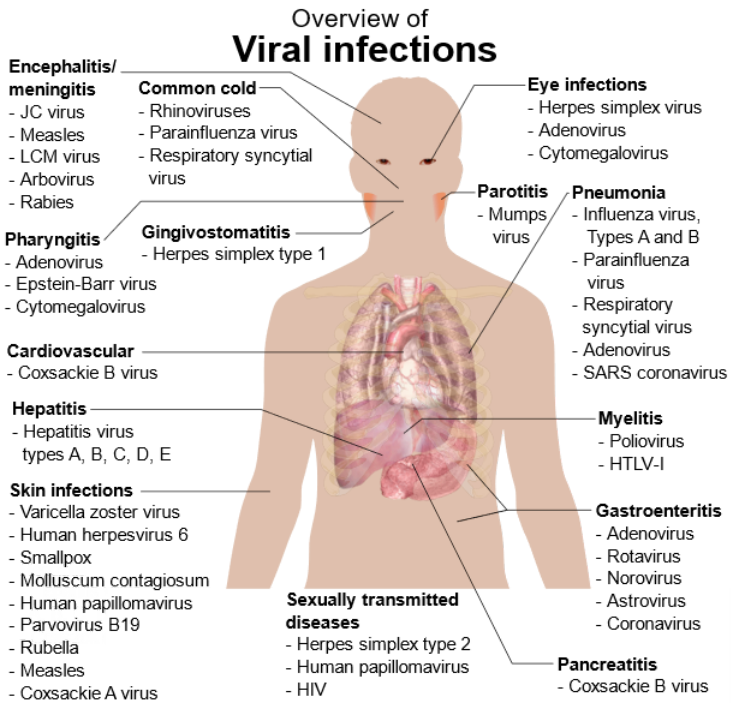


Figure 8.7.2: A simplified overview of human viral infections. (Copyright; Mikael Häggström, Public domain via Wikimedia Commons)

How do viruses infect and multiply?

The virus infection of cells begins with the attachment of the virus to the receptors at the host cell surface, followed by penetrating the cell membrane or cell wall. The protein coat (capsid) is removed, and the genetic material is injected into the host cell. The host cell replicates the viral DNA producing copies of it, and also translates it, making the proteins that the newborn virus needs. New viruses are assembled using the viral DNA and the protein coat produced by the host cell. They ultimately kill the host cell by bursting out of it in a lysis process, as illustrated in Figure 8.7.3. Some viruses take a portion of the host cell membrane to form an envelope around the capsid during lysis.

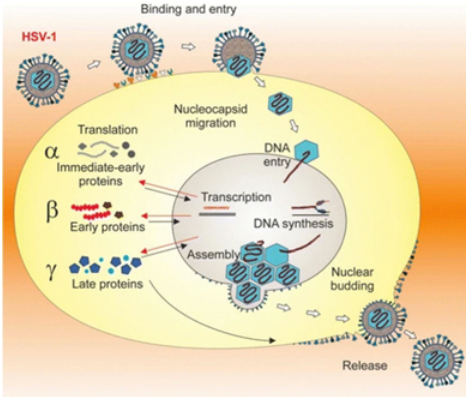


Figure 8.7.3: Illustration of virus multiplication using the host cell's machinery to replicate and translate its genetic material. (Copyright; Giovanna De Chiara, Maria Elena Marcocci, Rossella Sgarbanti, Livia Civitelli, Cristian Ripoli, Roberto Piacentini, Enrico Garaci, Claudio Grassi, Anna Teresa Palamara, CC BY 2.5, via Wikimedia Commons)

Reverse transcription

Viruses that contain viral RNA as their genetic material are called **retroviruses**. Once inside the host cell, the retroviruses first synthesize DNA using the viral RNA as a template, the reverse transcriptase enzyme in the virus, and the nucleotides in the host cell. The single DNA strand forms a double-strand viral DNA called a **provirus**. The provirus integrates with the DNA of the host cell using enzyme ligase, which is also present in the virus. Then the viral DNA directs the synthesis of viral RNA and the proteins

needed to assemble new virus particles, as illustrated in Figure 8.7.4. Finally, the newborn virus particles burst out of the host cell and infect other cells.

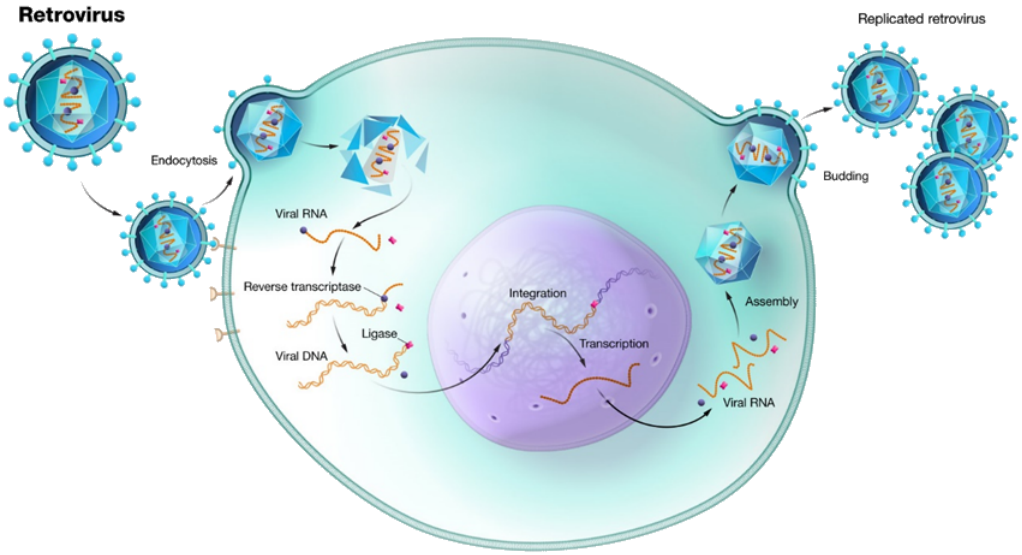


Figure 8.7.4: Illustration of reverse transcription of retroviruses. (Copyright; National Human Genome Research Institute, Public domain)

Examples of viral infections

Viruses are responsible for various human infections, as illustrated in Figure 8.7.2, and also in animals, plants, fungi, and bacteria. Two well-known example diseases caused by viruses, COVID-19 and AIDS, are described briefly here.

COVID-19

Coronavirus disease 2019, commonly known as COVID-19, is caused by SARS-CoV-2, which was discovered in Wuhan, China, in December 2019 and has spread worldwide. SARS-CoV2 is a member of a family of viruses that cause various diseases, from head or chest colds to more severe conditions like severe acute respiratory syndrome (SARS). SARS-CoV-2, also called coronavirus, spreads through droplets expelled or projected out of the mouth or nose of an infected person during breathing, coughing, sneezing, or speaking.

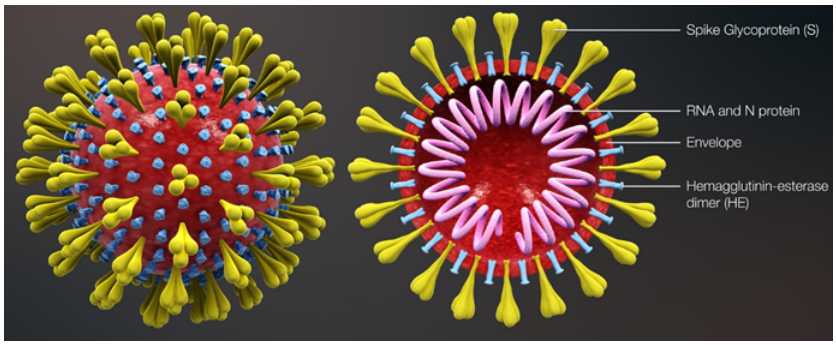


Figure 8.7.5: Images combined from a 3D medical animation depicting the coronavirus's shape and the cross-sectional view. The image shows the major elements, including the Spike S protein, HE protein, viral envelope, and helical RNA. (Copyright; https://www.scientificanimations.com, CC BY-SA 4.0, via Wikimedia Commons)

The common name coronavirus refers to the crown-shaped spike proteins sticking out of the virus, as illustrated in Figure 8.7.5. Coronavirus uses these spike proteins to attach to the host cells to infect them. Like other viruses, it replicates by reverse transcription inside the host cell, as illustrated in Figure 8.7.6. Genetic changes happen to the virus over time, resulting in variants that have different attributes regarding how fast the virus spreads or how severe the illness it causes. The virus and its variants are constantly monitored to update or improve its treatment as it changes over time.

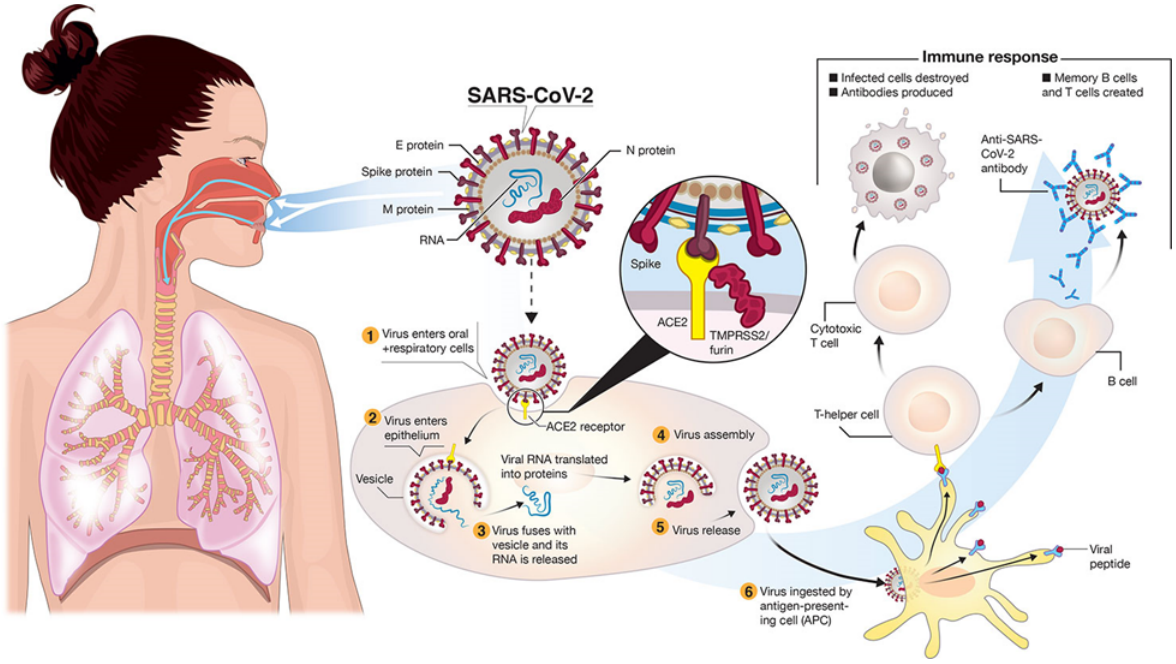


Figure 8.7.6: A simplified depiction of the virus's life cycle and potential immune responses elicited. (Copyright; Colin D. Funk, Craig Laferrière, and Ali Ardakani. Graphic by Ian Dennis - http://www.iandennisgraphics.com, CC BY 4.0, via Wikimedia Commons)

According to CDC, some antibodies can protect from coronavirus by targeting these spike proteins, but the best protection is vaccines. Several COVID-19 vaccines have been developed. The most common COVID-19 vaccines, i.e., Pfizer and Moderna vaccines, use either self-replicating RNA or mRNA, which cause the cells to produce SARS-CoV-2 spike protein. This process teaches the body's immune system how to identify and destroy the pathogen. RNA vaccines usually use modified nucleotides in mRNA.

AIDS

Acquired immunodeficiency syndrome (AIDS) is the late stage of the **human immunodeficiency virus** (HIV) when the virus badly damages the body's immune system. HIV infection is transmitted through sexual activity, blood transfusion, mother to child, etc., and primarily destroys $CD4^+$ T cells, components of the human immune system.

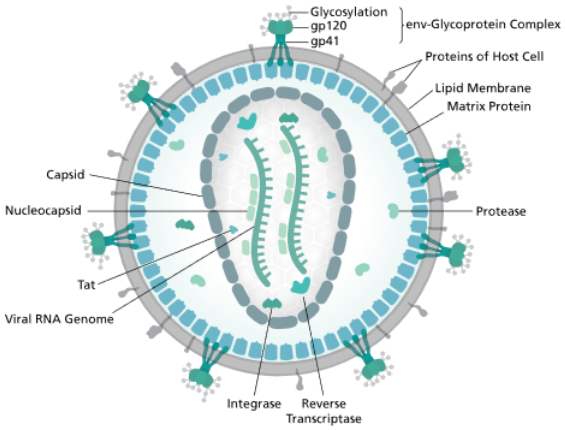


Figure 8.7.7: Diagram of the HIV virion. (Copyright; Thomas Splettstoesser (www.scistyle.com), CC BY-SA 4.0, via Wikimedia Commons)

HIV is a spherical retrovirus, as illustrated in Figure 8.7.7. It has two copies of RNA enclosed by a nucleocapsid and bound to enzymes needed for its development surrounded in a conical capsid. It also has an envelope of lipid bilayer taken from the cell membrane of the host cell. It also has proteins from the host cell, like glycoproteins, gp120 and gp41, that allow it to attach to the

target cell to begin the infection. Destruction of $CD4^+$ T cells by HIV infection makes the person's immune system ineffective. It increases the risk of common diseases such as tuberculosis, tumors, pneumonia, skin cancer, etc., that are rare, as illustrated in Figure 8.7.8.

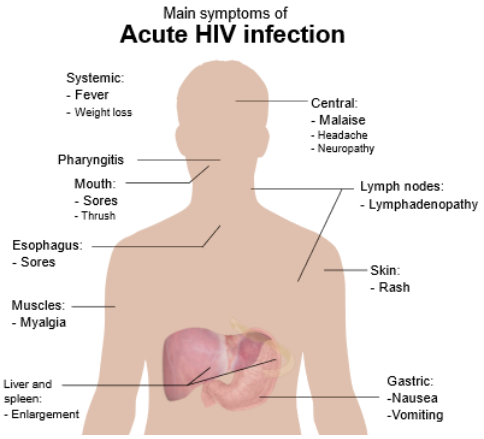
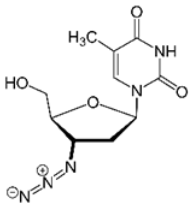
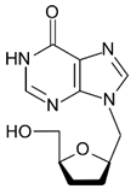
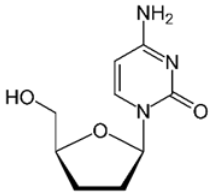
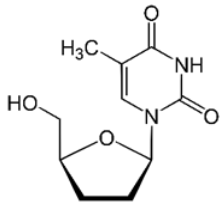


Figure 8.7.8: Main symptoms of acute HIV infection. (Copyright; Mikael Häggström, Public domain, via Wikimedia Commons)

There is no cure for AIDS, it stays for life long, but its progress can be slowed down by treatments that interfere with the life cycle of HIV at different stages. These include drugs that inhibit entry, reverse transcription, and translation. Entry inhibitors include enfuvirtide (Fuzeon) and maraviroc (Selzentry). Transcription inhibitors include nucleosides analogous to natural nucleosides, e.g., 3'-azido-2'deoxythymidine (AZT), 2',3'-dideoxyinosine (ddC), 2',3'-dideoxycytidine (ddC) and 2',3'-didehydro-2'3'-dideoxythymidine (d4T), illustrated below. They incorporate into viral DNA and prevent the formation of a sugar-phosphate backbone as they do not have $-OH$ groups at the 3' position of the sugar needed for this purpose. The translation inhibitors include saquinavir (Invirase), ritonavir (Norvir), fosamprenavir (Lexiva), etc.

			
3'-azido-2'deoxythymidine (AZT)	2',3'-dideoxyinosine (ddC)	2',3'-dideoxycytidine (ddC)	2',3'-didehydro-2'3'-dideoxythymidine (d4T)

8.8: Genetic engineering

Learning Objectives

- Understand recombinant DNA technology and its applications in biomedical technologies.
- Understand PCR and its applications in genetic testing, fingerprinting, and human genome project.

Recombinant DNA

DNA can be isolated from the cell and cut at a specific place using restriction enzymes. Fragments of DNA from different sources, e.g., an insulin gene from a human and a DNA from Escherichia coli (E. coli) cut using the same restriction enzyme, can be combined into a new **recombinant DNA**. The recombinant DNA can be introduced into the cell, e.g., back into E. coli, where it propagates using the cellular machinery of the host cell. It is then used to produce proteins of interest, e.g., human insulin used to treat diseases like diabetes.

How is recombinant DNA produced?

The DNA of host cells, e.g., small circular plasmids of DNA from E. coli, are isolated first. The host cells are soaked in a detergent solution that disrupts the plasma membrane and releases the plasmids. A restriction enzyme is used to cut the DNA of the host cell at a specific location. The same enzyme is used to cut a piece of donor DNA, e.g., an insulin gene from a human. The cut ends of the DNA are called **sticky ends**. When the cut donor and host DNA are mixed, they join at the sticky ends, as illustrated in Figure 8.8.1. The resultant recombinant DNA is introduced to a new culture of E. Coli. The host cells that have taken up the recombinant DNA are selected and grown. When these bacteria grow and divide, they produce the protein encoded in the inserted gene, i.e., insulin in this example. Insulin is extracted, purified, and used to treat diabetes.

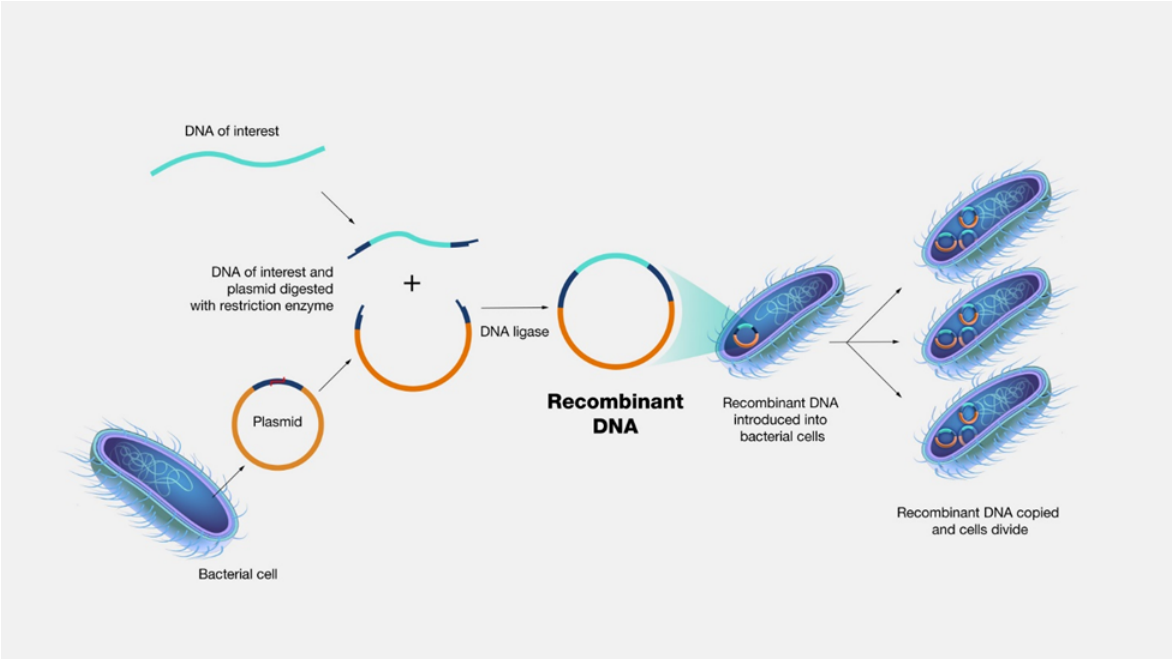


Figure 8.8.1: Illustration of a recombinant DNA formed by inserting a DNA fragment of interest with the plasmid of a bacterial cell, introducing it to the host cells where it replicates and produces the desired protein using the cell machinery fo the host cell. (Copyright; National Human Genome Research Institute, Public domain)

Some applications of recombinant DNA technology

The recombinant DNA technology is used in food production, medicine, agriculture, and bioengineering. Some examples are listed below.

- **Chymosin** is an enzyme used to manufacture chees. Currently, ~60% of hard cheese is produced in the US using genetically engineered chymosin.

- **Insulin** is usually synthesized by inserting the human insulin gene into E. Coli, or yeast, produces insulin processed and used to treat diabetes.
- **Human growth hormone** (HGT) is needed for patients whose pituitary glands do not produce sufficient hormones. Recombinant HGT is commonly used for this purpose.
- **Blood clotting factor VIII** is needed for patients suffering from the bleeding disorder hemophilia. It is produced these days by recombinant DNA technology.
- **The Hepatitis B vaccine** needed to treat hepatitis B infection is produced by recombinant DNA technology.
- **The recombinant HIV protein** tests the presence of antibodies that may have been produced in response to an HIV infection.
- **Herbicide-resistant crops**, including soy, corn, canola, alfalfa, cotton, etc., have been developed that contain recombinant genes resistant to the herbicide glyphosate found in Roundup. It allowed the use of Roundup to control weeds without affecting the crop.
- **Insect-resistant corps**: *Bacillus thuringeiensis* is a bacteria that naturally produces a protein with insecticidal properties. Crops containing the recombinant gene from the bacteria hold promise to control insect predators without using an insecticide.
- **Interferon** produced by recombinant DNA technology is used to treat cancer and viral disease.
- **The influenza vaccine** is used to prevent influenza.

Polymerase chain reaction (PCR)

polymerase chain reaction (PCR) quickly produces millions of copies of a DNA segment of interest. The selected DNA segment is heated to denature and separate the two strands and mixed with DNA polymerase, a short synthetic DNA called primer to the selected segment to be amplified, and nucleotides, to produce a complementary strand. The process is repeated several times to make millions of copies of the desired DNA fragment, as illustrated in Figure 8.8.2.

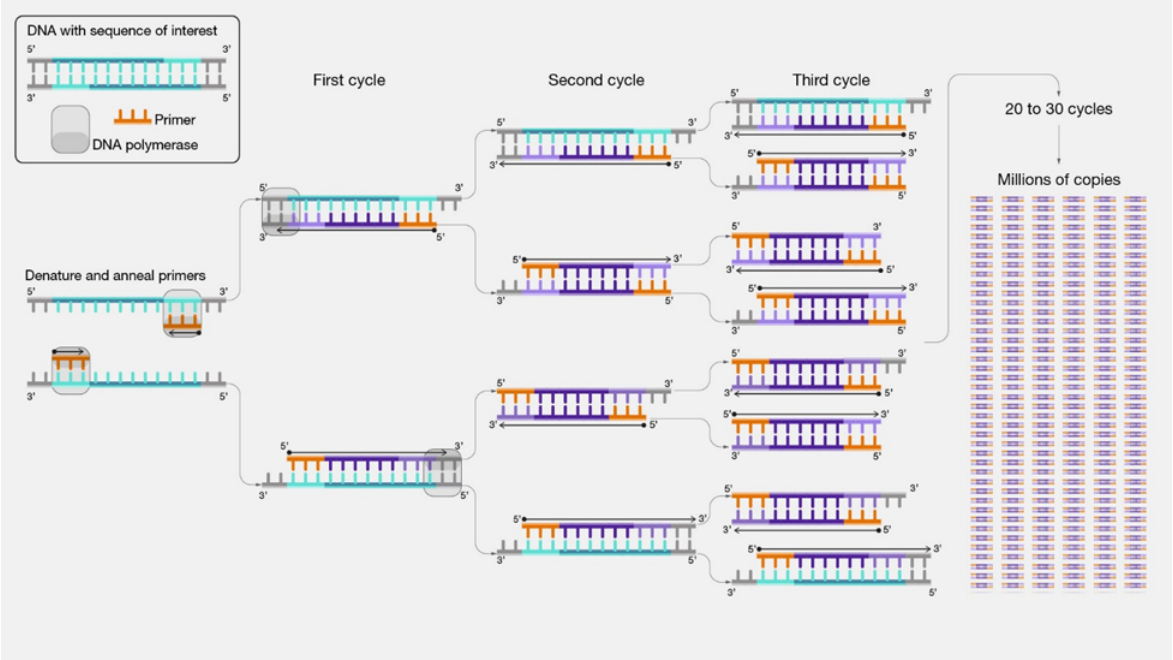


Figure 8.8.2: Illustration of Polymerase chain reaction (PCR): each cycle of PCR double the selected DNA segment. (Copyright; National Human Genome Research Institute, Public domain)

Genetic testing

Genetic testing is done in a laboratory to study an individual's DNA and diagnose a defective gene that may cause a congenital disease, as well as ancestry studies or forensic studies. For example, breast cancer may be related to defects in breast cancer genes BRCA1 and BRCA2. Patients are screened for these defects by using blood or saliva samples to extract the DNA, and the PCR technique is applied to amplify and study the defective genes. Genetic testing is also used to study the DNA of tumors or cancer cases.

Fingerprinting

Fingerprinting is a technique based on PCR that allows determining the nucleotide sequence of some areas of human DNA that are unique to individuals. A small sample from flood, skin, saliva, or semen is used to extract the DNA for amplification by PCR. Fluorescent or radioactive isotopes are incorporated in the amplified DNA for easy monitoring. The DNA is cut into pieces by restriction enzymes, placed on a gel, separated by electrophoresis, and studied to identify the individual, as illustrated in Figure 8.8.3. This technique is used for paternity tests, criminal investigations, and forensics.

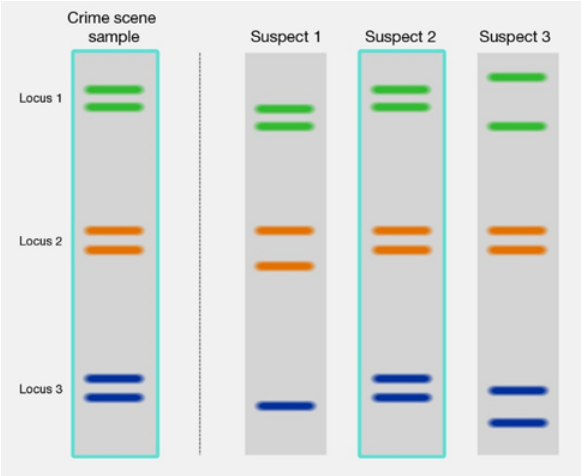


Figure 8.8.3: Fingerprint results can be used to identify a person. (Copyright; National Human Genome Research Institute, Public domain)

Human genome

The human genome project, which aimed to study all human DNA (human genome) comprehensively, was launched in October 1990 and completed in April 2003. It shows about 20,000 protein-coding genes in humans, representing only about 1.5% of the human genome. The role of the rest of the genome is still being explored. It is mainly related to regulating genes and serving as a protein recognition site. These studies provide fundamental information to study human biology, defects in DNA that lead to genetic diseases, and find cures for conditions.

CHAPTER OVERVIEW

9: Food to energy metabolic pathways

9.1: Basics of metabolism

Learning Objectives

- Define metabolism, its subclasses: catabolism and anabolism, and stages of food catabolism.
- Understand the structure of a typical cell and mitochondrion -the sites of catabolism.
- Understand the basic structural features and functions of some common compounds involved in the catabolism of food, including ATP/ADP, NAD^+/NADH, FAD/$FADH_2$ pairs.
- Understand the basic structural features of enzymes that are involved in their catalytic activity.

What is metabolism

A large number of chemical reactions take place in living things almost all the time. Some of these reactions synthesize the substances living things need from the raw materials available. Usually, these reactions convert simple molecules or compounds to more complex molecules or compounds at the expense of energy. For example, photosynthesis converts carbon dioxide (CO_2) and water (H_2O) into glucose ($C_6H_{12}O_6$) by utilizing energy from sunlight.

$$6\,CO_2 + 6\,H_2O + \text{Energy}(2808\,\text{kJ/mol}) \longrightarrow C_6H_{12}O_2$$

Other reactions break down the complex molecules to release the energy needed as heat, to do work, to provide energy for synthetic reactions, to dispose of waste byproducts, etc. For example, the reverse of photosynthesis happens in the digestion of glucose.

$$6\,CO_2 + 6\,H_2O + \text{Energy}(2808\,\text{kJ/mol})C_6H_{12}O_2 \longrightarrow 6\,CO_2 + 6\,H_2O + \text{Energy}(2808\,\text{kJ/mol})$$

Metabolism

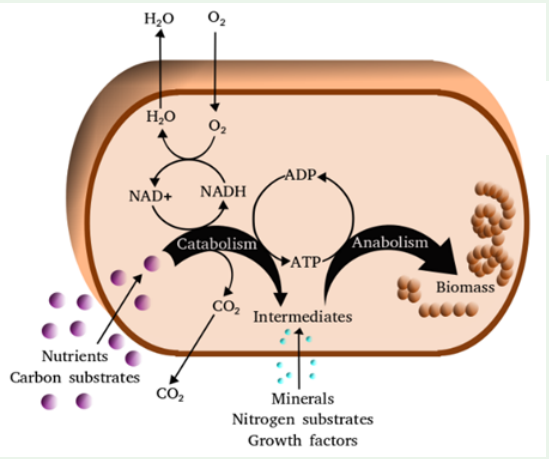
Metabolism is chemical reactions taking place in living things needed to sustain life.

Metabolism can be subdivided into two categories of reactions.

Catabolism and anabolism

- **Catabolism** is a set of chemical reactions in living things that breaks down molecules to release energy or obtain intermediates needed in other reactions.
- **Anabolism**, called biosynthesis, is a set of reactions in living things constructing molecules from smaller units.

The figure on the right summarizes metabolism and its sub-categories: catabolism and anabolism (*Copyright: Linares-Pastén, J. A. (2018), CCBYSA 4, via Wikimedia Commons*)

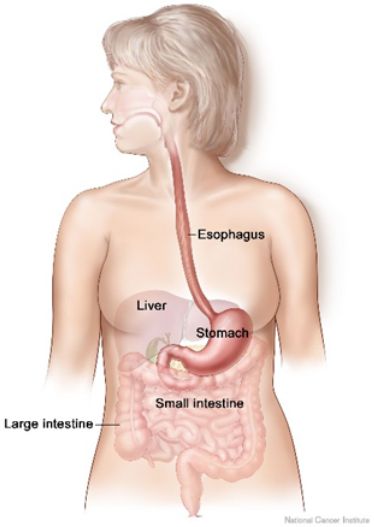


Metabolism usually happens through a series of interconnected chemical reactions called **metabolic pathways**. Metabolic pathways have a lot of similarities in different species. It indicates their common origin in the early stages of the evolution of species and their retention due to their efficiency. Some diseases, such as type II diabetes and cancer, disrupt normal metabolism. The difference in metabolic pathways from the normal allows scientists to find therapeutic interventions. Some of those metabolic pathways, particularly the catabolic pathways related to the digestion of food and its conversions to obtain energy, will be described in this chapter.

Stages of catabolism of food

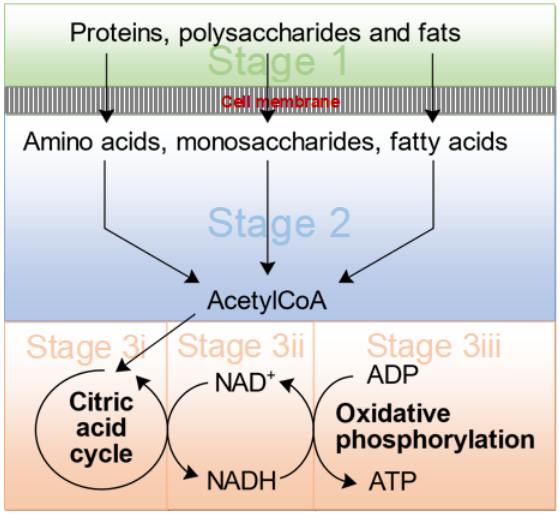
The catabolism of food starts from the digestion of food. It can be divided into three major stages, as illustrated in the figure on the right (*Copyright: modified from Tim Vickers, vectorized by Fvasconcellos, Public domain, via Wikimedia Commons*).

Stage 1



It is mainly the hydrolysis of food molecules in the digestive system, shown in the figure on the left (*Copyright: National Cancer Institute, Public domain, via Wikvia Wikimedians*).

Polysaccharides are hydrolyzed to monosaccharides, fats are hydrolyzed to glycerol and fatty acids, and proteins are hydrolyzed to amino acids. The products of the stage 1 reactions diffuse into the bloodstream and are transported to cells, entering the next phase of food catabolism.



Stage 2

It happens in the cells, which is the degradation of monosaccharides, fatty acids, and amino acids. It yields smaller groups, usually a two-carbon acetyl group or, in the case of some amino acids, four-carbon carboxylate groups, which enter the next stage.

Stage 3

It happens in mitochondria in the cells. Stage 3 can be divided into three sub-stages described below.

Stage 3i: It is the **citric acid cycle**. In this stage, the two-carbon acetyl or four-carbon carboxylate groups are oxidized to carbon dioxide (CO_2) at the expense of the reduction of coenzymes NAD^+ and FAD to $NADH$ and $FADH_2$.

Stage 3ii: It is **electron transport** where coenzymes $NADH$ and $FADH_2$ are reduced back to their oxidized forms NAD^+ and FAD at the expense of reduction of oxygen (O_2) into water (H_2O) via electron transport process with release of energy.

Stage 3iii: It is an **oxidative phosphorylation** process where the energy released during the electron transport stage is used to synthesize adenosine triphosphate (ATP), which is a high-energy molecule from adenosine diphosphate (ADP), and phosphate (Pi), which are lower-energy molecules. The energy is temporarily stored in the form of ATP and released wherever it may be needed by the reverse reaction, i.e., conversion of ATP into ADP and Pi.

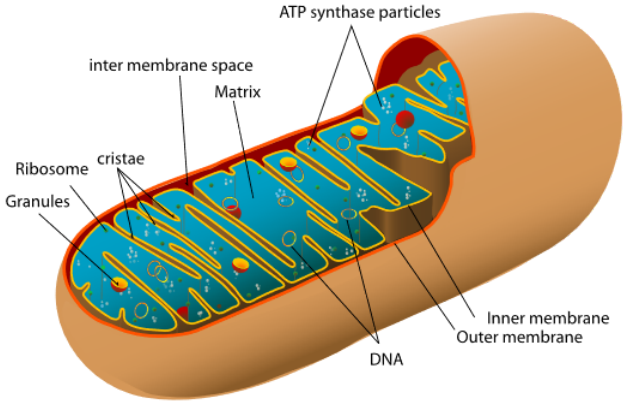
Cell structure related to food catabolism

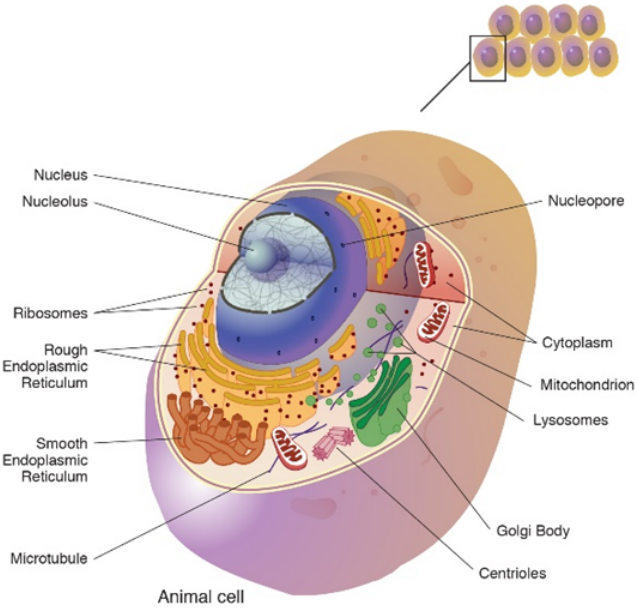
Stages 2 and 3 of food catabolism happen in the cells. A cell membrane surrounds a typical animal cell. **Organelles** are organized or specialized structures within the cells. A typical animal cell is illustrated in the right figure with some organelles labeled (*Copyright; National Human Genome Research Institute, Public domain*).

The **nucleus** in the cell contains hereditary material, i.e., DNA. The space between the cell membrane and the nucleus is called the **cytoplasm**. The **cytosol** is the fluid part of the cytoplasm containing an aqueous solution of electrolytes and enzymes that catalyze many of the cell's chemical reactions. Within the cytoplasm are organelles that perform specialized functions. For example, **ribosomes** are the sites for protein synthesis, and **mitochondria** are the cell's energy factory where stage 3 of food catabolism occurs.

Mitochondrion

The mitochondrion (plural mitochondria) has an **outer membrane** and **inner membrane** and an **inter-membrane space** between the two, as shown in the figure on the left (*Copyright; LadyofHats, Public domain, via Wikimedia Commons*). The fluid section surrounded by the inner membrane is called the **matrix**. The enzymes that catalyze the chemical reactions of stage 3 of food catabolism are located in the matrix and along the inner membrane. These reactions ultimately convert the food molecules into





CO_2, H_2O, and energy. The enzymes that catalyze stage 4 reactions that use this energy to produce high-energy ATP also occur in the matrix along the inner membrane.

The principal compounds involved in the common metabolic pathways

The principal compounds involved in the common metabolic pathways are adenosine triphosphate (ATP), which is the agent for the temporary storage of energy and transfer of phosphate (PO_4^{3-} or Pi) group; nicotinamide adenine dinucleotide (NAD^+) and flavin adenine dinucleotide (FAD) which are agents for the transfer of electrons during the biological oxidation-reduction sections; and coenzyme A (CoA) which is the agent for the transfer of acetyl (CH_3CO-) group. These compounds are described here.

ATP -the energy currency and agent for the transfer of phosphate groups

Adenosine triphosphate (ATP) is a nucleotide composed of adenine base bonded to ribose sugar by an N-glycosidic bond and triphosphate bonded to ribose by an ester bond, as shown in Figure 9.1.1. The N-glycoside of adenine with ribose is a nucleoside, i.e., adenosine. When adenosine is bonded to diphosphate, it makes adenosine diphosphate (ADP).

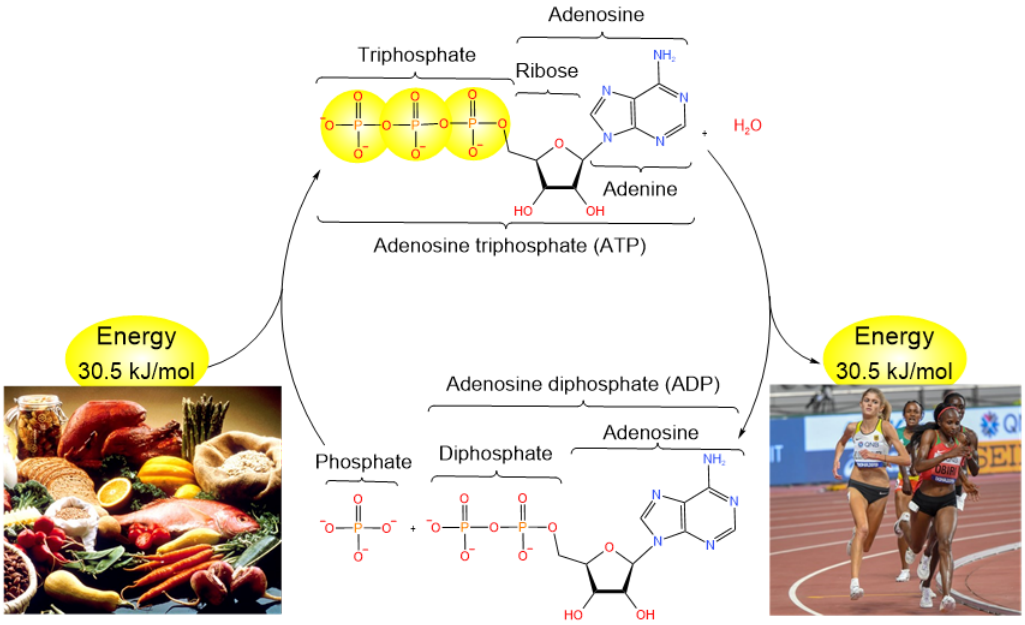


Figure 9.1.1: ATP $\rightleftharpoons$ ADP link the energy between energy-releasing relations, such as the catabolism of food, to the energy-consuming process. (Copyright; Public domain)

Hydrolysis of ATP splits one phosphate (PO_4^{3-} or Pi) and convert it to ADP with the release of 30.5 kJ/mol energy.

$$ATP + H_2O \longrightarrow ADP + Pi \quad \delta H = -30.5 \text{ kJ/mol}$$

This energy is used in the processes that require energy, such as muscle contraction, nerve signal conduction, and biosynthesis. Catabolism of food releases energy temporarily stored in the form of ATP by reversing the above reaction.

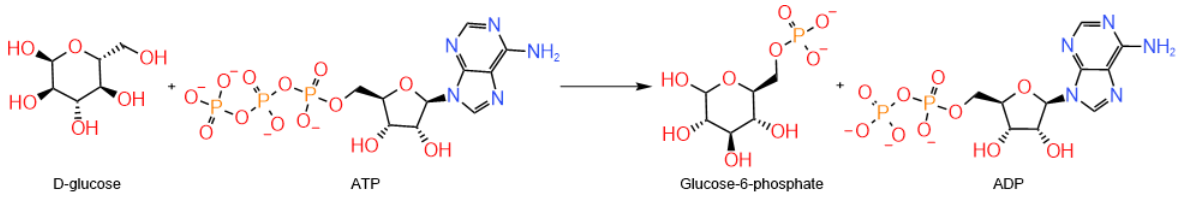
$$\mathrm{ADP + Pi \longrightarrow ATP + H_2O} \quad \delta\mathrm{H} = +30.5\ \mathrm{kJ/mol}$$

The cycle of reactions between ATP and ADP shown in Figure 9.1.1 happen rapidly, producing one to two million ATP′s per second. An average human body produces ATP about equal to the human body mass per day, but it contains about 1 g of ATP at a time.

ADP can also hydrolyze with water releasing energy and converting to adenosine monophosphate (AMP

$$\mathrm{ADP + H_2O \longrightarrow AMP + Pi} \quad \delta\mathrm{H} = -30.5\ \mathrm{kJ/mol}$$

ATP is also a phosphorylation agent in metabolic reactions. For example, D-glucose is phosphorylated by the reaction shown below.



In summary the main roles of ATP/ADP pair are:

1. **energy transfer**, i.e., releases energy when ATP converts not ADP + P_i and absorbs energy in the reverse reaction,
2. **transfer of phosphate** (P_i), i.e., releases phospahte when ATP converts not ADP + P_i and consumes phosphate in the reverse reaction.

NAD^+ and FAD -agents for the transform of electrons

Oxidation and reduction

Oxidation is:

- loss of electron (e^-),
- gain of oxygen (O), or
- loss of hydrogen (H).

Reduction is:

- gain of electron (e^-),
- loss of oxygen (O), or
- gain of hydrogen (H).

Generally, oxidation reactions release energy, and reduction reactions gain energy.

Clarifications on oxidation-reduction processes

Oxidation and reductions are coupled, i.e., if one reagent is oxidized, another is reduced simultaneously. For example, in the reaction shown below, pyruvate is oxidized by losing hydrogen coenzyme NAD^+ is reduced by gaining hydrogen. So, oxidation-reduction couples are collectively called **redox reactions**.

$$\underbrace{\mathrm{CH_3-\overset{O}{\overset{||}{C}}-\overset{O}{\overset{||}{C}}-O^-}}_{\text{Pyruvate}} + \mathrm{NADH + H^+} \xleftrightarrow{\text{Lactate dehydrogenase}} \underbrace{\mathrm{CH_3-\overset{OH}{\overset{|}{C}}H-\overset{O}{\overset{||}{C}}-O^-}}_{\text{Lactate}} + \mathrm{NAD^+}$$

Oxygen may be added to the compound, i.e., make a single or double bond with the substrate, or it may change from a single bond to a double bond in the substrate, e.g., $-C-O$ to $-C=O$, both ways it is oxidation. That is, an increase in the bond with

oxygen is oxidation.

In metabolic reactions, H is represented as a proton H^+ and an electron e^-. So, the addition of $H^+ + e^-$ is reduction, and their removal is oxidation. Usually, $2\,H^+ + 2\,e^-$ are transferred from or to a coenzyme in metabolic reactions.

If bonds with hydrogen added or removed are counterbalanced by bonds with oxygen removed or added, the overall reaction is not a redox. For example, an alkene's hydration, shown below, is not a redox reaction.

$$CH_3CH_2CH{=}CHCH_2CH_3 + H_2O \xrightarrow{H2SO4} CH_3CH_2CH_2{-}\overset{\overset{OH}{|}}{C}HCH_2CH_3$$

Broader definition of oxidation and reduction

Oxidation and reduction are not limited to making or breaking a bond with oxygen or hydrogen; broader definitions are the following.

Adding a bond with a more electronegative atom is oxidation, and its removal is reduction. The opposite is true for a less electronegative atom, i.e., The addition of a bond with a less electronegative atom is reduction, and its removal is oxidation. For example, conversion of $R{-}\overset{\overset{OH}{|}}{C}H{-}R'$ into $R{-}\overset{\overset{O}{||}}{C}{-}R'$ replaces bond on C from less electronegative H with more electronegative O is oxidation. Similarly, conversion of $R{-}\overset{\overset{O}{||}}{C}{-}CH_2R'$ into $R{-}\overset{\overset{O}{||}}{C}{-}SCH_2R'$ replaces bond on C from a C with more electronegative S is oxidation. The reverse of these is reductions.

Nicotinamide adenine dinucleotide (NAD^+)

Nicotinamide adenine dinucleotide is a coenzyme composed of two nucleotides, an adenosine diphosphate (ADP) and a second nucleotide in which the nitrogen base is nicotinamide provided by vitamin niacin. The two nucleotides are linked by diphosphate linkage, as shown in Figure 9.1.2. The oxidized form is represented as NAD^+ and its reduced form is represented as NADH. The NAD^+ is reduced to NADH by reacting with two hydrogen ($2\,H^+ + 2\,e^-$) leaving one H^+ in the products, as shown in Figure 9.1.2. Oxidation of NADH to NAD^+ is the reverse reaction.

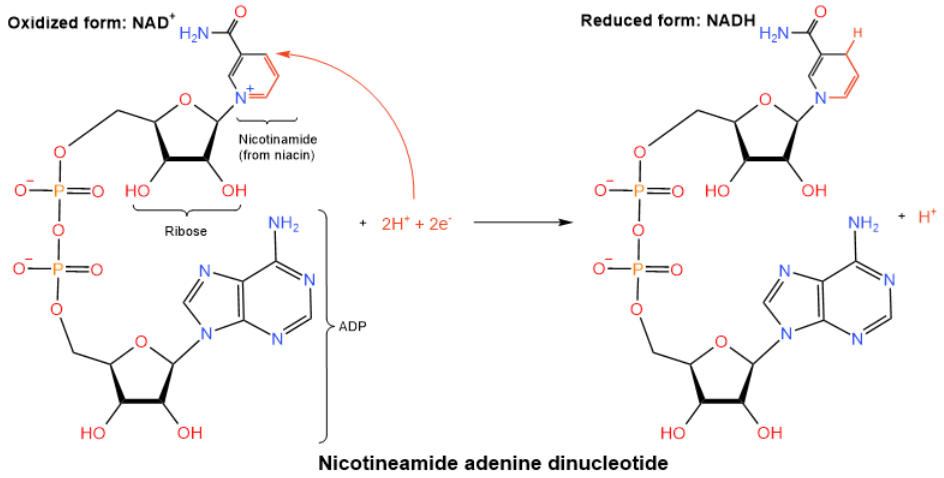
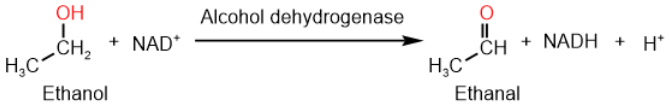


Figure 9.1.2: Illustration of the structure of the coenzyme nicotinamide adenine dinucleotide in its oxidized form NAD^+ and its conversion to the reduced form NADH by reacting with two H, i.e., $2\,H^+ + 2\,e^-$). (Copyright; Public domain)

The $NADH/NAD^+$ redox is coupled with $\underbrace{R{-}\overset{\overset{OH}{|}}{C}H{-}R'}_{Alcohol}$ / $\underbrace{R{-}\overset{\overset{O}{||}}{C}{-}R'}_{Carbonyl}$, redox reactions or it involves carbonyl (C=O) group in food catabolism .

An example of an oxidation reaction in metabolism is the oxidation of the alcohols ($-OH$) group to carbonyl ($C=O$) group. For example, ethanol is oxidized to ethanal in the liver at the expense of reduction of NAD^+ to $NADH$ as shown below.



Flavin adenine dinucleotide (FAD)

Flavin adenine dinucleotide is a coenzyme composed of two nucleotides, an adenosine diphosphate (ADP) and a second nucleotide which is riboflavin (vitamin B_2). Riboflavin is composed of flavin and ribitol, which is a sugar alcohol. The two nucleotides are linked by an ester linkage between ribitol and diphosphate, as shown in Figure 9.1.3. The oxidized form is represented as FAD, and its reduced form is represented as $FADH_2$. The oxidation-reduction happens in the N containing rings of the flavin part. The FAD is reduced to $FADH_2$ by reacting with two hydrogen ($2\,H^+ + 2\,e^-$), as shown in Figure 9.1.2. Oxidation of $FADH_2$ to FAD is the reverse reaction.

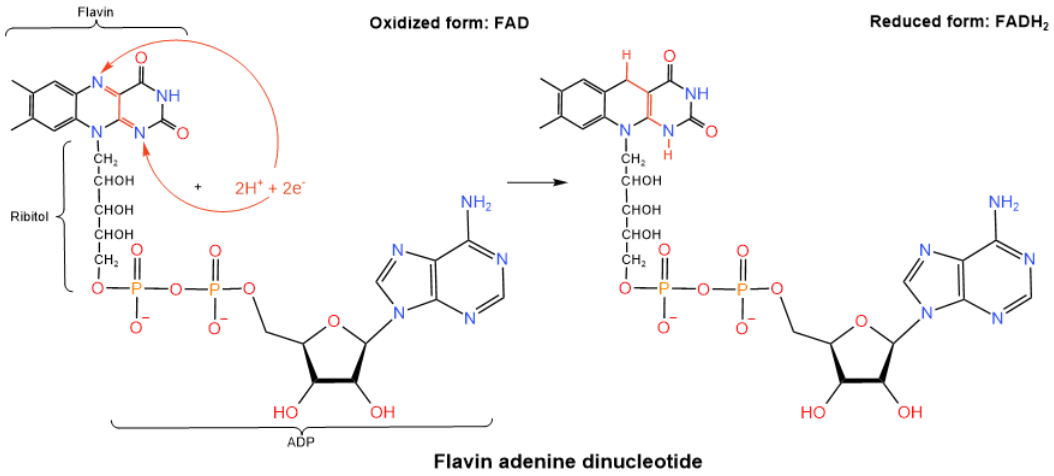
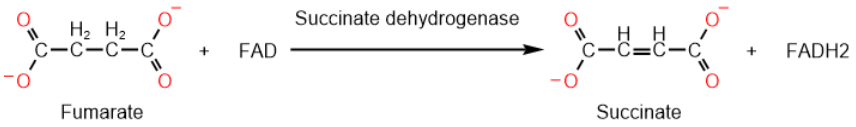


Figure 9.1.3: Illustration of the structure of the coenzyme flavin adenine dinucleotide in its oxidized form FAD and its conversion to the reduced form $FADH_2$ by reacting with two H, i.e., $2\,H^+ + 2\,e^-$. (Copyright; Public domain)

The $FADH_2/FAD$ redox is coupled with $\underbrace{RR'\overset{H}{\overset{|}{C}}-\overset{H}{\overset{|}{C}}RR'}_{Singlebond}$ / $\underbrace{RR'C=CRR'}_{\text{Doublebond}}$, redox reactions, i.e., it involves $C=C$ bond in food catabolism.

An example is the oxidation of a $C-C$ bond to $C=C$ bond of fumarate at the expense of reduction of FAD to $FADH_2$ as shown below.



Another example is the following step in the β-oxidation of fatty acids catalyzed by acyl-CoA dehydrogenase.

Coenzyme A (CoA) -the agent for transfer of acetyl (CH_3CO-) group

Coenzyme A (CoA) is composed of several components, as illustrated in Figure 9.1.4.

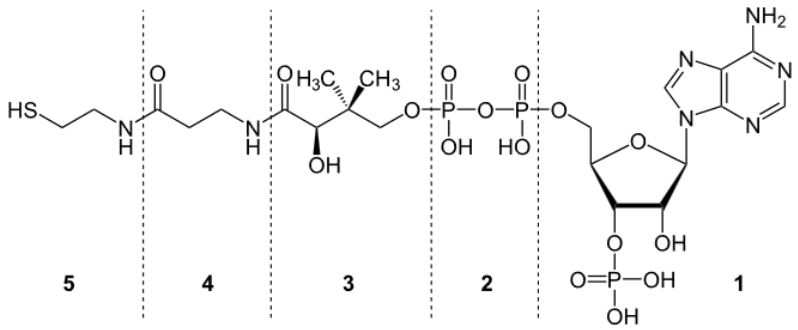
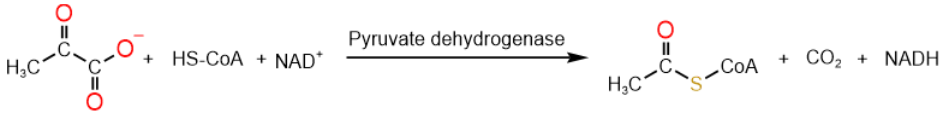


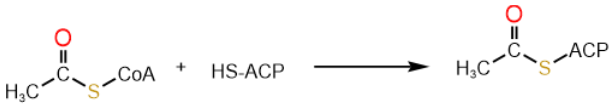
Figure 9.1.4: Structure of coenzyme A: 1 is 3′-phosphoadenosine; 2 is diphosphate, organophosphate anhydride; 3 is pantoic acid; 4 is β-alanine; 5 is cysteamine. Pantoic acid and β-alanine together are pantothenic acid, also called vitamin B_5. (Copyright; NEUROtiker, Public domain, via Wikimedia Commons)

The reactive part of CoA is the thiol ($-\mathrm{SH}$) group which forms a high-energy thioester bond with acyl groups. Free coenzyme is usually represented as $\mathrm{HS-CoA}$, and when it is as a thioester of the acyl group, it is represented as $\mathrm{Ac-S-CoA}$, where $\mathrm{Ac-}$ is an acyl group. For example, pyruvate transfers its acetyl ($\mathrm{CH_3CO-}$) groups to $\mathrm{HS-CoA}$ in the following reaction, which is a step in the catabolism of carbohydrates.



Pyruvate Acetyl CoA

The $-\mathrm{S-CoA}$ group in $\mathrm{Ac-S-CoA}$ is a good leaving group that easily transfers the acyl group to other compounds during biosynthesis. For example, the acetyl group is transferred from $\mathrm{Ac-S-CoA}$ to an acyl carrier protein (ACP) which is a step in the synthesis of fatty acids.



Acetyl CoA Acetyl ACP

Role of enzymes in metabolism

Nearly all metabolic reactions in living things are catalyzed by enzymes. Unlike chemical reactions in laboratories where different solvents, extreme temperature, and pressure conditions can be applied, and strong acids or bases can be used to catalyze the reactions, biomedical reactions in living things must occur under physiological conditions. Enzymes are specialized catalysts that catalyze the reactions under physiological conditions. The major factors responsible for the catalytic activity of the enzymes are the following.

1. Enzymes have functional groups in their active sites that act as handles to bind with and lock the substrate in a proper orientation such that the reagent can easily approach the reactive site on the substrate. This way, any steric hindrance to the reaction is minimized.
2. Enzymes and their co-factors also usually neutralize the ionic groups on the substrate that otherwise repel the charged or partially charged electrophile or nucleophile from approaching the substrate. This way, the electrostatic factors hindering the reaction are minimized.
3. Enzymes usually have functional groups in their active sites that catalyze the reaction of the substrate by activating it as a nucleophile, electrophile, acid, or base catalyst.

The following two examples from glucose catabolism illustrate the factors mentioned above.

The enzyme malate dehydrogenase catalyzes the oxidation of malate to oxaloacetate in the citric acid cycle. The substrate is bound by the enzyme in such a way that i) the charge of its two carboxylates ($(-COO^-)$ groups are neutralized by arginine side chains of the enzyme, ii) histidine side acts as a base

catalyst and removes the proton from the $-OH$ group of the substrate, and at the same time, iii) the H on the C carrying the $-OH$ group is exposed to nicotine amide ring of NAD^+ that removes the proton, as illustrated in Figure \(\PageIndex{1}\),

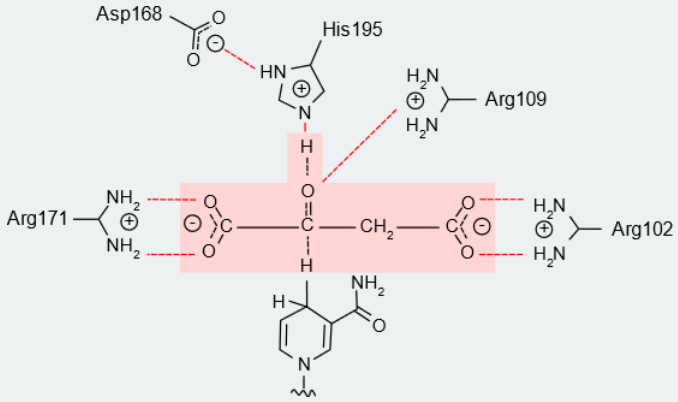


Figure 9.1.1: Active site of malate dehydrogenase with bound malate (highlighted in pink). (Copyright; Malate Dehydrogenase Active Site.jpg: Pssangderivative work:Miguelferig, CC0, via Wikimedia Commons)

Another example is the formation of citrate from oxaloacetate and acetyl-CoA in the citric acid cycle, catalyzed by the enzyme citrate synthase as illustrated in Figure 9.1.2.

1. The substrate is bonded by the enzyme in such a way that the two substrates are at a bonding distance. The $-CH_3$ hydrogens of acetyl-CoA are acidic protons due to being α to the carbonyl group. The asparagine side chain acts as a base catalyst and removes the acidic proton from $-CH_3$ group of acetyl-CoA. The proton transfer to asparagine is facilitated by another proton transfer from histidine (His^{274}) to the C=O group converting it into its enol form -a nucleophile.
2. The enol returns its proton to the histidine side chain in the second step activating the C=C bond of the enol for a nucleophilic attack on the electrophilic ketone (C=O group of oxaloacetate substrate, in the second step. This step is facilitated by the transfer of another proton from another histidine ((His^{320}), the (C=O) group involved in the reaction. The result is the conversion of the C=O group into an (−O−H group of citroyl-CoA ubterneduate.
3. Citroyl-CoA intermediate is hydrogen-bonded to a histidine (His^{274}). This hydrogen bond activates the thioester group for a hydrolysis reaction. Water molecule hydrolyzes the thioester group releasing the citrate product from the enzyme in this reaction step.

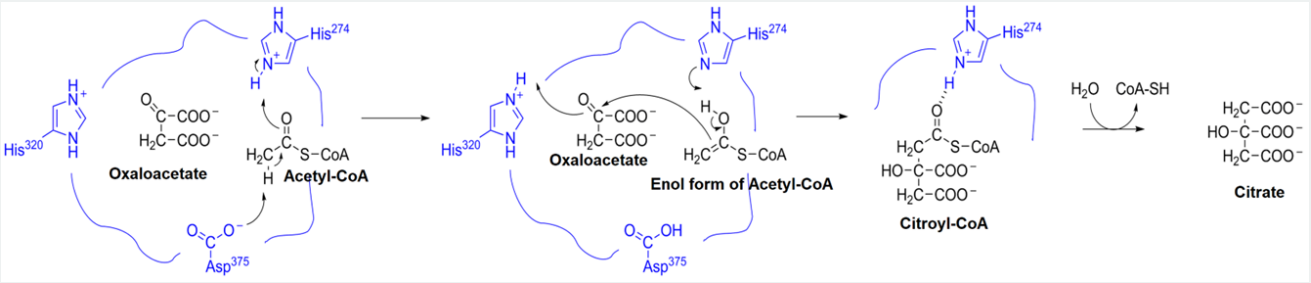


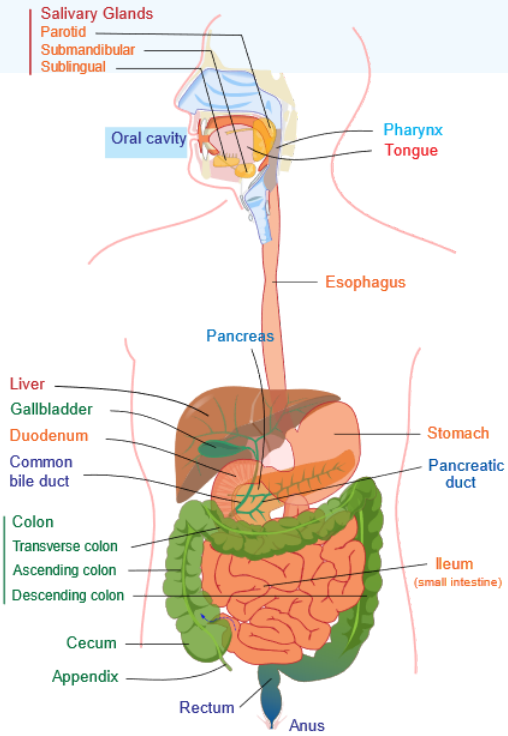
Figure 9.1.2: Illustration of enzyme action in the formation of citrate from oxaloacetate and acetyl-CoA in the citric acid cycle (Copyright; Modified from Maksim Fomich, via Wikimediaimedia Commons)

Nearly every metabolic reaction is catalyzed by an enzyme in a similar way as described above.

9.2: Digestion of food

Learning Objectives

- Understand what happens in stage 1 of food catabolism.



Food digestion, stage 1 of food catabolism, happens in the digestive system shown in the figure on the right (*Copyright: public domain, via Wikimedia Common*). It is primarily the hydrolysis of biopolymers or larger molecules in food into smaller molecules that can transfer from the digestive system to the bloodstream and transport to cells. Carbohydrates are hydrolyzed into monosaccharides, fats into fatty acids and mono- or di-glycerides, proteins into amino acids, and DNA and RNA into mononucleotides by different enzymes.

Digestion of carbohydrates

The carbohydrates in human food are starch, sucrose, and lactose. Their digestion begins in the mouth through enzymes in the saliva and continues in the small intestine by pancreatic amylase. The enzymes degrade the starch into smaller and smaller fragments that ultimately result in glucose and maltose, which the small intestine can absorb.

Sucrose (table sugar) is hydrolyzed into glucose and fructose by the enzyme sucrase. Lactose found in milk is hydrolyzed into glucose and galactose by the enzyme lactase. The majority of the adult population have problems digesting unfermented dairy because they can not produce sufficient amounts of the enzyme lactase -a condition called lactose intolerance.

Digestion of fats

Digestion of fats begins in the mouth through lingual lipase, but it mainly happens in the small intestine. Bile acids emulsify the fats, and the pancreatic enzyme lipase hydrolysis them into free fatty acids and mono- and diacyl-glycerides.

Digestion of proteins

Digestion of proteins happens in the stomach and the duodenum, i.e., the first portion of the small intestine. Three enzymes: pepsin, secreted by the stomach, and trypsin and chymotrypsin, secreted by the pancreas, hydrolyze proteins into smaller peptides which are then hydrolyzed into amino acids by various exopeptidases and dipeptidases.

9.3: Oxidation of glucose -the glycolysis

Learning Objectives

- Understand glycolysis and the essential reactions in this metabolic pathway responsible for glucose oxidation to two pyruvate molecules.
- Understand the reactions that pyruvate undergoes in the absence and presence of oxygen before entering the third stage of catabolism.

Glycolysis

Oxidation of glucose is the 2nd stage of the catabolism of carbohydrates. It happens in the cytoplasm of the cell. It is a metabolic pathway consisting of ten glycolysis reactions, as illustrated in Figure $9.3.1$.

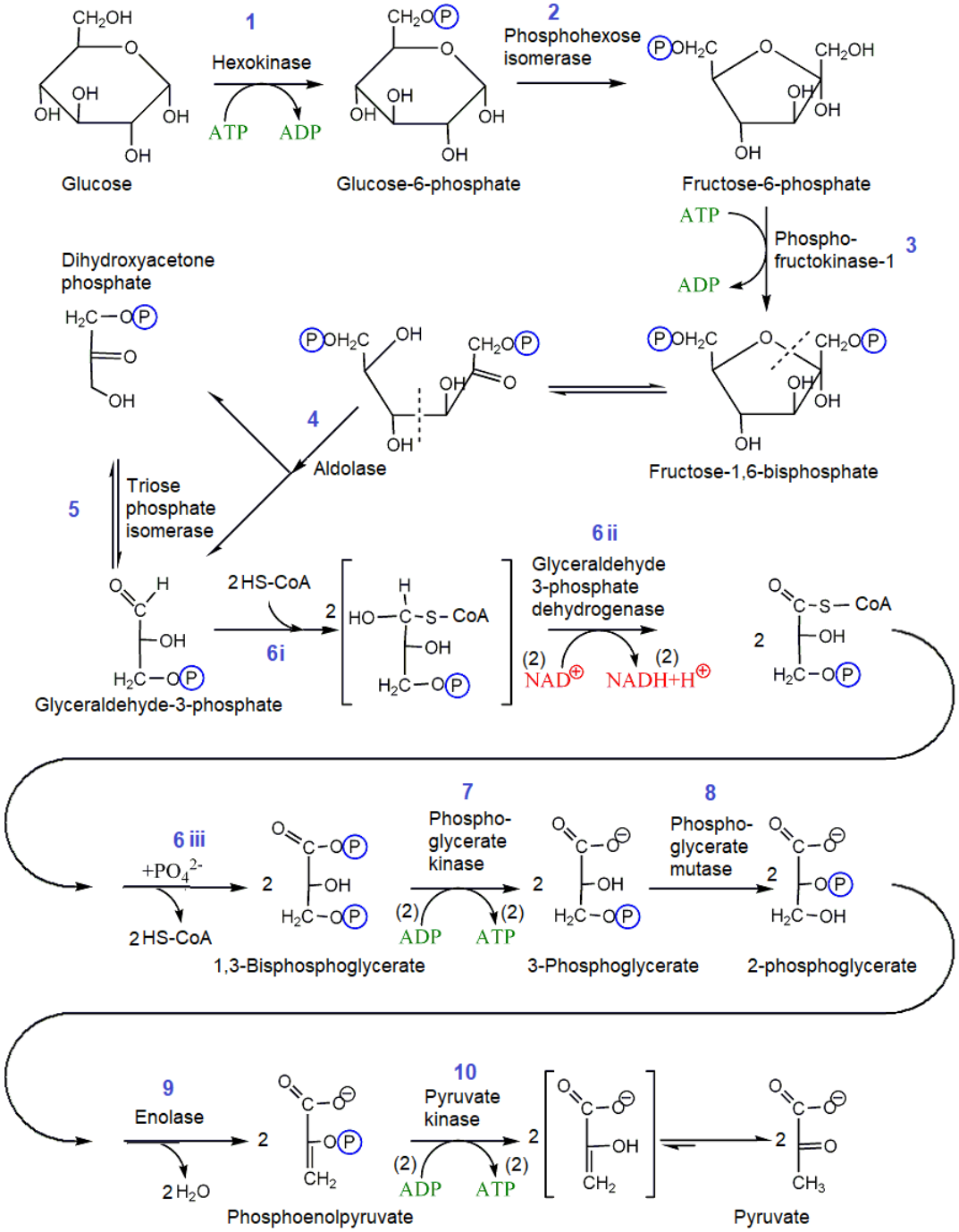
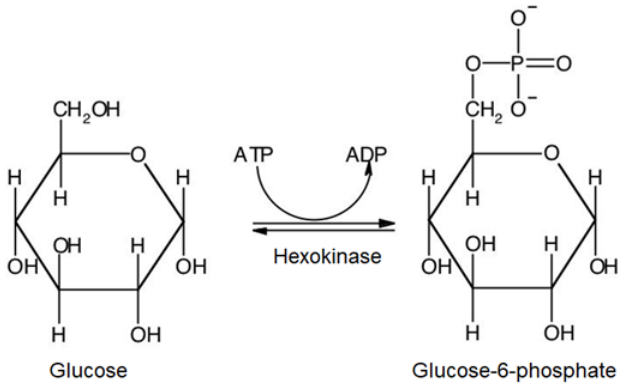


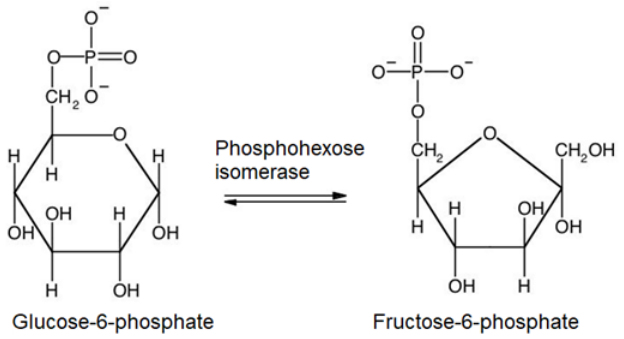
Figure 9.3.1: Illustration of ten steps of glycolysis with enzymes needed in each step. Phosphate (PO_4^{2-}) is shown as a letter P in a circle. (Copyright; Modified from created by Morglin, translated by Cryptex, Public domain, via Wikimedia Commons)

The ten reactions of glycolysis are the following (*copyright: modified from a Public Domain resource at Wikipedia*.)

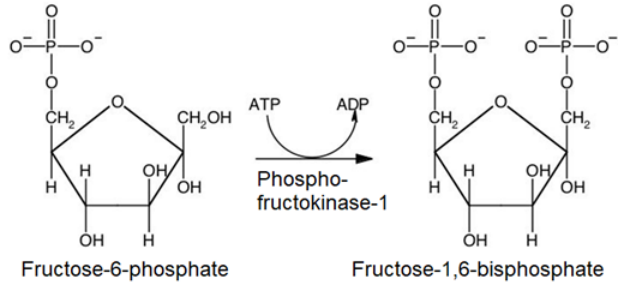
1. Glucose is converted to glucose-6-phosphate by an SN_2 reaction mechanism between a primary alcohol acting as a nucleophile and a P of ATP as an electrophile, where ADP acts as a good leaving group, as shown below.



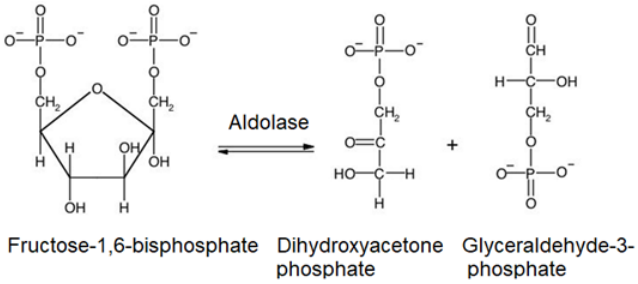
2. Glucose-6-phosphate, which is an aldohexose, isomerizes to fructose-6-phosphate, which is a ketohexose.



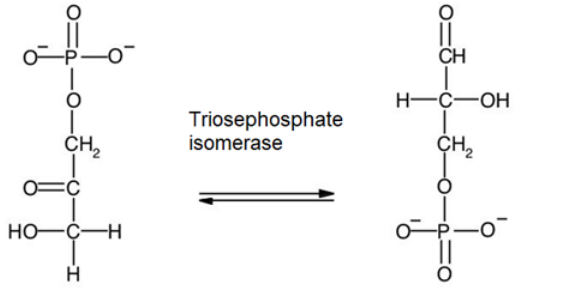
3. The same mechanism phosphorylates the primary alcohol group of fructose-6-phosphate as in the first step producing fructose-1,6-bisphosphate.



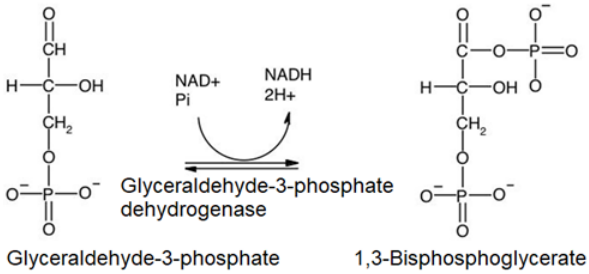
4. The open-chain form of fructose-1,6-bisphosphate, which is in equilibrium with the cyclic form, splits into two compounds, glyceraldehyde-3-phosphate and dihydroxyacetone phosphate.



5. This reaction is an isomerization reaction between glyceraldehyde-3-phosphate and its isomer dihydroxyacetone phosphate. The aldehyde or ketone in these two isomers tautomerizes to their enol forms, which either revert to the initial compound or to its isomers when the enol form changes back to an aldo or keto form.

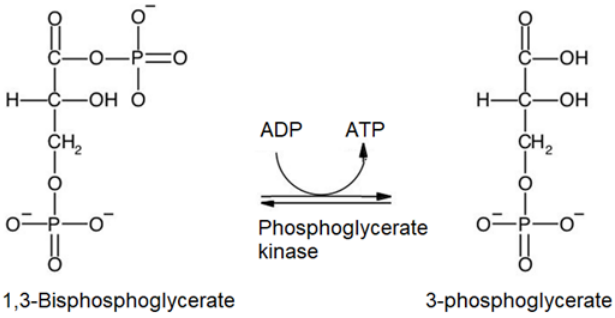


6. The aldehyde ($-HC=O$) group of glyceraldehyde-3-phosphate is oxidized in three steps: activated by the addition of $HS-CoA$ (step 6i), followed by oxidation of $-OH$ to a carbonyl ($C=O$ group at the expense of reduction of NAD^+ to $NADH$ (step 6ii)); and finally displacement of $HS-CoA$ by a phosphate ($-PO_4^{2-}$) group (step 6iii). The overall reaction is the following.



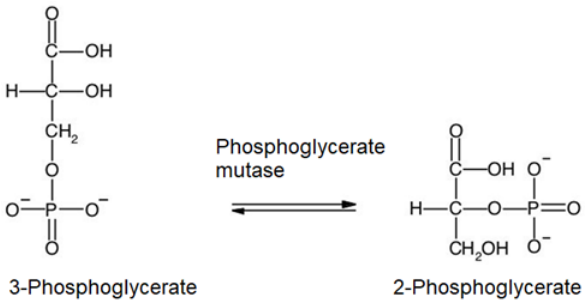
The ketone group in dihydroxyacetone phosphate does not oxidize directly but converts to glyceraldehyde-3-phosphate due to the equilibrium creation#5. So, reaction#6 and reactions after this happen twice for each glucose molecule.

7. A phosphate group is transferred from 1,3-bisphosphoglycerate to ADP producing an ATP and a 3-phosphoglycerate.

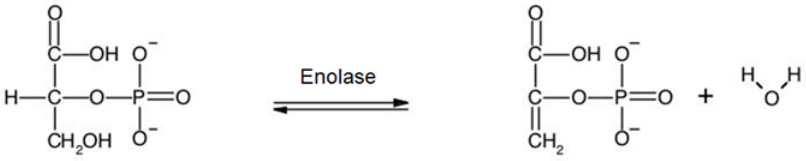


Since this step happens twice for each glucose molecule processed, it compensates for the two ATP consumed, one in step#1 and the other in step#3.

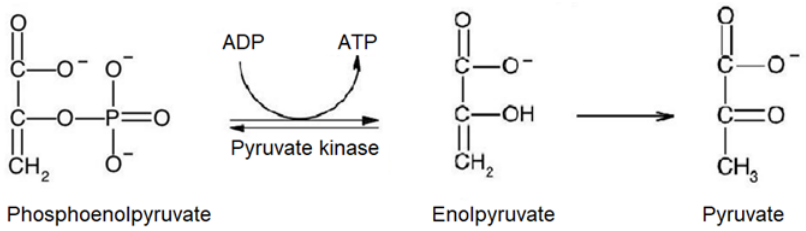
8. 3-Phosphoglycerate is isomerized to 2-phosphoglycerate. In this step, the enzyme phosphoglycerate mutase first transfers its phosphate group to the $-OH$ at position#2 and then receives the phosphate from position#3 of the intermediate, resulting in the isomerization of 3-Phosphoglycerate to 2-phosphoglycerate.



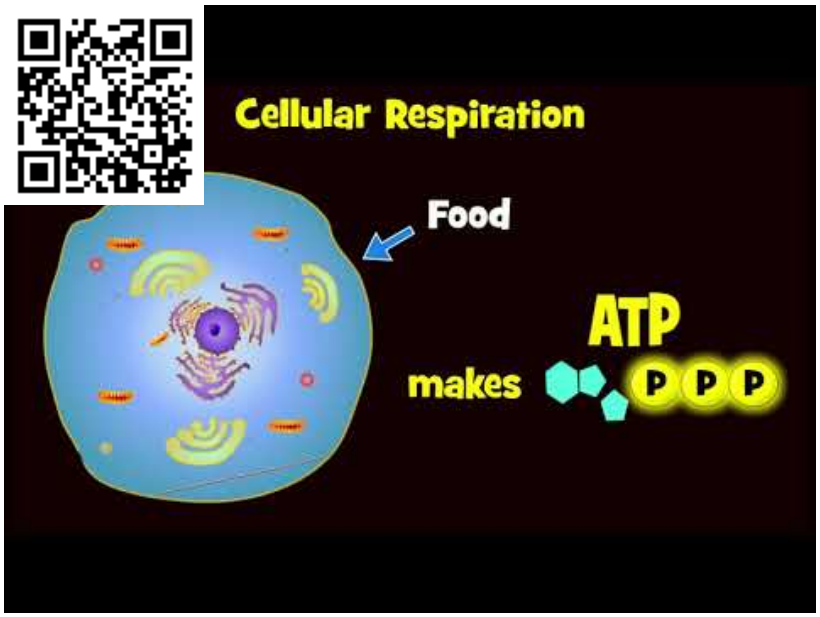
9. 2-Phosphoglycerate is dehydrated, i.e., H_2O is eliminated form it, producing phosphoenolpyruvate.



10. Phosphoenolpyruvate transfers its phosphate group to ADP producing pyruvate and ATP.



Since the reaction happens two times for each glucose, two ATP are produced in this step.



In summary, glycolysis is the oxidation of glucose without the need for oxygen, i.e., **anaerobic oxidation**, consisting of two phases. The first phase comprises reactions#1 to 5. It is the preparatory phase, which is also energy consuming phase in which a six C's glucose ($C_6H_{12}O_6$) is phosphorylated twice at the expense of two ATP's and converts to two glyceraldehyde-3-phosphate molecules having three C's each. The second phase, comprising reactions#6 to 10, is the payoff phase in which four ATP's are produced, i.e., two more than the ATP consumed in the first phase. Two NAD^+ are reduced to two $NADH$ along with the production of two pyruvates CH_3COCOO^- in the second phase, as shown in the following reaction.

$$\underbrace{C_6H_{12}O_6}_{\text{Glucose}} + 2\,NAD^+ + 2\,ADP + 2\,Pi \longrightarrow 2\,\underbrace{CH_3-\overset{O}{\overset{\|}{C}}-\overset{O}{\overset{\|}{C}}-O^-}_{\text{Pyruvate}} + 2\,NADH + 2\,ATP + 2\,H^+ + 2\,H_2O$$

Fate of pyruvate

Two NAD^+ are reduced to two $NADH$ which need to be oxidized back to NAD^+ for the process to continue. It happens in different ways depending on whether sufficient oxygen is present, i.e., **aerobic condition or aerobic respiration,** or if there is no oxygen, i.e., **anaerobic condition**. The fate of pyruvate also depends on whether the condition is aerobic or anaerobic, as described below.

Aerobic condition

In aerobic conditions, $NADH$ is not oxidized at this stage, it is oxidized to NAD^+ at the expense of oxygen in mitochondria. The energy released by the oxidation of $NADH$ is used to produce more $ATP's$ in the third stage of catabolism. Pyruvate is also transferred to mitochondria and undergoes oxidation at the cost of NAD^+ reduced to $NADH$ through a series of reactions catalyzed by a complex of three enzymes and five coenzymes, collectively called the pyruvate dehydrogenase complex. The overall reaction is the decarboxylation of pyruvate, i.e., carbon dioxide (CO_2) eliminated. The acetyl (CH_3-CO^-) group is transferred to $HS-CoA$ producing acetyl-CoA, as shown in the following reaction.

$$\underbrace{CH_3-\overset{O}{\overset{\|}{C}}-\overset{O}{\overset{\|}{C}}-O^-}_{\text{Pyruvate}} + HS-CoA + NAD^+ \xrightarrow{\text{Pyruvate dehydrogenase complex}} \underbrace{CH_3-\overset{O}{\overset{\|}{C}}-S-CoA}_{\text{Acetyl CoA}} + CO_2 + NADH + H^+$$

Acetyl-CoA enters into the citric acid cycle, also called the Krebs cycle, in mitochondria, the third stage of catabolism.

The processing of pyruvate under aerobic conditions is also called **oxidative decarboxylation** or **link reaction,** which links glycolysis to the citric acid cycle, as explained in the video below.

Two pyruvates are produced per glycolysis of one glucose molecule. The overall reaction of one glucose molecule under aerobic conditions before entry into the citric acid cycle is the following.

$$\underbrace{C_6H_{12}O_6}_{\text{Glucose}} + 2\,HS-CoA + 4\,NAD^+ + 2\,ADP + 2\,Pi \longrightarrow 2\,\underbrace{CH_3-\overset{O}{\overset{\|}{C}}-S-CoA}_{\text{Acetyl CoA}} + 2\,CO_2 + 4\,NADH + 2\,ATP + 4\,H^+ + 2\,H_2O$$

There is a net gain of six energetic molecules: $2\,ATP$ and $4\,NADH$ from glycolysis and link reaction under aerobic conditions.

Anaerobic condition

During vigorous exercises, as shown in the figure on the right, or strenuous physical work, oxygen depletes in muscles creating anaerobic (oxygen-free) conditions. Under anaerobic conditions, $NADH$ is oxidized to NAD^+ in the cytoplasm at the expense of the reduction of pyruvate to lactate, as shown in the reaction below.

$$\underbrace{CH_3-\overset{O}{\overset{\|}{C}}-\overset{O}{\overset{\|}{C}}-O^-}_{\text{Pyruvate}} + NADH + H^+ \xrightarrow{\text{Lactate dehydrogenase}} \underbrace{CH_3-\overset{OH}{\overset{|}{C}}H-\overset{O}{\overset{\|}{C}}-O^-}_{\text{Lactate}} + NAD^+$$

Overall, glycolysis of glucose under anaerobic conditions is the following reaction.

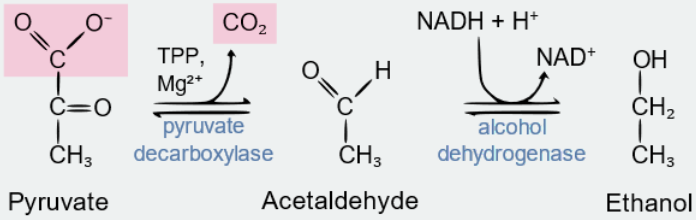
$$\underbrace{C_6H_{12}O_6}_{\text{Glucose}} + 2\,ADP + 2\,Pi \longrightarrow 2\,\underbrace{CH_3-\overset{OH}{\overset{|}{C}}H-\overset{O}{\overset{\|}{C}}-O^-}_{\text{Lactate}} + 2\,ATP$$

The two $NADH$ produced in the glycolysis of one glucose are consumed in the anaerobic conversion of two pyruvates into two lactates.

- There is a net gain of two energetic molecules, i.e., $2\,ATP$ in glycolysis of one glucose followed by anaerobic conversion of two pyruvates into two lactates.
- The accumulation of lactate makes muscles tire and sour. The person keeps beating heavily after the exercise to pay the oxygen debt. Most of the lactate is transported to the liver, where it is re-oxidized to pyruvate.

Fermentation

Instead of converting pyruvate to lactate, as in humans and animals, yeast has an enzyme called pyruvate decarboxylase that decarboxylates pyruvate to acetaldehyde. Then $NADH$ is oxidized to NAD^+ at the expense of the reduction of acetaldehyde to ethanol in a process called fermentation, as shown in the reaction below.



Copyright; User: Cwernert at en.Wikipedia, vectorized by User: GKFX., CC BY-SA 3.0, via Wikimedia Commons.

The overall reaction of glycolysis of glucose in the fermentation process by yeast.

$$\underbrace{C_6H_{12}O_6}_{\text{Glucose}} + 2\,ADP + 2\,Pi \longrightarrow 2\,\underbrace{CH_3-CH_2-OH}_{\text{Ethanol}} + 2\,CO_2 + 2\,ATP$$

9.4: Citric acid cycle

Learning Objectives

- Understand the citric acid cycle, its reactions, and the yield of energetic molecules.

What is the citric acid cycle?

The **citric acid cycle** is the first part of the third stage of food catabolism, including carbohydrates, fats, and proteins. It is called the citric acid cycle because a $2\,\mathrm{C'}$s acetyl-CoA produced in the second stage of catabolism of foods reacts with a $4\,\mathrm{C'}$s oxaloacetate and produces $6\,\mathrm{C'}$s-citrate in the first reaction. The citrate is oxidized through a series of eight reactions producing two carbon dioxide ($\mathrm{CO_2}$) and re-generates the oxaloacetate to repeat the next round of the cycle, as shown in Figure 9.4.1. It is also called the **tricarboxylic acid cycle** because citrate has three carboxylates ($-\mathrm{COO^-}$) groups, i.e., triacid. Another name for it is the **Krebs cycle** in honor of Hans Krebs, who discovered this metabolic pathway.

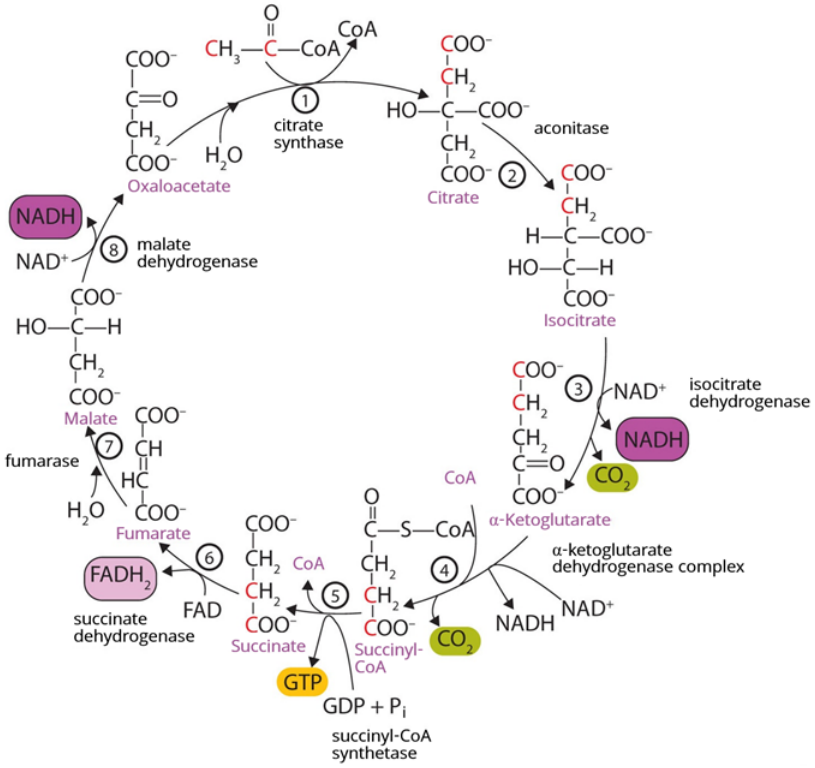


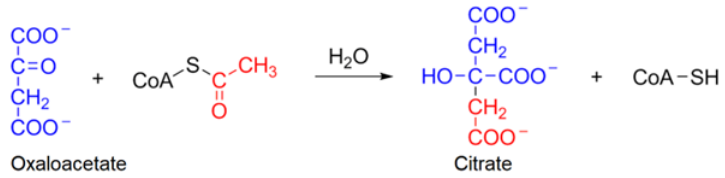
Figure 9.4.1: Citric acid cycle. (Copyright; Theislikerice, CC BY-SA 4.0, via Wikimedia Commons)

The oxidation reactions in the citric acid cycle release energy, which is coupled to the reduction of $\mathrm{NAD^+}$ to NADH, FAD to FADH, or conversions of ADP to high-energy ATP. Guanosine triphosphate (GTP) -another high-energy molecule is also produced from guanosine diphosphate (GDP), but this reaction ultimately reverses to transfer its energy to make ATP.

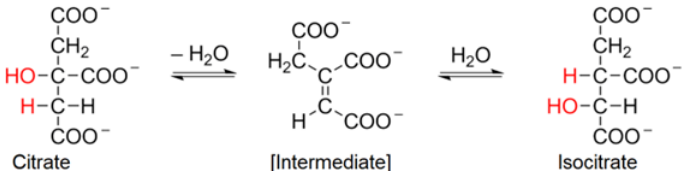
Reactions of the citric acid cycle

The following equations of the reactions are modified from a public domain at Wikipedia.

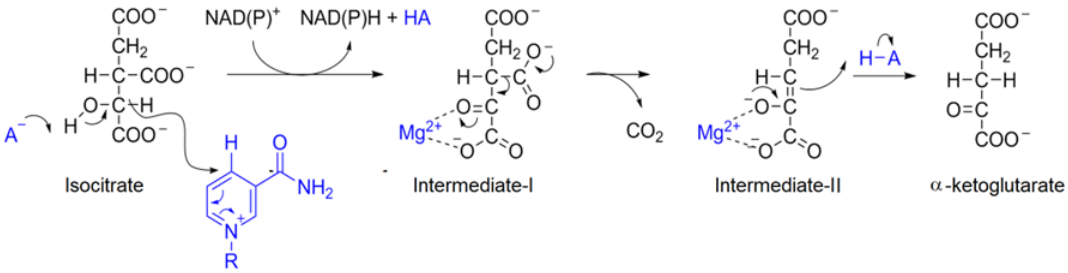
1. To understand the first reaction, recall that the $-\mathrm{CH_3}$ group of acetyl-CoA has relatively acidic protons as it is α to a carbonyl ($\mathrm{C{=}O}$) group. An enzyme, acting as a base, removes a proton from the $-\mathrm{CH_3}$ group making it a carbanion which, being a strong nucleophile, attacks the ketone-C of oxaloacetate, which is an electrophile. This nucleophilic addition reaction converts the $\mathrm{C{=}O}$ into an $-\mathrm{OH}$ group. It is followed by the hydrolysis of the thioester by $\mathrm{H_2O}$ through nucleophilic acyl substitution mechanism producing citrate and $\mathrm{HS{-}CoA}$, as shown in the following overall reaction.



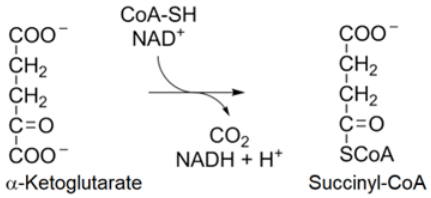
2. To understand the second reaction, recall that tertiary $-OH$ do not oxidize but are easily eliminated through E2-dehydration mechanics. The tertiary $-OH$ of citrate is eliminated, and then H_2O adds to the alkene intermediate, but it installs a secondary $-OH$ group in isocitrate product, as shown below.



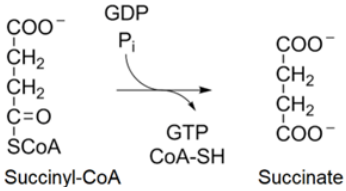
3. The third reaction has three steps. In the first step, the secondary $-OH$ in isocitrate is oxidized to a ketone group at the expense of reduction of a NAD^+ into $NADH$. In this step, one $-H$ is picked up by an enzyme (A^-) and the second by the nicotinamide ring (shown in blue color) of the coenzyme NAD^+. The COO^- group in the intermediate-I, being β to the $C=O$ group, eliminates (decarboxylate) easily as CO_2. The decarboxylation is facilitated by cofactor Mg^{2+} that binds with and draws electrons from the $O's$. The enzyme $H-A$ protonates the enole $C=C$ bond of the intermediate-II that produces α-ketoglutarate, as shown below.



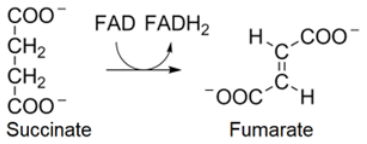
4. The fourth reaction resembles the previous link reaction's oxidative decarboxylation of pyruvate to acetyl-CoA. In this reaction, NAD^+ is reduced into $NADH$ at the expense of oxidative decarboxylation of α-ketoglutarate and the resulting acyl-group is transferred to $HS-CoA$ resulting in succinyl-CoA product as shown in the reaction below.



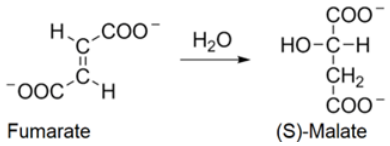
5. The fifth reaction has two steps. In the first step, a phosphate (Pi) group substitutes $-S-CoA$ group from succinyl-CoA. In the second step, the phosphate group is transferred to GDP that converts to a GTP by the same mechanism as in steps #7 and step#10 of glycolysis. GTP reverts back to GDP and transfers its phosphate group to ADP that converts to one ATP. So, production of one GTP is equal to production of one ATP. The overall reaction five is shown below.



6. In the sixth reaction, one FAD is reduced into $FADH_2$ at the expense of oxidation of succinate to fumarate by the following overall reaction.

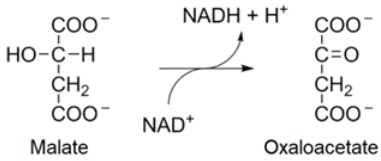


7. In the seventh reaction H_2O hydrates the $C{=}C$ of fumarate producing malate, as shown below.



Recall that the enzymes are stereospecific. In this case, only fumarate (not its cis-isomer) reacts with the enzyme, and only (S)-malate (not its enantiomers (R)-malate) is produced.

8. In the eighth step, one more NAD^+ is reduced into $NADH$ at the expense of oxidation of malate into oxaloacetate, as shown in the reaction below.



The oxaloacetate is the reactant of step#1 and begins the cycle again by reacting with another acetyl-. One citric acid cycle has the following overall reaction with one acetyl-CoA.

$$\underbrace{CH_3{-}\overset{\overset{O}{\|}}{C}{-}S{-}CoA}_{\text{Acetyl CoA}} + 3\,NAD^+ + FAD + GDP + P_i + 2\,H_2O \longrightarrow HS{-}CoA + 3\,NADH + 3\,H^+ + FADH_2 + GTP + 2\,CO_2$$

Recall that the catabolism of one glucose molecule produces two pyruvates that, intern, produce two acetyl-CoA in the oxidative decarboxylation of pyruvate under aerobic conditions. So, the citric acid cycle runs twice for the catabolism of one glucose molecule. One GTP ultimately converts into one ATP.

Two citric acids from one glucose molecule yield ten energetic molecules from two turns citric acid cycles: $6\,NADH$, $2\,FADH_2$, and $2\,ATP$

Predicting the production or consumption of energetic molecules in food catabolism

The information in this note is meant to help predict which energetic molecule will be produced or consumed during food catabolism reactions. Recall that oxidation generally release energy (exothermic), and reduction consumes energy (endothermic). The energy released from oxidation reactions in food catabolism is utilized to produce one of the two energetic coenzymes, i.e., either $NADH$ from the reduction of NAD^+ or $FADH_2$ from the reduction of FAD. Recall that $C{=}O$ is stronger bond (~800 kJ/mol bond dissociation energy) compared to $C{=}C$ bond (~600 kJ/mole bond dissociation energy). $NADH$ is more energetic than $FADH_2$. Therefore, any oxidation reaction involving carbonyl ($C{=}O$) group produces $NADH$ and any oxidation involving $C{=}C$ bond produces $FADH_2$. There is only one reaction described in food catabolism in this book, i.e., reaction#6 in the citric acid cycle that involved oxidation of $C{-}C$ to $C{=}C$ bond and produced $FADH_2$. All other oxidation reactions in the citric acid cycle and glycolysis involve $C{=}O$ group, i.e., either alcohol ($CH{-}OH$) to carbonyl ($C{=}O$) conversion, or ketone ($R{-}\overset{\overset{O}{\|}}{C}{-}R'$ to thioester ($R{-}\overset{\overset{O}{\|}}{C}{-}S{-}R$ ") conversions produce NAD. Conversionerisons of pyruvate to lactate and ethanol under anaerobic conditions are reduction reactions involving $C{=}O$ group and consume $NADH$.

Reactions that are not redox reactions, e.g., acyl ($\mathrm{R-\overset{\overset{\mathrm{O}}{\|}}{C}-}$) transfer, hydration (addition of $\mathrm{H_2O}$) , dehydration (elimination of $\mathrm{H_2O}$), and isomerization, involve a small amount of energy. They are usually not coupled with the formation or consumption of energetic molecules like NADH, $\mathrm{FADH_2}$, or ATP.

Exceptions include phosphate ester or thioester group, which are high-energy groups. The addition or removal of phosphate ester or thioester groups without redox reaction is usually coupled with ATP/ ADP conversion. Two reactions, i.e., reaction#1 and reaction#3 in glycolysis (Figure 9.3.1) add phosphate ester bond and consume ATP, and two reactions, i.e., reaction#& and reaction#10, remove phosphate ester and produce ATP. One reaction, i.e., reaction#5 in the citric acid cycle (Figure 9.4.1), involves the replacing a thioester group with a carboxylate group without redox. It is coupled with the production of GTP, which, in turn, is associated with the output of ATP.

In summary:

- redox reactions involving $\mathrm{C=O}$ group relate to NADH,
- redox reactions involving $\mathrm{C=C}$ group relate to $\mathrm{FADH_2}$, and
- all other reactions that involve energetic groups like phosphate esters or thioesters relate to ATP in food catabolism.

NADH and FADH are high-energy molecules that are oxidized by $\mathrm{O_2}$ in part two of the third stage of catabolism of food. The energy released from the oxidation of NADH and FADH is used to produce more ATPs, which is described in the next section.

The following video summarizes the citric acid cycle, also called the Krebs cycle.

9.5: Oxidative Phosphorylation

Learning Objectives

- Understand parts 2 and 3 of the third stage of food catabolism, i.e., oxidative phosphorylation, and the basics of the redox chemistry of the molecules involved, including flavin mononucleotide, ubiquinone, and cytochrome c.
- Understand the electron transport process basics, including basic redox reactions and proton pumping in complexes I through IV.
- Understand the ATP synthesis process, accounting for the ATP yield, body heating, and reactive oxygen species produced.

What is oxidative phosphorylation?

In food catabolism up to this point, the energy is transferred from food to high-energy molecules, such as energetic electron carriers NADH and $\mathrm{FADH_2}$. Recall that these electron carrier molecules do not deliver elections alone; they deliver protons and electrons, i.e., $2\,\mathrm{H^+} + 2\,\mathrm{e^-}$. The electrons in NADH and FADH are transported through a series of enzymes located in the inner membrane of mitochondria, and ultimately, the $2\,\mathrm{H^+} + 2\,\mathrm{e^-}$ reduce O into $\mathrm{H_2O}$ by the reaction shown below.

$$4\,\mathrm{H^+} + 4\,\mathrm{e^-} + \mathrm{O_2} \longrightarrow 2\,\mathrm{H_2O}$$

In each step, the electrons are transferred from a substance with a higher reduction potential to one with a lower reduction potential. Energy is released in each step proportional to the difference in the reduction potential. The energy is converted to electrochemical potential energy by pumping protons ($\mathrm{H^+}$) from the matrix to intermembrane spaces through sets of enzymes called complexes I, III, & IV. The electrochemical energy is harnessed by ATP synthase, which allows protons to flow back from the intermembrane space to the matrix and couples the energy released to synthesize higher energy ATP by phosphorylating lower energy ADP with inorganic phosphate (Pi). These interconnected processes of electron transport and ATP synthesis are collectively called **oxidative phosphorylation**. It happens in the mitochondria of cells as illustrated in Figure $9.5.1$ and is explained in the following sections.

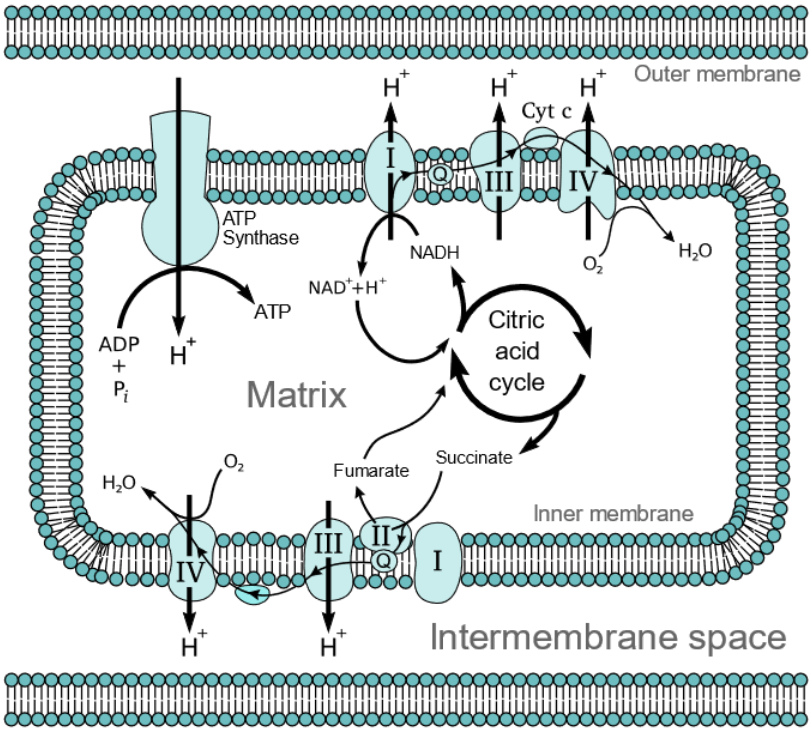
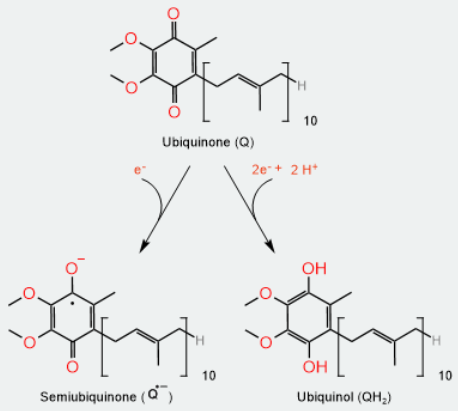
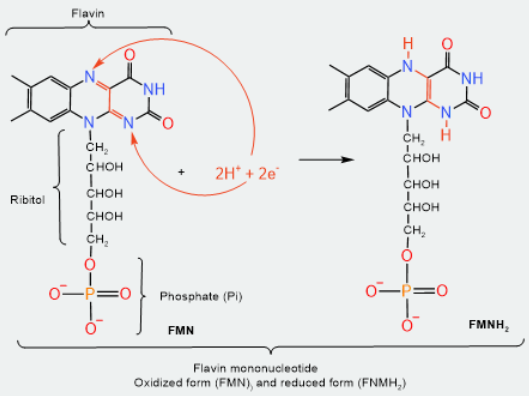


Figure $9.5.1$: Schematic diagram of the mitochondrial electron transport chain comprising of complexes I through IV and production of ATP by ATP-synthase. (Copyright; Fvasconcellos 22:35, 9 September 2007 (UTC), Public domain, via Wikimedia Commons)

Electron and proton transport molecules

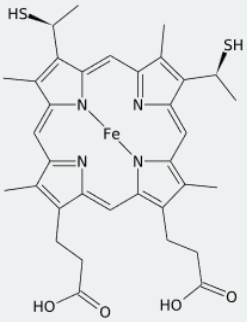


In addition to NAD^+ and FAD, the following molecules are involved in the transport of electrons and protons in the electron transport chain. The first electron acceptor from $NADH$ or $FADH_2$ in the electron transport chain is **flavin mononucleotide** which changes from its oxidized form represented as FMN to its reduced form represented as $FMNH_2$ as illustrated in the figure on the right. Structure of FMN comprises three parts, phosphate, ribitol, and flavin. It is the same as one of the two nucleotides in flavin adenine dinucleotide. Nitrogen atoms in flavin part of FMN take part in the redo process, as illustrated in the figure on the right. $FMNH_2$ pass on the electrons to iron-sulfur clusters, which include $[2\,Fe{-}2\,S]$ and $[4\,Fe{-}4\,S]$ clusters. The redox couple in iron-sulfur clusters is Fe^{3+}/Fe^{2+}, i.e., the following redox reaction:



$$Fe^{3+} + e^- \underset{Oxidation}{\overset{Reduction}{\rightleftharpoons}} Fe^{2+}.$$

Finally, the electrons are transferred to coenzyme-Q, which in oxidized form is ubiquinone represented as Q and in its reduced form is ubiquinol represented as QH_2, as illustrated in the figure above on the left. Coenzyme-Q is hydrophobic, i.e., lipid-soluble. Coenzyme-Q moves freely in the hydrophobic environment within the biliary inner membrane to pass on electrons from complex I and II to complex III. Coenzyme-Q not only can accept two electrons and two protons, but it can also accept one electron to become semiubiquinone, a resonance-stabilized radical anion.



Cytochrome c (Cyt c) is another electron carrier protein with a heme group attached. Heme group is illustrated in the figure on the right. Iron in the heme group of cytochrome c is the redox couple (Fe^{3+}/Fe^{2+}) as in the case of iron-sulfur clusters. Fe^{3+} changes to Fe^{2+} when an electron is added to it, and the reverse happens when an electron is removed. Cytochrome c is water soluble and travels between complex III and complex IV in the watery environment along the inner membrane's surface facing intermembrane space.

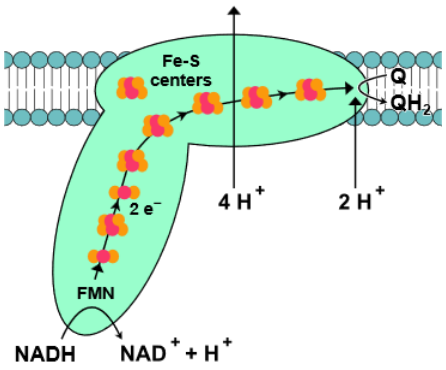
Electron transport chain

In this part 2 of stage 3, electrons are removed from $NADH$ and $FADH_2$ and passed on to reduce O_2 into H_2O through a series of enzyme complexes, i.e., complexes I through IV and mobile electron carriers like ubiquinone (Q) and cytochrome c ($Cyt\ c$) by a process called **electron transport chain**. At each step, electrons pass from a substance with a higher reduction potential to one with a lower one. Energy is released proportional to the difference in the reduction potentials.

The complex I, III, and IV extend from the matrix to the intermembrane space through the inner membrane, as illustrated in Figure $9.5.1$. These three complexes utilize the energy released to pump protons (H^+) from the matrix to the intermembrane space. Complex II extends to the matrix from the inner membrane but does not cross the membrane to intermembrane space. Therefore complex II transports electrons but does not pump protons. Coenzyme-Q (Q) is a hydrophobic electron carrier molecule that moves in a hydrophobic environment inside the lipid bilayer of the inner membrane from complexes I and II to complex III. Cytochrome c ($Cyt\ c$) is a hydrophilic electron carrier molecule that moves along the surface of the inner membrane facing the intermembrane space from complex III to complex IV. The enzyme complexes and the mobile electron carriers involved in the electron transport chain and ATP synthesis are described in the following sub-sections, where the chemical equations and figures are taken from the public domain source via Wikipedia.

Complex I

Complex I, also known as NADH-coenzyme Q oxidoreductase or NADH dehydrogenase, is the first enzyme complex in the electron transport chain. Complex I is illustrated in the figure on the right. The electron transport begins when an NADH binds with and donates two electrons to complex I. The electrons are received by flavin mononucleotide (FMN) attached to the complex, which is reduced to $FMNH_2$ form. Then the electrons are transferred through a series of iron-sulfur clusters within complex I to coenzyme-Q (Q) that, in turn, reduces to its reduced form QH_2. QH_2 travels through the lipid bilayer to and delivers the electrons to complex III.



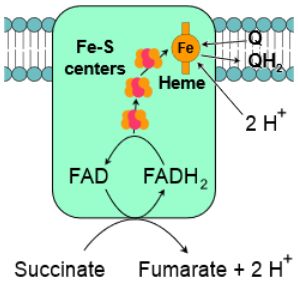
As the electrons move through complex I and release part of their energy, four protons ($4\,H^+$) are pumped from the matrix to intermembrane space by utilizing the energy released. It generated polarity across the inner membrane, with the positive side (P-side) facing the intermembrane area and the negative side (N-side) facing the matrix. The exact mechanics of proton pumping is not yet clearly understood. Still, the research indicates that it happens through conformational changes in complex I such that the protein bind protons on the N-side and releases them on the P-side of the membrane. The overall redox reaction in complex I is the following.

$$NADH + Q + 5\,H^+_{matrix} \longrightarrow NAD^+ + QH_2 + 4\,H^+_{intermembrane}$$

NAD^+ liberated in this reaction becomes available to oxidize more substrates in the metabolic pathways.

Complex II

Complex II, also known as succinate-Q oxidoreductase or succinate dehydrogenase, is a second entry point to the electron transport chain. It is part of the citric acid cycle and the electron transport chain. It contains protein subunits bonded with flavin adenine dinucleotide FAD, iron-sulfur clusters, and a heme group, as illustrated in the figure on the right. Succinate is oxidized to fumarate in the citric acid cycle at the expense of oxidation of FAD to $FADH_2$ in complex II. The electrons and protons transfer from $FADH_2$ through iron-sulfur clusters to coenzyme-Q in complex II by the following overall reaction.



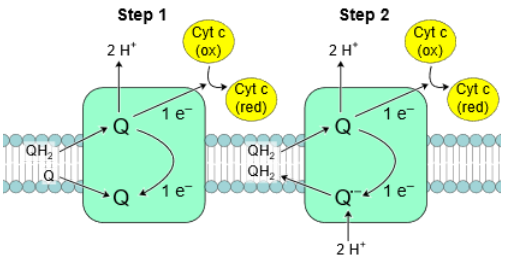
$$Succinate + Q \longrightarrow Fumarate + QH_2$$

Coenzyme-Q passes on the electrons from complex II to complex III as from complex I to complex III.

Recall that (\ce{FADH2}\) is less energetic than (\ce{NADH}\). Since oxidation $FADH_2$ to FAD at complex II releases less energy than oxidation of $NADH$ to NAD^+ at complex I, complex II do not pump protons from matrix to intermembrane space. The energy released by electron transport is utilized to pump protons in complex III and IV, as described in the next section.

Complex III

Complex III, also known as cytochrome C reductase or cytochrome bc_1 complex, contains protein subunits, an iron-sulfur cluster, and cytochromes. A cytochrome is a protein with at least one heme group that transfers one electron by accepting and releasing an electron based on the following reaction:



$$Fe^{3+} + e^- \rightleftharpoons Fe^{2+}$$

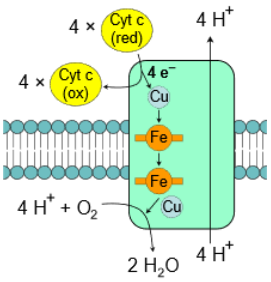
Complex III passes on electrons from coenzyme-Q (QH_2) to cytochrome c ($Cyt\ c$). Since QH_2 delivers two electrons (e^-) while $Cyt\ c$ can accept only one, the process happens in two steps, as illustrated in the figure on the right. In the first step, QH_2 transfers one e^- to $Cyt\ c$ and one to another coenzyme-Q in quinone (Q) form and releases $2\,H^+$. Q converts to a semiubiquinone ($Q^{\bullet -}$). In the second step, another QH_2 transfers one electron to a second $Cyt\ c$ and one to $Q^{\bullet -}$ converting it to Q which is released from the complex and repeats the above two steps. This way, two electrons are transferred from a Q to two $Cyt\ c$ and $4\,H^+$ are pumped out into intermembrane space by complex III as shown in the following reaction.

$$QH_2 + 2\,Cyt\ c_{ox} + 2\,H^+_{matrix} \longrightarrow Q + 2\,Cyt\ c_{red} + 4\,H^+_{intermembrane}$$

$Cyt\ c$ carries the electron to complex IV.

Complex IV

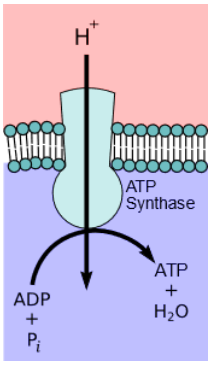
Complex IV, also known as cytochrome c oxidase, is the final complex in the electron transport chain. It contains several subunits, two heme groups, and several metal ion cofactors, including three atoms of copper, one magnesium, and one zinc. Complex IV reduces oxygen into water at the expense of oxidation of Cyt c and pumps $4\,H^+$ from matrix to intermembrane space as illustrated in the figure on the right and shown in the following overall reaction.

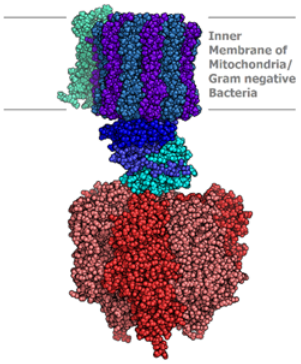


$$4\,\text{Cyt c}_{\text{red}} + O_2 + 8\,H^+_{\text{matrix}} \longrightarrow 4\,\text{Cyt c}_{\text{ox}} + 2\,H_2O + 4\,H^+_{\text{intermembrane}}$$

ATP-synthase

ATP-synthase, also known as complex V is the final complex in the oxidative phosphorylation process, i.e., part 3 of stage 3 of food catabolism. It comprises two parts in the shape of a mushroom, as illustrated in the figure on the left. The first part, F_O, is the hydrophobic part embedded in the inner membrane (a blue and purple region in the model image on the right (*Copyright; Alex. X, CC BY-SA 3.0, via Wikimedia Commons*). The F_O acts as a pore for the flow of protons (H^+) across the membrane. The second part is F_1 which is hydrophilic and spheroidal and protrudes into the matrix (red area in the image on the right). ATP synthesis takes place in the F_1 part of ATP-synthase. It operates on the principle of chemiosmosis, which is described below.





Chemiosmosis

Water movement from the direction of higher to lower concentration across a semipermeable membrane is called **osmosis.** When protons (H^+) are pumped from the matrix to intermembrane space by complex I, III, & IV, a proton concentration gradient is developed across the membrane. Since protons are positive charge ions (H^+), there is also an electrical gradient; collectively, it is called an electrochemical gradient. **Chemiosmosis** is the flow of ions (H^+) across a semipermeable membrane down their electrochemical gradient.

An electrochemical gradient developed due to protons (H^+) is a form of electrochemical potential energy. ATP-Synthase harnesses the electrochemical potential energy to synthesize ATP as the protons flow through it, just as electricity is generated as water from a dam flows through a turbine. The higher energy ATP is synthesized by condensing lower energy ADP and phosphate P_i in a process called **oxidative phosphorylation**. It takes three to four H^+ to synthesize one ATP by the following overall reaction.

$$ADP + P_i + 4\,H^+_{\text{intermembrane}} \rightleftharpoons ATP + H_2O + 4\,H^+_{\text{matrix}}$$

This is an equilibrium reaction, i.e., when the electrochemical gradient is low, ATP-synthase consumes ATP and returns protons from the matrix to the intermembrane space.

Structure and mechanisms of operation of ATP-synthase

ATP-synthase comprises several protein sub-units. It is in the shape of a mushroom with two major parts: a stem-like part which is a hydrophobic portion embedded in the inner membrane, is called $\mathbf{F_O}$, and a mushroom head-like part that protrudes into the matrix, is called $\mathbf{F_1}$, as illustrated in Figure 9.5.2 a. Note: the subscript in F_O is the letter O, not the number zero.

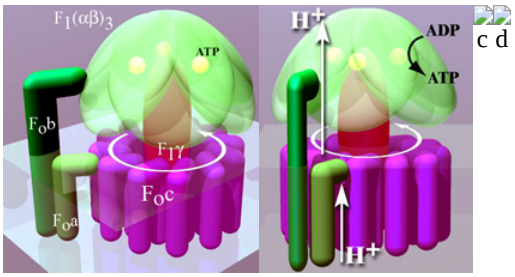


Figure 9.5.2: **a)** Simplified model of ATP-synthase (Copyright; Asw-hamburg, CC BY-SA 2.5, via Wikimedia commons), **b)** illustration proton flow in ATP-synthase (Copyright; Asw-hamburg at German Wikipedia., CC BY-SA 3.0, via Wikimedia Commons), **c)** Mechainsm of rotation of F_0c ring of ATP-synthase (Copyright: Asw-hamburg at German Wikipedia., CC BY-SA 3.0, via Wikimedia Commons), and **d)** Mechanism of ATP production in $F_1\beta$ subunits of ATP synthase. ATP is shown in red, ADP and phosphate in pink and the rotating γ subunit in black (Copyright; By US gov - US gov, Public Domain, https://commons.wikimedia.org/w/inde...?curid=3588598)

F_O portion of \(\ce{ATP}-synthase

F_O part consists of six c subunits (F_Oc) arranged in a ring shape with proton channels between them that make the rotor part of this molecular machinery, as illustrated in Figure 9.5.2 a & b. The b subunit (F_Ob) connects to the F_1 portion and prevents it from rotating. The subunit F_Oa connects F_Ob to the F_Oc ring. F_Oc part is embedded in the membrane and couples the energy released as the protons flow through the channels to kinetic energy in the form of rotation of the F_Oc ring, like a turbine moves by water in a hydroelectric dam. Other subunits are not described here.

F_1 portion of \(\ce{ATP}-synthase

F_1 portion has a set of three α and thee β subunits arranged alternately like carpels of an orange ($F_1(\alpha\beta)_3$) with $F_1\gamma$ in the middle. The $F_1\gamma$ is like an axil connected with the rotary F_Oc, as shown in Figure 9.5.2 a. The $F_1(\alpha\beta)_3$ is prevented from rotating by the F_Ob subunit, but $F_1\gamma$ rotates along with the rotor F_Oc ring. The rotation of $F_1\gamma$ within $F_1(\alpha\beta)_3$ causes conformation changes in the β subunits, as illustrated in the video that will play by clicking this link. The conformational changes in the β subunits are linked to the mechanism of ATP synthesis. Other subunits of the F_1 portion are not described here.

Mechanism of F_Oc ring rotation in \(\ce{ATP}-synthase

The steps of the rotation of the F_Oc ring are illustrated below.

Steps in the rotation mechanism of F_Oc ring in ATP-synthase (Copyright; Asw-hamburg at German Wikipedia., CC BY-SA 3.0, via Wikimedia Commons)

Step a: F_Oc subunits are arranged in a circle. All of them have an amino acid asp residue with an acidic side chain but in neutral ($\text{peptide}-\text{COOH}$) form except one (labeled #1 in step a), which, being near a cation group ($\text{peptide}-\text{NH}_3^+$ of arg residue of F_Ob subunit, is an anion, i.e., $\text{peptide}-\text{COO}^-$. Electrostatic interaction between opposite charges builds mechanical tension in the F_Oc ring like a spiral spring.

1. Step b: A H^+ from outside (intermembrance space) neutralizes the anion ($\text{peptide}-\text{COO}^- + \text{H}^+ \longrightarrow \text{peptide}-\text{COOH}$).
2. Step c: The strained structure relaxes as the electrostatic tension is removed, causing the peptide to rotate.
3. Step d: Twisting motion bring F_Oc#1 back to the other side of the stator F_Ob $\text{peptide}-\text{NH}_3^+$ group causing the rotor F_Oc to rotate by 30^o.
4. Step e: F_Oc#1 subunit is neutral like other subunits in the ring.
5. Step f: The rotor F_Oc is rotated by 30^o, and F_Oc#2 is under the spell of the positive charge of F_Ob $\text{peptide}-\text{NH}_3^+$ group.
6. Step g: The asp residue of F_Oc#2 ionizes the \(\ce{-COOH) ($\text{peptide}-\text{COOH} \longrightarrow \text{peptide}-\text{COO}^- + \text{H}^+$) and the spring comes under tension again.
7. Step h: The asp residue releases the H^+ that moves inside (the matrix).
8. Step i: The initial position of step a is restored and the following has happed in the process:

1. F_Oc#2 has taken the place of F_Oc#1,
2. electrochemical energy of one H^+ movement transferred to kinetic energy through 30° rotation of the rotor, and
3. one H^+ is smuggled from the intermembrane space to the matrix.

The simulation of the movement described above is illustrated in Figure 9.5.2 c.

Mechanism of catalysis of ATP synthesis in $F_1(\alpha\beta)_3$ complex

The $F_1(\alpha\beta)_3$ complex catalyzes ATP synthesis. It has three α and three β subunits arranged alternately as three $\alpha\beta$ dimers in the shape of carpels in an orange. It is the β subunits that catalyze the ATP synthesis. Rotation of $F_1\gamma$ subunit inside the $F_1(\alpha\beta)_3$ complex causes conformational changes in the $F_1\beta$ subunits that are linked with the mechanism of ATP synthesis.

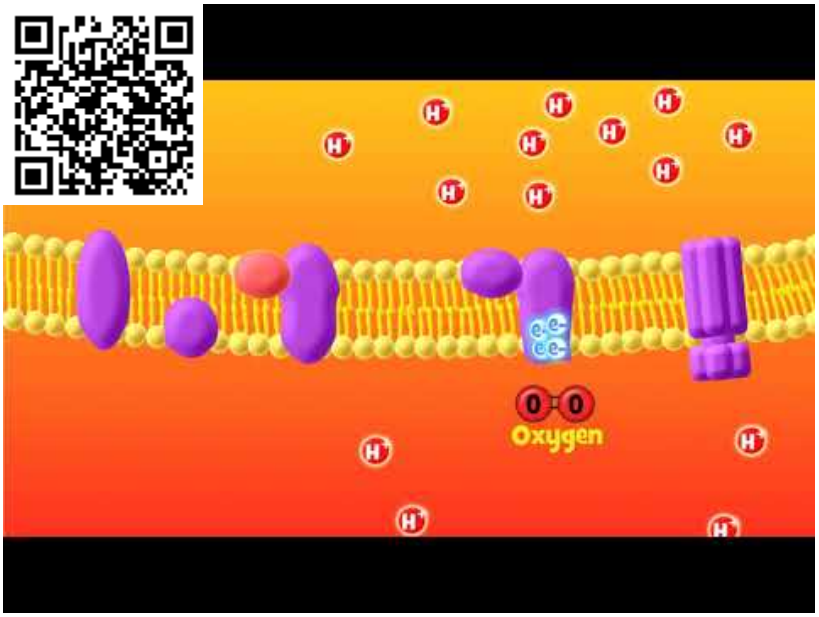
$F_1\beta$ switched between three states with each 360° rotation of $F_1\gamma$ subunit inside the $F_1(\alpha\beta)_3$ complex, as illustrated in Figure 9.5.2 d.

1. First is the **open-state**, shown in brown, where ADP and P_i enter the active state.
2. Second is the **loose-state** in which the $F_1\beta$ closes up around the substrate ADP and P_i binding them loosely, shown in red.
3. Third, is the **tight-state** in which $F_1\beta$ tightens around the substrate ADP and P_i forcing them to condense into an ATP product.
4. Finally, $F_1\beta$ reverts to an open-state that releases the ATP and binds ADP and P_i to repeat the process.

Since one H^+ cause a 30° turn, a complete (360°) turn of $F_1\gamma$ subunit inside $F_1(\alpha\beta)_3$ complex needs 12 H^+ and produces three ATP′s from three $F_1\beta$ subunits in it, that gives the following overall reaction.

$$ADP + P_i + 4\,H^+_{intermembrane} \rightleftharpoons ATP + H_2O + 4\,H^+_{matrix}$$

The next video presents the oxidative phosphorylation of glucose in a summary form.



Heating the body

NADH is reduced in the electron transport chain at the expense of reduction of O_2 by the following overall reaction.

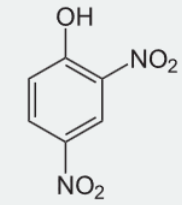
$$\tfrac{1}{2}O_2 + NADH + H^+ \longrightarrow H_2O + NAD^+$$

The potential difference between these redox pairs is 1.14 volts, equivalent to -218 kJ/mol. Reduction of one NADH can produce three ATP. Production of ATP costs 30.5 kJ/mole, which is equivalent to 30.5 kJ/mole x 3 = 91.5 kJ/mole of NADH. So the percentage of energy conserved as ATP, i.e., the **energy efficiency** is: $\frac{91.5}{218} \times 100 = 42\%$. The remaining 58% of energy ends up heating the body.

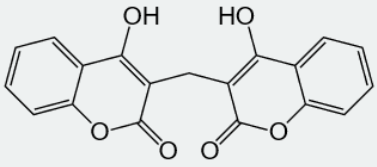
Some compounds, known as **uncouplers**, uncouple electron transport from ATP synthase, i.e., the electron transport pumps H^+ as usual but H^+ return to the matrix without producing ATP′s. So, all energy released during the electron transport

process becomes heat energy. Examples of uncouplers are 2,4-dinitrophenol and dicumarol that combine with H^+ and, being hydrophobic, carry them through the inner membrane. Some compounds, like oligomycin A prevent ATP synthesis by blocking H^+-channels in ATP synthase. Salicylic acid (aspirin), if taken in extreme excess, also blocks H^+-channels in ATP synthase. When the electron transport chain operates without ATP synthesis, all of the electron transport energy is used as heat.

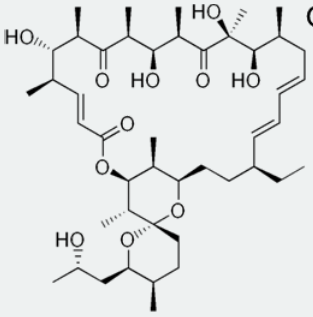
Certain animals adapted to the cold environment have developed uncoupling systems to generate more heat for heating their body. They contain a large amount of **brown fat,** a tissue with many mitochondria, and are brown due to iron in the cytochrome in the mitochondria. The electron transport works in the brown fat. Still, the mitochondria have certain proteins embedded in the inner membrane that allow H^+ to return to the matrix without ATP production. The body of newborn babies loses more heat per unit mass because of the larger surface area to mass ratio. Newborn babies have brown fat deposits on arteries that heat the blood circulating in their bodies. Adults usually do not have brown fat except those who live and work in a cold climate.



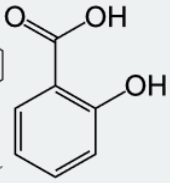
2,4-Dinitrophoenol



Dicumarol



Oligo
mycin A



Salicylic acid

Reactive oxygen species

The electron transport chain uses four electrons and four protons to reduce oxygen by the following overall reaction,

$$O_2 + 4\,e^- + 4\,H^+ \longrightarrow 2\,H_2O\ .$$

However, some electrons, particularly during the reduction of coenzyme-Q in the complex-III, leak and cause the following reaction.

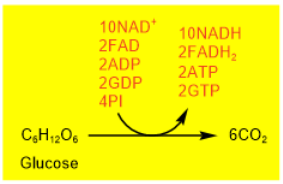
$$O_2 \xrightarrow{e^-} \underset{\text{Superoxide}}{O_2^{\bullet -}} \xrightarrow{e^-} \underset{\text{Peroxide}}{O_2^{2-}}$$

Species like superoxide ($O_2^{\bullet -}$), peroxide (O_2^{2-}) and their product hydroxyl radical ($HO^{\bullet}$) are called **reactive oxygen species** which are harmful. They damage proteins, cause mutations by reacting with DNA, and are responsible for aging, as illustrated in the figure on the right (*taken from* *https://www.hiclipart.com/free-trans...-clipart-npvtr*). The body has mechanisms to suppress the production of reactive species and destroy them if formed. Production of reactive oxygen species increases with an increase in the inter-member potential. The reactive oxygen species, oxidants, activate uncoupling proteins that reduce the membrane potential. Some substances in the body, e.g., vitamins C, E and antioxidant enzymes, react with and destroy the reactive oxygen species.

Summary of glucose catabolism

Aerobic catabolism of a glucose molecule converts a 6 C′s glucose molecule to six carbon dioxide along with ATP′s, GTP′s, and reduced coenzyme NADH′s, and $FADH_2$′s, as shown in the figure on the right.

Glucose first goes through glycolysis that produces two pyruvate and overall two ATP and two NADH. Two pyruvates go through oxidative decarboxylation producing two acetyl-CoA and two NADH. Each acetyl-CoA goes through one turn of the citric acid cycle. So, the two turns of the citric acid cycles convert two acetyl-CoA into four CO_2, six NADH, two $FADH_2$, and two GDP. These steps are summarized in Figure 9.5.3 and Figure 9.5.4 shown below.



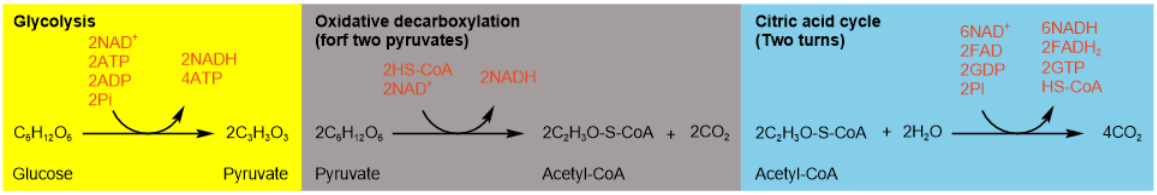


Figure 9.5.3 illustrates the summary of aerobic catabolism of glucose and production of ATP′s, GTP′s, and reduced coenzyme NADH′s, and $FADH_2$′s at different stages.

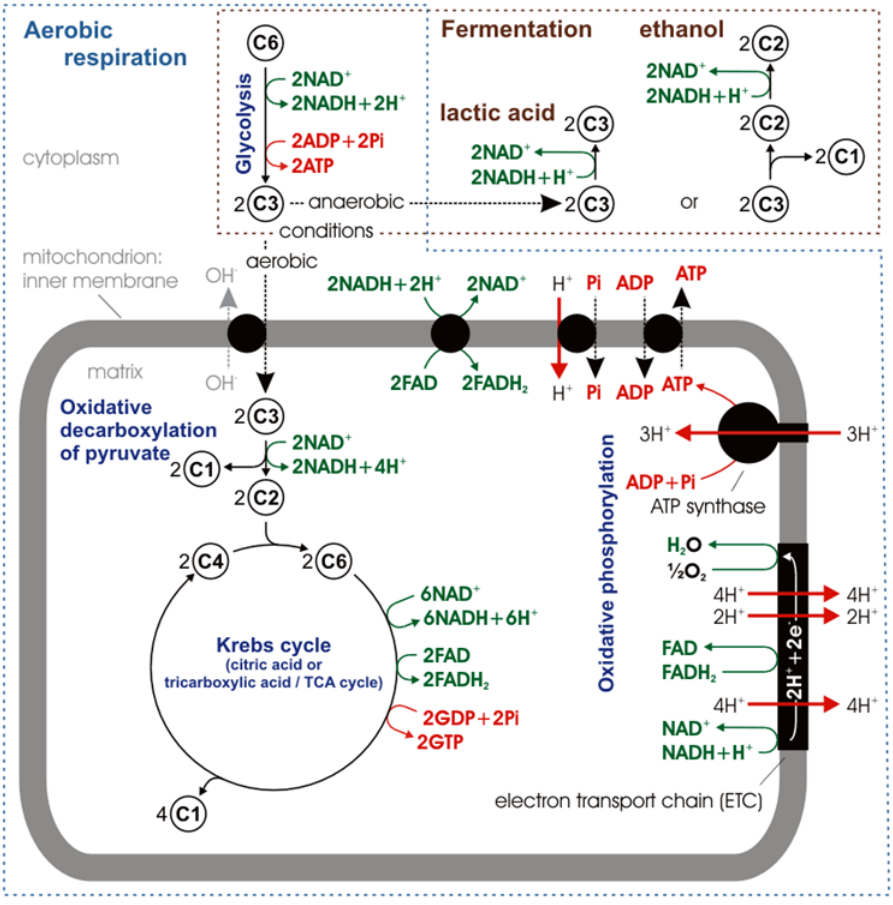


Figure 9.5.4: Stoichiometry of aerobic respiration and most known fermentation types in the eucaryotic cell. Numbers in circles indicate counts of carbon atoms in molecules; C6 is glucose $C_6H_{12}O_6$, and C1 carbon dioxide CO_2. The mitochondrial outer membrane is omitted. (Copyright; Darekk2, CC BY-SA 3.0, via Wikimedia Commons)

Insulin -the blood glucose-regulating hormone

Insulin is a hormone that regulates blood glucose levels. The glucose level increases after a meal, and the insulin promotes its absorption into the liver, fat, and muscle cells, converting it into glycogen or fats. High levels of insulin also inhibit the production of glucose by the liver. Low insulin levels have the opposite effects, which happen when a person has diabetes.

Comparison of ATP yield in aerobic and anaerobic catabolism of glucose

Anaerobic catabolism, i.e., the two forms of fermentation shown in Figure 9.5.3 yields a net two ATP′s from a glucose molecule.

Aerobic catabolism yields about thirty ATP′s. Two ATP′s are produced before the citric acid cycle and two GTP′s, ten NADH′s, and two $FADH_2$′s are produced during citric acid cycle. GTP′s later convert to ATP′s. the reduced coenzymes, i.e., NADH′s and $FADH_2$′s enter the electron transport chain and produce heat and ATP′s as explained in the next section.

Account of ATP production from complete oxidation of glucose

Two ATP′s are produced before the citric acid cycle along with four NADH. The citric acid cycle yields GTP′s, six NADH′s, and two $FADH_2$′s. GTP′s convert to ATP′s, i.e., one GTP is equivalent to one ATP. NADH′s and $FADH_2$′s enter the electron transport chain and produce heat and ATP′s. Recall that NADH is more energetic and pumps more proton through complex I, III, and IV than $FADH_2$ which is less energetic and pumps less protons through complex III and IV only. Therefore, NADH yields more ATP′s than $FADH_2$. Earlier books state that one NADH′s is equivalent to three ATP′s, but currently accepted average values are the following.

$$\mathrm{GTP} \approx 1\mathrm{ATP}$$

$$\mathrm{NADH} \approx 2.5\mathrm{ATP}$$

$$\mathrm{FADH_2} \approx 1.5\mathrm{ATP}$$

Based on these conversions, complete aerobic respiration of glucose produces approximately 32 ATP′s, as shown in the following equation and detailed in Table 1.

$$\underbrace{\mathrm{C_6H_{12}O_6}}_{\text{Glucose}} + 6\,\mathrm{O_2} + 10\,\mathrm{NAD^+} + 2\,\mathrm{FAD} + +2\,\mathrm{GTP} + 2\,\mathrm{ADP} + 4\,\mathrm{Pi} \longrightarrow 6\,\mathrm{CO_2} + 6\,\mathrm{H_2O} + 10\,\mathrm{NADH} + 2\,\mathrm{FADH_2}$$
$$+ 2\,\mathrm{GTP} + 2\,\mathrm{ATP}$$

$$\underbrace{\mathrm{C_6H_{12}O_6}}_{\text{Glucose}} + 6\,\mathrm{O_2} + 32\,\mathrm{ADP} + 32\,\mathrm{Pi} \longrightarrow 6\,\mathrm{CO_2} + 6\,\mathrm{H_2O} + 32\,\mathrm{ATP}$$

Table 1: ATP Production from complete aerobic catabolism of one glucose molecule.

Reaction	ATP or reduced coenzymes	Total ATP′s output
Glycolysis		
glucose → glucose-6-phosphate		-1
fructose 6-phosphate → fructose 1,6-bisphosphate		-1
2 glyceraldehyde-3-phosphate → 2 1,3-bisphosphoglycerat	2 NADH	5
2 1,3-bisphosphoglycerate → 2 3-phosphoglycerate	2 ATP	2
2 phosphoenolpyruvate → 2 pyruvate	2 ATP	2
Oxidative phosphorylation of two pyruvate		
2 pyruvate → 2 acetyl-CoA	2 NADH	5
Two turns citric acid cycles to consume to 2 acetyl-CoA		
2 isocitrate → 2 α-ketoglutarate	2 NADH	5
2 α-ketoglutarate → 2 succinyl-CoA	2 NADH	5
2 succinyl-CoA → 2 succinate	2 GTP	2
2 succinate → 2 fumarate	2 $FADH_2$	3

Reaction	ATP or reduced coenzymes	Total ATP′s output
2 malate → 2 oxaloacetate	2 NADH	5
Total		**32**

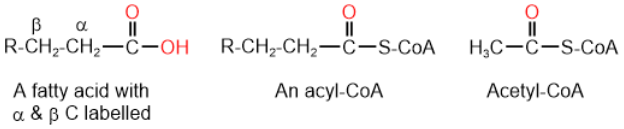
9.6: Oxidation of fatty acids

Learning Objectives

- Understand the catabolism of fatty acids, i.e., β-oxidation of fatty acids, the reactions involved in the process, the number of cycles needed, and the calculation of ATP yield per fatty acid.
- Understand obesity, ketogenesis, ketosis, ketoacidosis, and catabolism of glycerol, associated with β-oxidation of fatty acids.

β-Oxidation of fatty acids

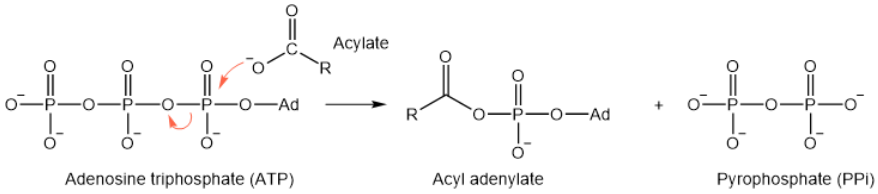
Although glucose is a quick energy source that animals need, fatty acids have higher energy per unit mass. Therefore, fats are the main energy storage compounds in animals. Fatty acids ($\mathrm{R-COOH}$) in fats are a major source of energy for animals. Hydrolysis of fats, i.e., the 1st stage of catabolism of fats, begins in the digestive tract and completes in the cytosol of the cells. The second stage of fatty acid catabolism starts with the conversion of ($\mathrm{R-COOH}$) to the thioester group of coenzyme A, ($\mathrm{R-CO-S-CoA}$), i.e., an acyl-CoA, at the cost of energy of ATP to AMP conversion. The thioester is called an activated fatty acid. Some of the important terms used in this section are illustrated in the figure below.



Acyl-CoA is transported from the cytosol into mitochondria for the second stage of its catabolism. In the second stage, the Acyl-CoA is fragmented into two C′s fragments in the form of acetyl-CoA. The activated fatty acid goes through a set of four reactions in which the βC is oxidized to a carbonyl C=O group. Another coenzyme A makes thioester group with the β-C=O, and an acetyl-CoA group is split off, leaving behind an acyl-CoA with two C′s less than the initial acyl-CoA. This process is called **β-Oxidation of fatty acids**. The process repeats on the fragment acyl-CoA again and again until the last acyl-CoA fragment left is acetyl-CoA. The acetyl-CoA enters the citric acid cycle, which is the third stage of the catabolism, as in the case of the catabolism of glucose. These reactions are explained next.

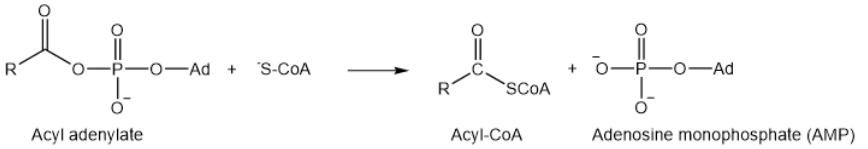
Activation of fatty acids

The activation of a fatty acid begins with an SN_2 reaction of acylate anion as a nucleophile with ATP. It produces acyl-adenylate and pyrophosphate, as shown below.



Pyrophosphate is removed by hydrolysis, making the reaction irreversible: $\mathrm{PPi + H_2O \longrightarrow 2\,Pi + 2\,H^+}$.

Acyl-adenylate goes through a nucleophilic acyl substitution reaction with coenzyme A ($\mathrm{HS-CoA}$) attacking as a nucleophile and AMP as leaving group.

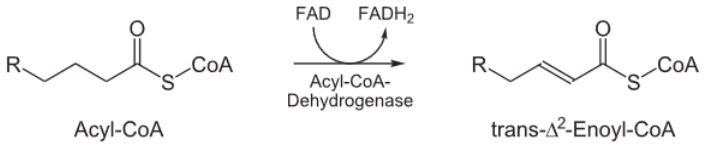


As described next, acyl-CoA product is transported from the cytosol into mitochondria for the β-oxidation process.

Reaction 1 -oxidation

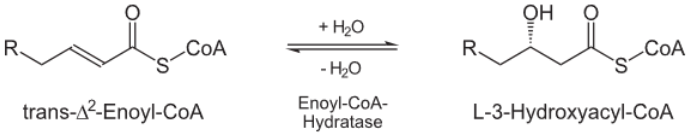
Reaction 1 is the dehydrogenation of of acyl-CoA by enzyme acyl-CoA-dehydrogenase that removes H from α and β C′s generating a trans C=C bond. An FAD is reduced to $FADH_2$ as the acyl-CoA is oxidized in this reaction. Resonance

stabilization of the $C=C$ bond by conjugated $C=O$ group makes the product stable, lowering the activation energy for the reaction.



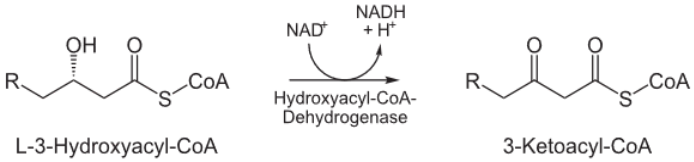
Reaction 2 -hydration

The $C=C$ is hydrated by addition of H_2O that selectively installs an $-OH$ group at βC, creating a new chiral center in L-configuration. This is an electrophilic addition reaction. The $C=O$ makes the βC more nucleophilic by withdrawing electrons from it by resonance. That is why the βC selectively reacts with the incoming H_2O electrophile.



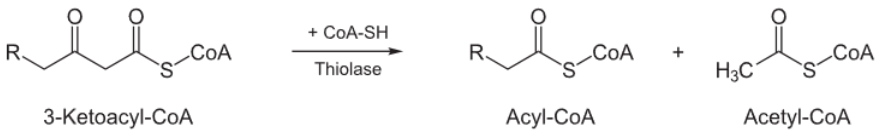
Reaction 3 -oxidation

The secondary $-OH$ group is oxidized to a ketone ($C=O$) group at the expense of reduction of NAD^+ to $NADH$ by the following reaction.



Reaction 4 -cleavage by thiolysis

The fourth reaction is a nucleophilic acyl substitution reaction in which an anion of coenzyme A (^-S-CoA) acts as a nucleophile, the ketone ($C=O$) as electrophile and $^-CH_2-CO-S-CoA$ as the leaving group. $^-CH_2-CO-S-CoA$ is a good leaving group because the negative charge is resonance stabilized and mainly resides on O of carbonyl ($C=O$) group. A proton later neutralizes the negative charge.



The second product of the above reaction is an acyl-CoA with two $2\ C's$ less than the initial acyl-CoA.

Number of β-oxidation cycles

The product acyl-CoA of the first β-oxidation cycle goes through the process of four reactions again and again until the last acyl-CoA product is acetyl-CoA, i.e., $CH_3CO-S-COA$. The process is illustrated in Figure 9.6.1, with the help of an example of β-oxidation of stearic acid which is a typical $18\ C's$ saturated fatty acid.

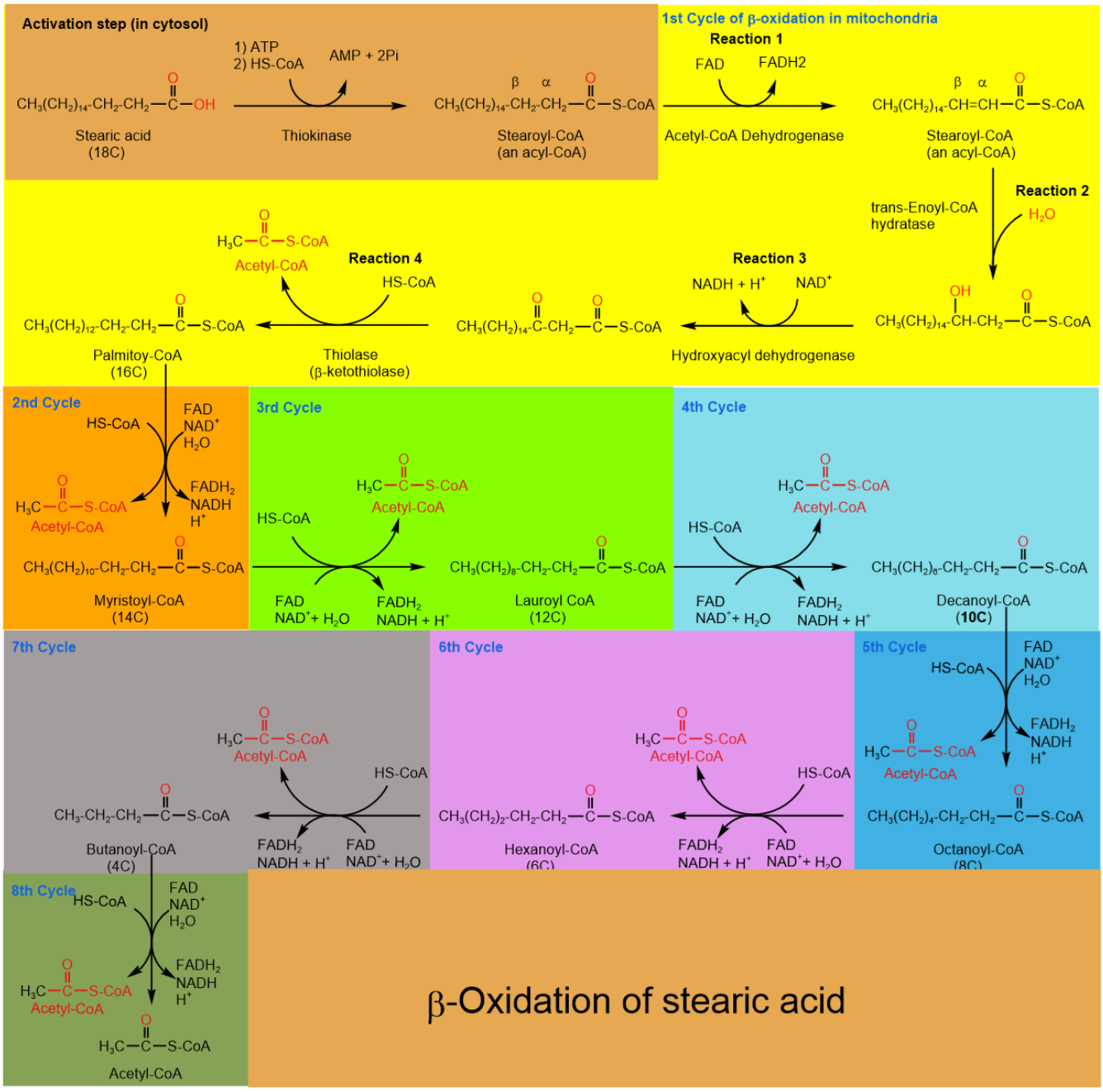


Figure 9.6.1: β-Oxidation of fatty acids illustrated with the help of β-oxidation of stearic acid. (Copyright; Public domain)

If the starting fatty acid has nC′s, there are: $\frac{n}{2}$ acetyl-CoA produced. Since the last cycle produces two acetyl-CoA, the number of cycles for a fatty acid containing nC′s is: $\frac{n}{2}-1$. For example, stearic acid has $18\,\mathrm{C}'\mathrm{s}$ and it goes through: $\frac{18}{2}-1=8\ \beta$-oxidation cycles.

Summary of β-oxidation of fatty acids

ATP Yield from fatty acid oxidation

Each β-oxidation cycle of a fatty acid yields one $\mathrm{FADH_2}$, one NADH and one acetyl-CoA. These molecules produce $\mathrm{ATP'}$s when they enter the citric acid cycle and oxidative phosphorylation, as described earlier in the catabolism of glucose. One $\mathrm{FADH_2}$ ≈ 1.5 $\mathrm{ATP'}$s, one NADH ≈ 2.5 $\mathrm{ATP'}$s, and one acetyl-CoA ≈ 10 $\mathrm{ATP'}$s that makes 14 $\mathrm{ATP'}$s per β-oxidation cycle. Since there are $(\frac{n}{2}-1)$ or $(0.5\times n-1)$ cycles for a fatty acid containing nC′s, there are: $((0.5\times n-1)\times 14))\ \mathrm{ATP'}\mathrm{s}$ produced. The last cycle yields two acetyl-CoA, so there are addition 10 $\mathrm{ATP'}$s from the the last acetyl-CoA. The activation step converts one ATP to one AMP which is equivalent to conversion of 2 ATP to 2 ADP. So, after adding ten from the last acetyl-CoA to the formula ($(0.5\times n-1)\times 14)\ \mathrm{ATP'}\mathrm{s}$ and subtracting 2 for the 2 $\mathrm{ATP'}$s consumed in the activation step, the formula for the number of $\mathrm{ATP'}$s produced per fatty acid containing nC′s become the following:

$$\mathrm{ATP'}\text{s produced per fatty acid containing nC's} = ((0.5\times n-1)\times 14+10-2)\ \mathrm{ATP'}\mathrm{s}\quad .$$

It simplifies to:

$$\mathrm{ATP'}\text{s produced per fatty acid containing nC's} = (7\times n-6)\ \mathrm{ATP'}\mathrm{s}\quad .$$

For example, stearic acid contains 18 C′s and produces $(7\times 18-6)=120\ \mathrm{ATP'}\mathrm{s}$. So, one mole of stearic acid (284.48 g/mol) produces 120 mol $\mathrm{ATP'}$s, which is significantly higher than 32 mol $\mathrm{ATP'}$s produced by one mole of glucose (180.156 g/mol). When converted to moles of $\mathrm{ATP'}$s produced per unit mass, stearic acid produces 0.42 mol $\mathrm{ATP'}$s/g, which is more than two times higher than 0.18 mol $\mathrm{ATP'}$s/g of glucose. This is because the average oxidation state of glucose C′s is higher than those of fatty acid C′s. Glucose is a quick energy source, but animals use fats as relatively longer-term energy storage molecules due to their higher energy density.

Unsaturated fatty acids

Unsaturated fatty acids have C=C bond in their alkyl chain. β-Oxidation of unsaturated fatty acids happens the same way as for saturated fatty acids except for the following changes. If C=C bond is a *trans*-C=C bond between α and β C′, reaction 1 in the β-oxidation is not needed and $FADH_2$ from reaction 1 is not produced. If C=C bond is a cis-C=C bond or it is between β and γ C′, It goes through isomerization reactions catalyzed by isomerase enzymes to convert the C=C bond into a *trans* C=C bond between α and β C′ before reaction 3 happens on it. Unsaturated fatty acids yield a little less ATP′s due to one $FADH_2$ less produced pre C=C bond, but the difference is small and the formula: $(7 \times n - 6)$ ATP′s , still gives a reasonably accurate estimate of ATP yield.

Fat and obesity

Although carbohydrates and glucose are quick energy sources, fat has more energy per mass or volume. Storage of fats for long-term energy supply is an important survival feature for several animals. For example, hibernating animals like polar bear converts food into fats during summer when food is plenty and utilize stored fats to survive for months without food during hibernation in winter. Whales and penguins are kept warm by a layer of body fat called blubber and use it as an energy source to survive when food is unavailable. Camel store fat in their hump and can survive for months without food and water by utilizing the fat reserves. Migrator birds also store fat and use it as an energy source during migration.

Humans can store fats. In earlier times, major human food was vegetables, and fats in food accounted for about 20% of food calories, but these days more than 60% of food calories are from fats. The cumulation of too much fat is associated with overweight and obesity. **Obesity** is an excess accumulation of body fat that negatively affects health. It is measured in terms of **body mass index (BMI)**, which is the ratio of a person's mass to the square of a person's height. BMI 18.5 kg/m^2 to 24.9 kg/m^2 is normal, less is underweight, more is overweight, and over 30 kg/m^2 is obese, as shown in Figure 9.6.2. Obesity is associated with medical conditions like diabetes, high blood pressure, stroke, heart disease, gallstones, arthritis, and some cancers. According to NIH, nearly 3 out of 4 adults aged 20 or above in the US are overweight or obese. It is a preventable condition, as described in the following video message from NIH.

US national research authority video message.

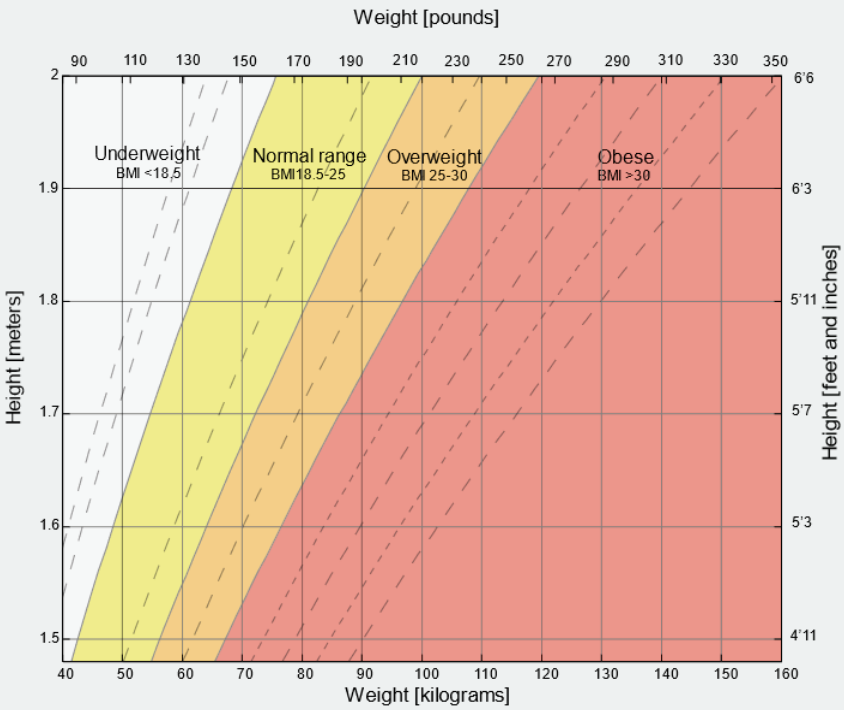
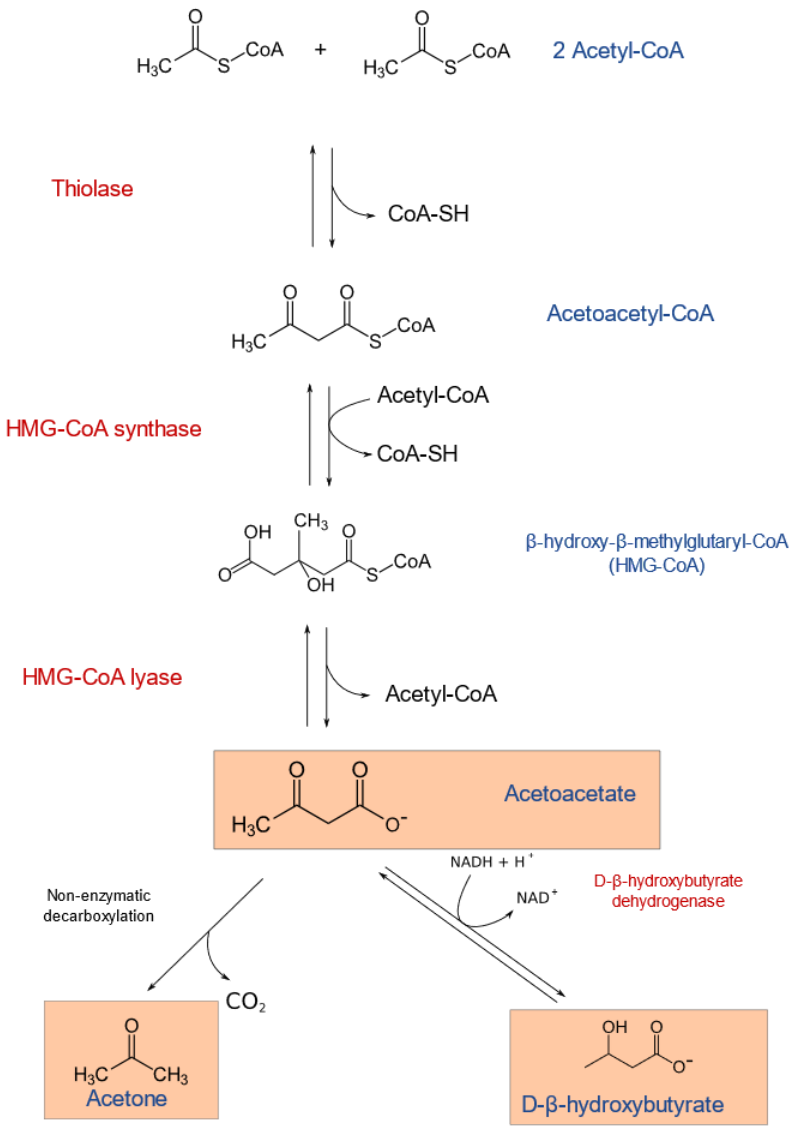


Figure 9.6.2: Body mass versus height chart showing ranges of underweight, normal weight, overweight, and obese regions based on body mass index (BMI) that can be read from the region marked by dotted lines in the chart at the cross-point of vertical lines starting from weight and horizontal lines starting from person's height. (Copyright; InvictaHOG, Public domain, via Wikimedia Commons)

Like insulin regulates blood glucose, **leptin** is a hormone adipose tissues produce. Leptin governs the long-term balance between food intake and energy expenditure. When there is more adipose tissue (fatty tissues), more leptin is released, which decreases hunger by giving a feeling of fullness, and less fatty tissue means less fat and less leptin, which has the opposite effect. Like diabetes disturbs the function of insulin, obesity may be related to the malfunctioning of leptin. This is an active area of research these days for finding obesity treatment.

Ketone bodies

Acetyl-CoA produced during the β-oxidation of fatty acids enter the citric acid cycle. When many fatty acids degrade, the citric acid cycle can not take all of the acetyl-CoA. The excess acetyl-CoA accumulates in the liver and converts into ketone bodies by a pathway known as **ketogenesis**, as shown in the figure on the right (*Copyright; Sav vas, CC0, via Wikimedia Commons*). In this figure, the enzymes are colored red, and the substrates/products are colored blue. The three **ketone bodies**, i.e., acetoacetate, acetone, and β-hydroxybutyrate, are marked within orange boxes.



Ketone bodies are transported from the liver to other tissues where acetoacetate and β-hydroxybutyrate are converted back to acetyl-CoA and enter the citric acid cycle to produce energy. Normally there is constant production and consumption of acetone bodies maintaining ~1 mg/dL concentration in blood.

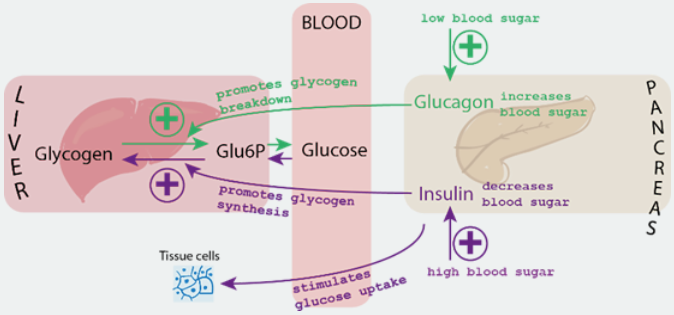
Acetone cannot be converted back into acetyl-CoA except in the liver, where it can be converted into lactate and then to pyruvate.

Ketosis

When the synthesis rate of ketone bodies exceeds their consumption rate, they start accumulating and may start excreting in urine and via breathing -a condition called **ketosis**. Since two of the three ketone bodies are acids, their accumulation in the blood lowers the blood pH -a condition called **ketoacidosis**. The smell of acetone, fruity or like nail polish remover smaller, is detectable in the breath of persons suffering from ketosis. Ketone bodies can be tested in blood, urine, or breath, e.g., by using test strips with a color chart to read the results, as shown in the figure on the left. In ketosis, the ketone bodies in the blood are between 0.5 to 3.0 millimole/L (mmol/L). In ketoacidosis, the concentration may go above 10 mmol/L.

Ketosis can happen during fasting, starvation, prolonged low carbohydrate diets, prolonged intense exercises, alcoholism, or due to uncontrolled diabetes.

Diabetes and ketone bodies

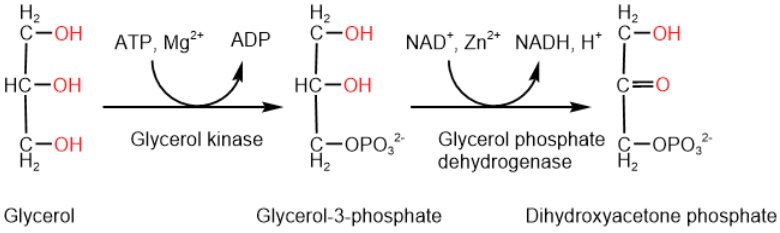


The liver and pancreas play important roles in maintaining blood glucose levels, illustrated in the figure on the right (*copyright: C. Muessig, CC BY-SA 3.0, via Wikimedia Commons*) and described next. The normal glucose level in the blood is 4.5 to 5.5 mM. When glucose level is above normal, usually after a meal, the pancreas senses it and secrets **insulin** that increases the flow of glucose into muscles and fatty tissues, where it is converted into glycogen, lowering blood glucose. When the blood glucose level is low, the pancreas secretes another hormone, **glucagon**, into the bloodstream. Glucagon triggers glycogen breakdown, particularly in the liver, releasing glucose into the bloodstream.

In **diabetes**, there is either insufficient insulin or not functioning properly. Less glucose is sent to muscles and, as a result, muscles cause β-oxidation of fatty acids to meet the energy needs. Higher levels of acetyl-CoA from the β-oxidation produce more ketone bodies than their consumption, accumulating ketone bodies. It causes **ketosis** or **ketoacidosis**. Low pH of blood plasma due to ketoacidosis triggers kidneys to excrete urine with high acid levels. Glucose and ketone bodies also spill into the urine. High glucose concentration in the blood causes higher osmotic pressure, increasing dehydration. The symptoms include frequent urination and excessive thirst. These conditions can be treated with low carbohydrate diets, insulin therapy, or other medications to control diabetes.

Catabolism of glycerol

Glycerol, a part of triglycerides (fats), accounts for about 5% of their energy. Glycerol is first converted into glycerol-3-phosphate by an SN_2 reaction of a primary alcohol group of glycerol acting as a nucleophile, phosphorus in the phosphate of ATP as a nucleophile, and ADP as a leaving group. Then, the secondary alcohol of glycerol is oxidized to a ketone group at the expense of reduction of a NAD^+ coenzyme into NADH as shown below.



Dihydroxyacetone phosphate produced by the second reaction is an intermediate in the glycolysis of glucose and enters the same catabolic pathway.

9.7: Degradation of amino acids

Learning Objectives

- Understand the catabolism of amino acids, including transamination, oxidative amination, and urea cycle that takes care of the N and processing of the C skeleton of the amino acid to intermediates that enter into citric acid cycle for energy production.

Amino acids are the products of stage 1 of protein catabolism. Usually, amino acids are used to synthesize proteins and other substances that need nitrogen, e.g., nucleotides. Only about 10% of our energy is usually derived from amino acids. Still, when there is a shortage of carbohydrates and fats, e.g., during fasting or starvation, amino acids are used as an alternate energy source. Prolonged use of amino acids as an energy source may lead to the destruction of essential tissues.

Degradation of amino acids has two sets of catabolic pathways: one that deals with the processing of N and the other that deals with the processing of the remaining C skeleton of the amino acids.

Processing of N of amino acids

Processing of N usually happens in the liver and has three major stages: i) transamination, ii) oxidative deamination, and iii) urea cycle, as illustrated with some examples in Figure 9.7.1.

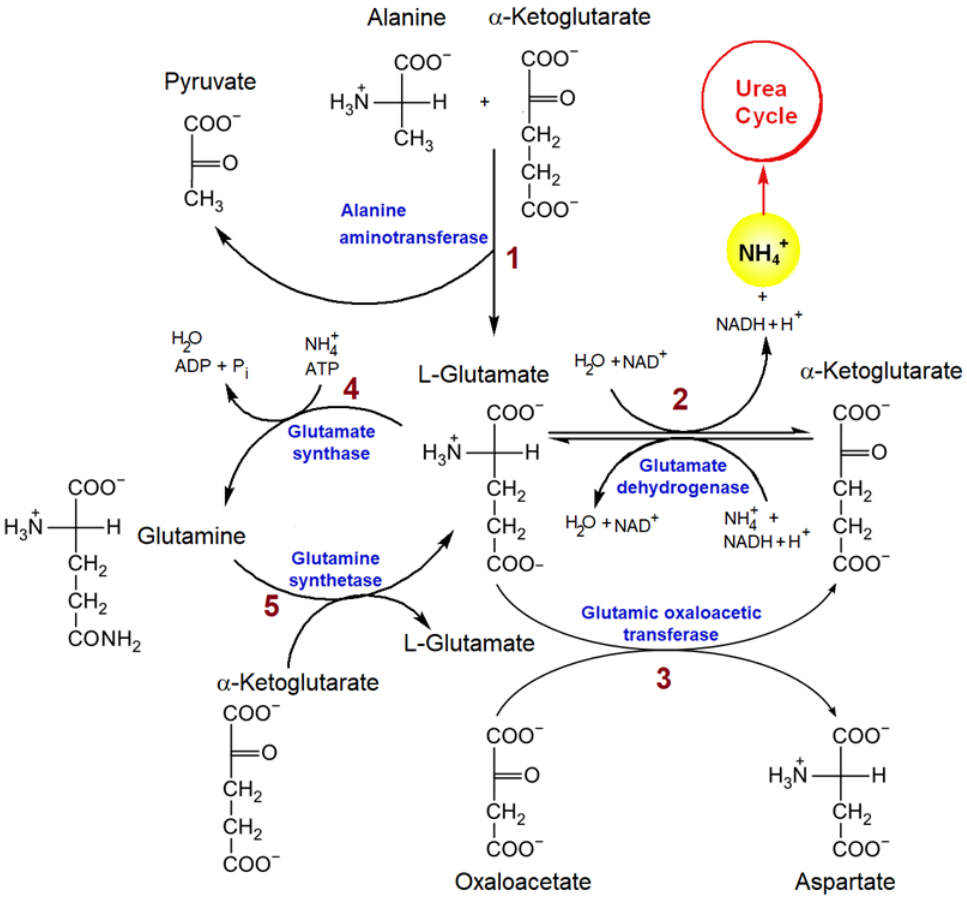


Figure 9.7.1: Processing of N of amino acids Illustrated, including examples of transamination (Reactions 1, 3), oxidative deamination (reaction 2), and some other ways of $-NH_2$ transfer (reactions 4 and 5) and production of NH_4^+ (highlighted yellow) that enters the urea cycle. (Copyright; Public domain)

Transamination

Transamination is the process of exchange of an ammonium ($-NH_3^+$) group of amino acids with ketone (C=O) group of an α-ketoacid. Usually α-amino acids exchange $-NH_3^+$ group with ketone (C=O) group of an α-ketoglutarate. For example, alanine

exchanges its $-NH_3^+$ group with a $C{=}O$ of an α-ketoglutarate producing a new α-keto acid (pyruvate in this example) and a new α-amino acids (L-glutamate in this example), as shown in reaction 1 in Figure 9.7.1. Other α-keto acids can also receive $-NH_3^+$ of amino acids also. For example, L-glutamate can transfer its $-NH_3^+$ to oxaloacetate that regenerates α-ketoglutarate and produces aspartate, as shown in reaction 3 in Figure 9.7.1.

Oxidative deamination

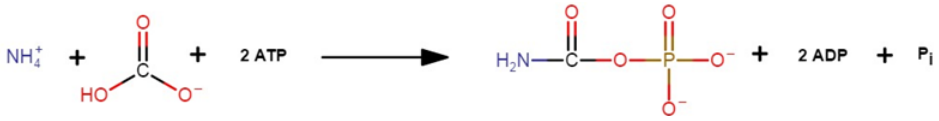
Oxidative deamination reaction replaces $-NH_3^+$ with a $C{=}O$, producing an α-keto acid and an ammonium (NH_4^+) ion, as shown in reaction 2 in Figure 9.7.1. This is an oxidation reaction that is coupled with the reduction of NAD^+ to $NADH$. It happens in mitochondria where $NADH$ enters oxidative phosphorylation pathway to produce $ATP's$. The other product, i.e., NH_4^+, is a toxic substance that needs to be disposed off. Some of NH_4^+ may be used in anabolic reactions, e.g., reactions 4 and 5 in Figure 9.7.1, but most of it enters the urea cycle.

Urea cycle

Ammonium ion NH_4^+ is a toxic substance that is converted to less toxic urea in the urea cycle by the following overall reaction.

$$2\,NH_4^+ + CO_2 \longrightarrow \underbrace{H_2N-\overset{\overset{O}{\|}}{C}-NH_2}_{\text{Urea}} + 2\,H^+ + H_2O \quad .$$

Before entry into the urea cycle, NH_4^+ reacts with two $ATP's$ and bicarbonate (HCO_3^-), producing carbamoyl phosphate, two $ADP's$ and phosphate (Pi, as shown below.



This reaction is catalyzed by an enzyme carbamoyl phosphate synthetase I. Blood contains HCO_3^- which is the product of dissolution of CO_2. The product of the above reaction, i.e., Carbamoyl phosphate, goes through four reactions called the **urea cycle** illustrated in Figure 9.7.2.

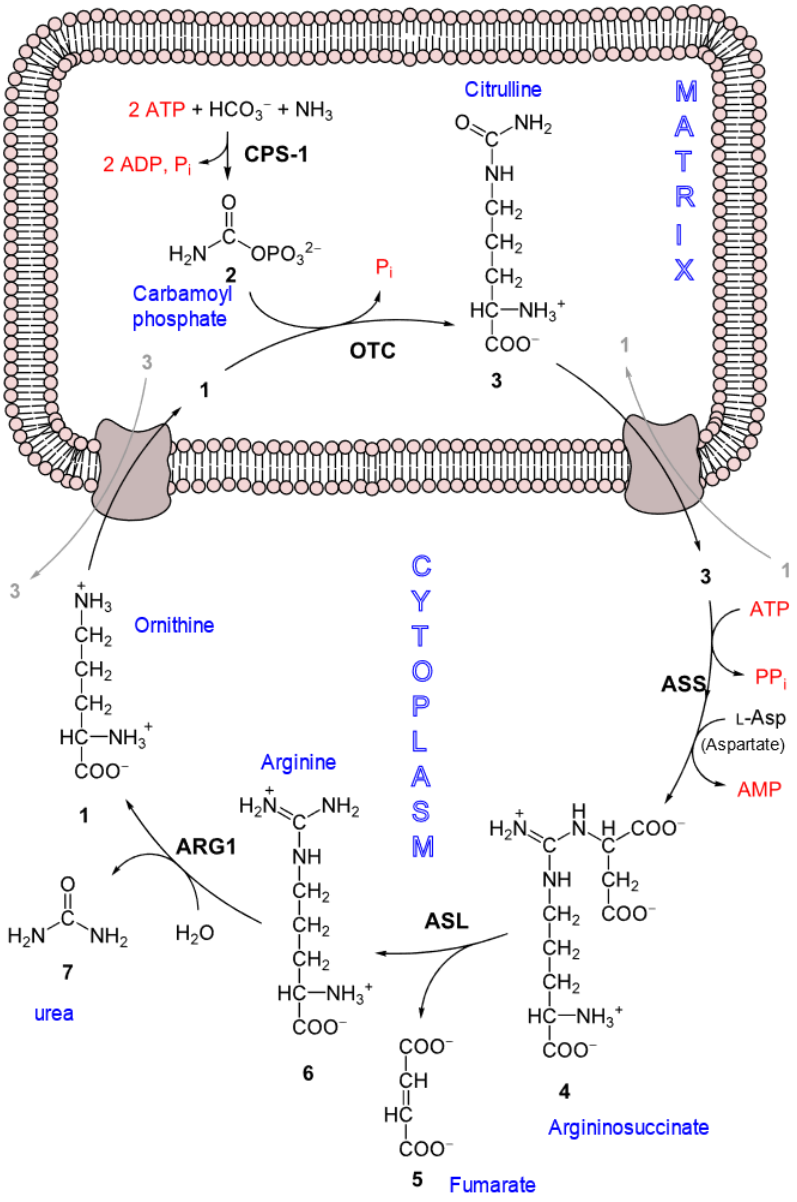
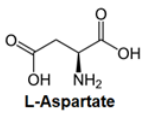


Figure 9.7.2: Urea cycle: CPS-1 is carbamoyl phosphate synthetase-1, OTC is ornithine transcarbamoylase, ASS is argininosuccinate synthetase, ASL is argininosuccinate lyase, and ARG1 is Arginase. Note: Inner membrane is shown, but the outer membrane of the mitochondrion is ignored in this diagram. (Copyright; Pk0001, CC BY-A 4.0, via Wikimedia Commons)

- The first reaction of the urea cycle is catalyzed by ornithine transcarbamylase. The $-\mathrm{COO}^-$ group of ornithine substitutes phosphate ($\mathrm{P_i}$ from carbamoyl phosphate producing citrulline in mitochondria. Citrulline moves out from the matrix into the cytoplasm.
- The second reaction is a condensation reaction between the carbonyl ($\mathrm{C{=}O}$) group of citrulline and $-\mathrm{NH_3^+}$ group of aspartate that produces argininosuccinate. This reaction takes place in the cytosol and is catalyzed by argininosuccinate synthetase.
- In the third reaction, argininosuccinate is cleaved by the enzyme argininosuccinase to form arginine, which stays in the cycle, and fumarate, which leaves the cycle.
- In the fourth reaction, arginase cleaves arginine to form urea and ornithine. Ornithine is transported back to the matrix in mitochondria to start the next urea cycle, and urea is released into the blood.



Since $\mathrm{HCO_3^-}$ is produced by the dissolution of $\mathrm{CO_2}$ in water, the $\mathrm{NH_4^+}$ + $\mathrm{HCO_3^-}$ are equivalent to $\mathrm{NH_3}$ + $\mathrm{CO_2}$ + $\mathrm{H_2O}$. So, the overall equation of the urea cycle becomes:

$$\mathrm{NH_3 + CO_2 + H_2O + aspartate + 3\,ATP + 3\,H_2O \longrightarrow urea + fumarate + 2\,ADP + 2\,P_i + AMP + PP_i + H_2O}$$

Since NH_3 is removed to convert aspartate into fumarate along with $PP_i + H_2O \longrightarrow 2\,P_i$, substituting these for aspartate and fumarate simplifies the above equation into:

$$2\,NH_3 + CO_2 + 3\,ATP + 3\,H_2O \longrightarrow \text{urea} + 2\,ADP + 4\,P_i + AMP$$

One $NADH$ is produced during oxidative deamination of L-glutamate (reaction 2 in Figure $9.7.1$). Another $NADH$ is produced when fumarate from reaction 3 of the urea cycle is processed in the citric acid cycle, i.e., fumarate is converted into malate by enzyme fumarase and then malate is oxidized to oxaloacetate at the expense of reduction of NAD^+ to $NADH$ by enzyme malate dehydrogenase. Adding these reactions to the above reaction of the urea cycle results in the following overall reaction.

$$CO_2 + \text{glutamate} + \text{aspartate} + 3\,ATP + 2\,NAD^+ + 3\,H_2O \longrightarrow \text{urea} + \alpha\text{−ketoglutarate} + \text{oxaloacetate} + 2\,ADP + 2\,P_i + AMP + 2\,PP_i + 2\,NADH$$

Three ATP are consumed, one of them to AMP that makes total equivalent to four ATP's consumed but two $NADH$ enter oxidative phosphorylation and yield ~5ATP, i.e., there is net production of an ATP in the process of N of an amino acid.

The urea released into the blood is later filtered out by the kidneys and excreted with urine. An adult passes about 25 to 30 g of urea in urine per day. If urea is not eliminated properly, it builds to toxic levels and needs medical treatment, like dialysis. Reducing protein intake also lowers urea output.

Urea cycle

The link between the urea cycle and the citric acid cycle

The urea cycle and citric acid cycle are two separate cycles but are linked with each other. Aspartate that provides one N in the urea cycle is produced by oxaloacetate transamination, which is an intermediate in the citric acid cycle. Fumarate is a product of the urea cycle that is an intermediate of the citric acid cycle and returns to it.

processing of C skeleton of amino acids

α-Ketoacids are left as C skeleton of amino acid after transamination or oxidative deamination. For example, pyruvate is left after transamination of alanine and α-ketoglutarate is left after oxidative deamination of glutamate, as shown in reactions 1 and 2, respectively, in Figure $9.7.1$. α-Ketoacids are either used as precursors for the synthesis of other compounds, or they enter the citric acid cycle to produce CO_2, H_2O, and energy. α-Ketoglutarate and oxaloacetate are α-keto acids that are intermediates in the citric acid cycle and directly enter the cycle. Other α-keto acids go through a series of reactions to convert into one of the intermediates in the citric acid cycle, or they convert to pyruvate or acetyl-(\ce{CoA}\) that enter the citric acid cycle, as illustrated in Figure $9.7.3$. Details of the conversation of α-keto acids into intermediates of the citric acid cycle or pyruvate or acetyl-(\ce{CoA}\) are not described here.

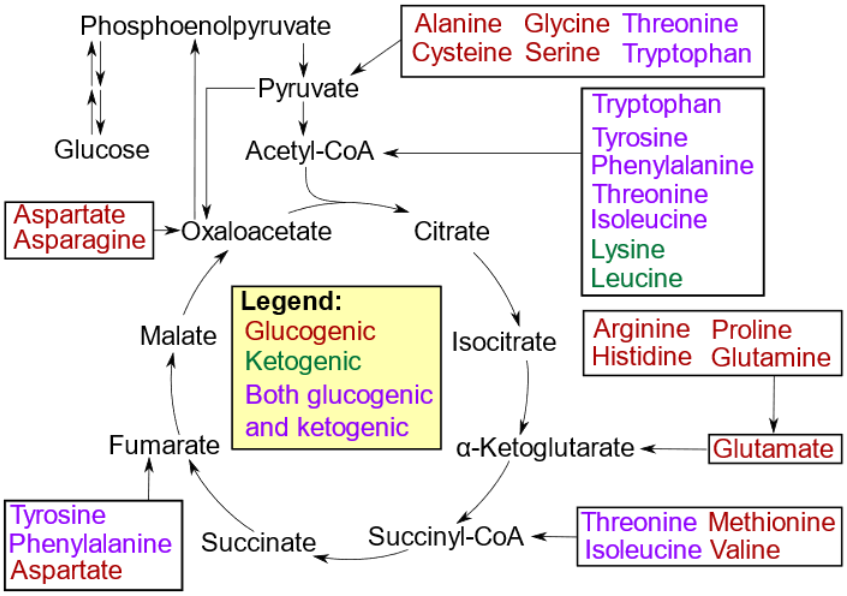


Figure 9.7.3: Illustration of the feeding point of amino acid degradation products into the citric acid cycle and classification of amino acids as glucogenic, ketogenic, or both. (Copyright; Mikael Häggström, CC0, via Wikimedia Commons)

Amino acids that can degrade to pyruvate or oxaloacetate are called **glucogenic** because these products can form glucose through the glucogenesis pathway. For example, alanine yields pyruvate, and aspartate yield oxaloacetate, as shown in reactions 1 and 3 in Figure 9.7.1. Amino acids that degrade to acetyl-(\ce{CoA}\) or acetoacetic acid, which can not form glucose but can be converted into ketone bodies, are called **ketogenic**. Lysine and leucine are ketogenic amino acids. Several amino acids can catabolize to produce both glucogenic and ketogenic intermediates, and they fall into both classes, as shown in Figure 9.7.3.

Index

D

I

J

K

L

M

Glossary

Acetyl-CoA | A molecule that participates in many biochemical reactions in protein, carbohydrate and lipid metabolism. Its main function is to deliver the acetyl group to the citric acid cycle (Krebs cycle) to be oxidized for energy production.

addition polymerization | A reaction in which monomers add to one another to produce a polymeric product that contains all the atoms of the starting monomers.

addition reactions | A reaction in which substituent groups join to hydrocarbon molecules at points of unsaturation—the double or triple bonds.

alcohol | An organic compound with an OH functional group on an aliphatic carbon atom.

aldehyde | An organic compound with a carbonyl functional group that has an hydrogen atom attached and either a hydrocarbon group or a second hydrogen atom.

alkaloid | A nitrogen-containing organic compound obtained from plants that has physiological properties.

alkanes (or saturated hydrocarbons) | A hydrocarbon with only carbon-to-carbon single bonds and existing as a continuous chain of carbon atoms also bonded to hydrogen atoms

alkenes | A hydrocarbon with one or more carbon–carbon double bonds.

alkyl group | A hydrocarbon group derived from an alkane by removal of a hydrogen atom.

alkyl halide (or haloalkane) | A compound resulting from the replacement of a hydrogen atom of an alkane with a halogen atom.

Alkynes | A hydrocarbon with a carbon–carbon triple bond.

alloy | A solid solution of a metal with other substances dissolved in it.

alpha particle | A type of radioactive emission that is equivalent to a helium atom nucleus.

amide | An organic compound with a carbonyl group joined to a nitrogen atom from ammonia or an amine.

amine | An organic compound derived from ammonia by the replacement of one, two, or three of the hydrogens atoms by alkyl or aryl groups.

amino group | An NH_2 unit.

amorphous | A solid with no regular structure.

amphiprotic | A substance that can either donate or accept a proton, depending on the circumstances.

amylopectin | A branched polymer of glucose units found in starch.

anabolism | Metabolic reactions in which molecules are synthesized.

anaerobic metabolism | A biochemical process that takes place in the absence of oxygen.

anions | A negatively charged ion.

anomeric carbon | The carbon atom that was the carbonyl carbon atom in the straight-chain form of a monosaccharide.

anticodon | A set of three nucleotides on the tRNA that is complementary to, and pairs with, the codon on the mRNA.

antioxidants | A substance in foods that acts as a reducing agent.

aromatic compound | Any compound that contains a benzene ring or has certain benzene-like properties.

aromatic hydrocarbons | A hydrocarbon with a benzene-like structure.

aryl group | A group derived from an aromatic hydrocarbon by the removal of a hydrogen atom.

atomic bomb | A weapon that depends on a nuclear chain reaction to generate immense forces.

atomic mass | A weighted average of the masses of all the element's naturally occurring isotopes.

atomic mass unit | One-twelfth the mass of a ^{12}C atom.

atomic radius | The approximate size of an atom.

autoionization of water | The process by which water ionizes into hydronium ions and hydroxide ions as it acts as an acid and a base.

Avogadro's number | The value 6.022×10^{23}.

balanced | A property of a chemical equation when there are the same number of atoms of each element in the reactants and products.

base | A compound that increases the concentration of hydroxide ion (OH^-) in aqueous solution.

Base (or basic) units | A fundamental unit of SI.

beta particle | A type of radioactive emission that is equivalent to an electron.

Bilayers | A double layer of lipids arranged so that nonpolar tails are found between an inner surface and outer surface consisting of hydrophilic heads.

Bile | The yellowish green liquid produced in the liver.

biochemistry | The chemistry of molecules found in living organisms.

boiling point | The temperature at which a substance goes from a liquid to a gas (or from a gas to a liquid).

boiling point elevation | The raising of the boiling point of a solution versus the pure solvent.

bond length | The distance between two nuclei in a covalent bond.

Boyle's law | The gas law that relates pressure and volume.

Brønsted-Lowry base | A compound that accepts a hydrogen ion (H^+) in a reaction; a proton acceptor.

buffer | A solution that resists dramatic changes in pH.

calorie | A unit of energy widely used in the health professions and everyday life.

capacity | The amount of strong acid or base a buffer can counteract.

carbohydrates | A compound composed of carbon, hydrogen, and oxygen atoms that is a polyhydroxy aldehyde or ketone or a compound that can be broken down to form such a compound. It is one of the three main components of the human diet.

carboxyl group | A functional group that contains a carbon–oxygen bond and an OH group also attached to the same carbon atom.

carboxylic acids | An organic compound that has a carboxyl functional group.

cations | A positively charged ion.

cerebrosides | A sphingolipid that contains a fatty acid unit, a sphingosine unit, and galactose or glucose.

chain reaction | An exponential growth in a process.

Charles's law | The gas law that relates volume and absolute temperature.

chemical bond | A very strong attraction between two atoms.

chemical equilibrium (or equilibrium) | The condition in which the extent of a chemical reaction does not change any further.

chemical formula | A concise list of the elements in a compound and the ratios of these elements.

Chemical properties | A characteristic that describes how matter changes its chemical structure or composition.

chemical reaction | A representation of a chemical change.

chemical symbol | A one- or two-letter abbreviation for an element.

chiral carbon | A carbon atom that has four different groups attached to it.

Cholesterol | A steroid that is found in mammals.

cis-trans isomers (or geometric isomers) | Isomers that have different configurations because of the presence of a rigid structure such as a double bond or ring.

citric acid cycle | A cyclic sequence of reactions that brings about the oxidation of a two-C unit to carbon dioxide and water.

codon | A set of three nucleotides on the mRNA that specifies a particular amino acid.

coefficient | A number that gives the number of molecules of a substance in a balanced chemical equation.

coenzymes | A cofactor that is an organic molecule.

colligative properties | A characteristic of solutions that depends only on the number of dissolved particles.

combination (composition) reaction | A chemical reaction that makes a single substance from two or more reactants.

combined gas law | The gas law that relates pressure, volume, and absolute temperature.

combustion reaction | A chemical reaction in which a substance combines with molecular oxygen to make oxygen-containing compounds of other elements in the reaction.

competitive inhibitor | A compound that resembles a particular substrate and competes with the substrate for binding at the active site of an enzyme to slow the rate of the reaction.

complementary bases | Specific base pairings in the DNA double helix.

compound | A substance that can be broken down into chemically simpler components.

concentration | How much solute is dissolved in a certain amount of solvent.

condensed structural formulas | An organic chemical formula that shows the hydrogen atoms (or other atoms or groups) right next to the carbon atoms to which they are attached.

conversion factor | A fraction that has equivalent quantities in the numerator and the denominator but expressed in different units.

core electrons | An electron in a lower-numbered shell of an atom.

covalent network bonding | A type of interaction in which all the atoms in a sample are covalently bonded to other atoms.

curie (Ci) | A unit of radioactivity equal to 3.7×10^{10} decays per second.

cyclic hydrocarbons | A hydrocarbon with a ring of carbon atoms.

cycloalkanes | A cyclic hydrocarbon with only single bonds.

cytochromes | A protein that contains an iron porphyrin in which iron can alternate between Fe(II) and Fe(III).

cytoplasm | Everything between the cell membrane and the nuclear membrane.

decomposition reaction | A chemical reaction in which a single substance is converted into two or more products.

Denaturation | Any change in the three-dimensional structure of a macromolecule that renders it incapable of performing its assigned function.

Density | The mass of an object divided by its volume.

Derived units | A combinations of the SI base units.

diatomic molecules | A two-atom grouping that behaves as a single chemical entity.

digestion | The breakdown of food molecules by hydrolysis reactions into the individual monomer units in the mouth, stomach, and small intestine.

dipole-dipole interaction | An attraction between polar molecules.

Dispersion forces | A force caused by the instantaneous imbalance of electrons about a molecule.

dispersion forces (or London forces) | A force caused by the instantaneous imbalance of electrons about a molecule.

dissociation | The process of cations and anions of an ionic solute separating when the solute dissolves.

Disulfide linkages | A covalent bond that forms by the oxidation and linkage of two sulfur atoms from the side chains of two cysteine residues.

double bond | Two pairs of electrons being shared by two atoms in a molecule.

electron | A subatomic particle with a negative electric charge.

electron configuration | A shorthand description of the arrangement of electrons in an atom.

electron transport chain (or respiratory chain) | An organized sequence of oxidation-reduction reactions that ultimately transports electrons to oxygen, reducing it to water.

electronegativity | A relative measure of how strongly an atom attracts electrons when it forms a covalent bond.

emulsion | A dispersion of two liquids that do not normally mix.

enantiomers | Stereoisomers that are nonsuperimposable mirror images of each other.

endothermic | A process that absorbs energy.

Energy | The ability to do work.

equivalents (Eq) | One mole of charge (either positive or negative).

essential fatty acids | A fatty acid that must be obtained from the diet because it cannot be synthesized by the human body.

ester | An organic compound derived from a carboxylic acid and an alcohol in which the OH of the acid is replaced by an OR group.

esterification | The formation of an ester from a carboxylic acid and an alcohol.

ether | An organic compound that has an oxygen atom between two hydrocarbon groups.

Exact numbers | A number that is defined or counted.

exothermic | A process that gives off energy.

fats | A compound, composed largely of hydrocarbon chains, that supplies energy for the body.

Feedback inhibition | A normal biochemical process that makes use of noncompetitive inhibitors to control some enzymatic activity.

fibrous proteins | A protein that is elongated or fiberlike and insoluble in water.

formula mass | The sum of the masses of the elements in the formula of an ionic compound.

formula unit | A set of oppositely charged ions that compose an ionic compound.

freezing point depression | The lowering of the freezing point of a solution versus the pure solvent.

functional group | A structural arrangement of atoms and/or bonds that imparts a wide range of important properties to organic compounds.

Fusion | A nuclear process in which small nuclei are combined into larger nuclei, releasing energy.

galactosemia | A genetic disease caused by the absence of one of the enzymes needed to convert galactose to glucose.

gamma rays | A type of radioactive emission that is a very energetic form of electromagnetic radiation.

gangliosides | A sphingolipid that contains a fatty acid unit, a sphingosine unit, and a complex oligosaccharide.

gas law | A simple mathematical formula that relates two or more properties of a gas.

Gastric juice | A mixture of water, inorganic ions, hydrochloric acid, and various enzymes and proteins found in the stomach.

Geiger counter | An electrical device that detects radioactivity.

genes | The basic unit of heredity.

genetic code | The identification of each group of three nucleotides and its particular amino acid.

genetic diseases | A hereditary condition caused by an altered DNA sequence.

Globular proteins | A protein that is generally spherical in structure and soluble in water.

glycols | An alcohol with two OH functional groups.

glycolysis | The metabolic pathway in which glucose is broken down to two molecules of pyruvate with the corresponding production of ATP.

glycosidic linkage | The carbon–oxygen-carbon linkage between monosaccharide units in more complex carbohydrates, such as disaccharides or polysaccharides.

groups (or families) | A column of elements on the periodic table.

half reactions | A chemical reaction that shows only oxidation or reduction.

half-life | The amount of time it takes for one-half of a radioactive isotope to decay.

halogenated hydrocarbons | A hydrocarbon in which one or more hydrogen atoms has been replaced by a halogen atom.

halogenation | A reaction in which a halogen reacts at a carbon-to-carbon double or triple bond to add halogen atoms to carbon atoms.

heat | The transfer of energy from one part of the universe to another due to temperature differences.

heat of vaporization | The amount of heat per gram or per mole required for a phase change that occurs at the boiling point.

heterocyclic compounds | A cyclic compound in which one or more atoms in the ring is an element other than a carbon atom.

homogeneous mixtures | A mixture that acts as a single substance so that it is not obvious that two or more substances are present.

homologous series | Any family of compounds in which adjacent members differ from each other by a definite factor.

hydration | Solvation by water molecules.

Hydrogen bonding | Bonding between a highly electronegative oxygen atom or nitrogen atom and a hydrogen atom attached to another oxygen atom or nitrogen atom.

hydrogenation | A reaction in which hydrogen gas reacts at a carbon-to-carbon double or triple bond or a carbon-to-oxygen double bond to add hydrogen atoms to carbon atoms.

hydrolysis | The reaction of a substance with water.

ideal gas law | The gas law that relates volume, pressure, temperature, and amount of a gas.

ideal gas law constant | The constant the appears in the ideal gas law.

immiscible | Liquids that do not dissolve in each other.

induced-fit model | A model that says an enzyme can undergo a conformational change when it binds substrate molecules.

inner transition metals | An element in the two rows beneath the main body on the periodic table. Such metals are also called the lanthanide and actinide elements.

inorganic chemistry | The study of the chemistry of all other elements.

intermolecular interactions | A force of attraction between different molecules.

Ionic bonding | Bonding that results from electrostatic attractions between positively and negatively charged groups.

ionic compounds | A compound formed with an ionic bond.

ionic interactions | An attraction due to ions of opposite charges.

irreversible inhibitor | A substance that inactivates an enzyme by bonding covalently to a specific group at the active site.

isoelectric point | The pH at which a given amino acid exists in solution as a zwitterion.

isomers | Compounds having the same molecular formula but different structural formulas and properties.

isothermal | A process that occurs at constant temperature.

isotopes | Atoms of the same element that have different numbers of neutrons.

IUPAC System of Nomenclature | A systematic way of naming chemical substances so that each has a unique name.

joule | The SI unit of energy, work, and heat.

ketogenic amino acids | An amino acid that is converted to acetoacetyl-CoA or acetyl-CoA, which can be used for the synthesis of ketone bodies but not glucose.

ketone | An organic compound whose molecules have a carbonyl functional group between two hydrocarbon groups.

ketoses | A monosaccharide that contains a ketone functional group on the second carbon atom.

kinetic theory of gases | The fundamental theory of the behavior of gases.

L sugars | A sugar whose Fischer projection terminates in the same configuration as L-glyceraldehyde.

lattice energy | The strength of interactions between atoms that make ionic bonds.

law | A general statement that explains a large number of observations.

law of conservation of matter | In any given system that is closed to the transfer of matter (in and out), the amount of matter in the system stays constant.

Lewis diagrams | A representation that shows valence electrons as dots around the chemical symbol of an atom (also called Lewis electron dot diagrams).

line-angle formula | An organic chemical formula in which carbon atoms are implied at the corners and ends of lines. Each carbon atom is understood to be attached to enough hydrogen atoms to give each carbon atom four bonds.

lipids | A compound isolated from body tissues that is more soluble in organic solvents than in water.

lock-and-key model | A model that portrays an enzyme as conformationally rigid and able to bond only to a substrate or substrates that exactly fit the active site.

mass number | The sum of the numbers of protons and neutrons in a nucleus of an atom.

mass-mass calculations | A stoichiometry calculation converting between the mass of one substance and the mass of a different substance in a chemical reaction.

mass/mass percent | A concentration unit that relates the mass of the solute to the mass of the solution.

mass/volume percent | A concentration unit that relates the mass of the solute to the volume of the solution.

Matter | Anything that has mass and takes up space.

metabolic pathway | A series of biochemical reactions by which an organism converts a given reactant to a specific end product.

microscopic | A view of the universe in which one is working with a few atoms or molecules at a time.

mitochondria | Small, oval organelles with double membranes; the “power plants” of a cell.

modern atomic theory | The fundamental concept that all elements are composed of atoms.

molar mass | The mass of 1 mol of atoms or molecules.

Molarity | Number of moles of solute per liter of solution.

mole-mass calculations | A stoichiometry calculation converting between masses and moles of different substances in a chemical reaction.

mole-mass conversion | The conversion from moles of material to the mass of that same material.

molecular formulas | A chemical formula for a covalent compound.

molecular mass | The mass of a molecule, which is the sum of the masses of its atoms.

molecule | A discrete group of atoms connected by covalent bonds.

mutagens | A chemical or physical agent that cause mutations.

mutarotation | The ongoing interconversion between anomeric forms of a monosaccharide to form an equilibrium mixture.

mutation | Any chemical or physical change that alters the nucleotide sequence in DNA.

neutralization | The reaction of acid and base to make water and a salt.

neutron | A subatomic particle with no electric charge.

nomenclature | The systematic naming of chemical compounds.

nonbonding pairs (or lone pairs) | Electron pair that does not participate in covalent bonds.

noncompetitive inhibitor | A compound that can combine with either the free enzyme or the enzyme-substrate complex at a site distinct from the active site to slow the rate of the reaction.

nonelectrolytes | A compound that does not ionize at all when it dissolves.

nonpolar covalent bond | A covalent bond with a balanced electron distribution across the bond.

Nuclear energy | The controlled harvesting of energy from fission reactions.

nuclear reactor | An apparatus designed to carefully control the progress of a nuclear reaction and extract the resulting energy for useful purposes.

nucleotides | A monomer unit that is linked together to form nucleic acids.

nucleus | The central part of an atom that contains protons and neutrons.

octet rule | The idea that atoms tend to have eight electrons in their valence shell.

oil | A triglyceride that is a liquid at room temperature.

optimum pH | The pH at which a particular enzyme exhibits maximum activity.

Organic chemistry | The study of the chemistry of carbon compounds.

organic compound | A compound containing carbon atoms.

Osmolarity | A way of reporting the total number of particles in a solution to determine the osmotic pressure.

osmotic pressure | The tendency for solvent molecules to move from the more dilute solution to the more concentrated solution until the concentrations of the two solutions are equal.

oxidative deamination | A reaction in which glutamate loses its amino group as an ammonium ion and is oxidized back to α-ketoglutarate.

oxidative phosphorylation | The process that links ATP synthesis to the operation of the electron transport chain.

oxidizing agent | A species that causes oxidation, which is itself reduced.

parts per billion (ppb) | The mass of a solute compared to the mass of a solution times 1,000,000,000.

peptide bond | The amide bond joining two amino acid units in a peptide or protein.

period | A row of elements on the periodic table.

periodic table | A chart of elements that groups the elements by some of their properties.

peripheral proteins | A protein that is more loosely associated with the membrane surface.

pH scale | A logarithmic scale that relates the concentration of the hydrogen ion in solution.

phase | A form of matter that has the same physical properties throughout.

phase change | A physical process in which a substance goes from one phase to another.

phases | A certain form of matter that includes a specific set of physical properties.

phenols | An aromatic compound with an OH group attached directly to a benzene ring.

photosynthesis | The process by which plants use solar energy to convert carbon dioxide and water to glucose.

point mutations | A change in which one nucleotide is substituted, added, or deleted.

polar | A molecule with a net unequal distribution of electrons in its covalent bonds.

polyamide | A condensation polymer in which the monomer units are joined by an amide linkage.

polyatomic ions | An ion with more than one atom.

polycyclic aromatic hydrocarbons (PAHs) | An aromatic hydrocarbon consisting of fused benzene rings sharing a common side.

polymers | A giant molecule formed by the combination of monomers in a repeating manner.

polypeptides | A chain of about 50 or more amino acids.

polysaccharides | A carbohydrate containing many monosaccharide units.

polyunsaturated fatty acids | A fatty acid that has two or more carbon-to-carbon double bonds.

power | The exponent in a number expressed in scientific notation.

pressure | Force divided by area.

primary (1°) alcohol | A compound with an OH group on a carbonatom that is attached to only one other carbon atom.

primary structure | The sequence of amino acids in a polypeptide chain or protein.

products | A substance on the right side of the arrow in a chemical equation.

Proteins | A compound of high molar mass consisting largely or entirely of amino acids linked together.

proton | A subatomic particle with a positive charge.

purines | A heterocyclic amine consisting of a pyrimidine ring fused to a five-member ring with two nitrogen atoms.

quantized | Having a fixed value.

quantum mechanics | The modern theory of electron behavior.

quaternary structure | The arrangement of multiple subunits in a protein.

rad | A unit of radioactive exposure equal to 0.01 J/g of tissue.

radioactivity | Emanations of particles and radiation from atomic nuclei.

reducing sugar | Any carbohydrate capable of reducing a mild oxidizing agent, such as Tollens’ or Benedict’s reagents, without first undergoing hydrolysis.

rem | A unit of radioactive exposure that includes a factor to account for the type of radioactivity.

replication | The process in which the DNA in a dividing cell is copied.

respiration | The process by which cells oxidize organic molecules in the presence of gaseous oxygen to produce carbon dioxide, water, and energy in the form of ATP.

retroviruses | An RNA virus that directs the synthesis of a DNA copy in the host cell.

ribonucleic acid | The nucleic acid responsible for using the genetic information encoded in DNA.

ribosomes | A cellular substructure where proteins are synthesized.

round | The process of assessing the final significant figure of a quantity to determine if it should be kept or moved higher.

saponification | The hydrolysis of fats and oils in the presence of a base to make soap.

science | The process by which we learn about the natural universe by observing, testing, and then generating models that explain our observations.

scientific method | An organized procedure for learning answers to questions.

secondary (2°) alcohol | A compound with an OH group on a carbon atom that is attached to two other carbon atoms.

secondary structure | The fixed arrangement of the polypeptide backbone.

semimetals | An element whose properties are intermediate between metals and nonmetals.

shells | A grouping of electrons within an atom.

significant figures | All the digits of a measured quantity known with certainty and the first uncertain, or estimated, digit.

single bond | A covalent bond formed by a single pair of electrons.

solubility | The limit of how much solute can be dissolved in a given amount of solvent.

solute | The minor component of a solution.

solution | Another name for a homogeneous mixture.

specific heat | A proportionality constant that relates heat to a temperature change.

Sphingolipids | A lipid that contains the unsaturated amino alcohol sphingosine.

Sphingomyelins | A sphingolipid that contains a fatty acid unit, a phosphoric acid unit, a sphingosine unit, and a choline unit.

spontaneous fission | The breaking apart of an atomic nucleus into smaller nuclei.

standard temperature and pressure | 273 K (0°C) and 1.00 atm pressure.

steroids | A lipid with a four-fused-ring structure.

Stock system | The system of indicating a cation's charge with roman numerals.

stoichiometry | The study of the numerical relationships between the reactants and the products in a balanced chemical equation.

strong acid | An acid that is 100% ionized in aqueous solution.

strong base | A base that is 100% ionized in aqueous solution.

structural formula | A chemical formula that shows how the atoms of a molecule are attached to one another.

subshells | A grouping of electrons within a shell.

substrate-level phosphorylation | The synthesis of ATP by the direct transfer of a phosphate group from a metabolite to ADP.

substrates | A compound on which an enzyme acts.

tertiary (3°) alcohol | A compound with an OH group on a carbon atom that is attached to three other carbon atoms.

tertiary (3°) amine | A compound that has three alkyl or aryl groups on the nitrogen atom.

tertiary structure | The unique three-dimensional shape of a polypeptide chain as a whole.

theory | A general statement that describes a large set of observations and data.

Thiols | A compound with an SH functional group.

torr | Another name for millimeters of mercury.

tracer | A substance that can be used to follow the pathway of that substance through some structure.

transamination | An exchange of functional groups between any amino acid and an α-keto acid.

transcription | The process in which RNA is synthesized from a DNA template.

translation | The process in which the information encoded in mRNA is used to direct the sequencing of amino acids to synthesize a protein.

triglycerides | An ester composed of three fatty acid units linked to glycerol and found in fats and oils.

triple bonds | Three pairs of electrons being shared by two atoms in a molecule.

unit | The scale of measurement for a quantity.

unsaturated | A solution whose solute is less than its solubility limit.

unsaturated hydrocarbons | An alkene or alkyne having one or more multiple (double or triple) bonds between carbon atoms.

valence shell electron pair repulsion | The general concept that estimates the shape of a simple molecule.

vapor pressure | The pressure of a vapor that is in equilibrium with its liquid phase.

vapor pressure depression | The lowering of the vapor pressure of a solution versus the pure solvent.

Viruses | An infectious agent that is much smaller and simpler than bacteria.

vitamins | An organic compound that is essential in very small amounts for the maintenance of normal metabolism.

volume | The amount of space that a given substance occupies.

volume/volume percent | A concentration unit that relates the volume of the solute to the volume of the solution.

weak base | A base that is less than 100% ionized in aqueous solution.

zwitterion | An electrically neutral compound that contains both negatively and positively charged groups.

β-oxidation | A sequence of four reactions in which fatty acyl-CoA molecules are oxidized, leading to the removal of acetyl-CoA molecules.

Detailed Licensing

Overview

Title: Introduction to Organic and Biochemistry (Malik)

Webpages: 83

Applicable Restrictions: Noncommercial

All licenses found:

- Public Domain: 97.6% (81 pages)
- Undeclared: 1.2% (1 page)
- CC BY-NC-SA 3.0: 1.2% (1 page)

By Page

- Introduction to Organic and Biochemistry (Malik) - *Public Domain*
  - Front Matter - *Public Domain*
    - TitlePage - *Public Domain*
    - InfoPage - *Public Domain*
    - Table of Contents - *Undeclared*
    - Licensing - *Public Domain*
    - Dedication - *Public Domain*
  - 1: Bonding in organic compounds - *Public Domain*
    - 1.1: What is organic chemistry? - *Public Domain*
    - 1.2: What is a chemical bond - *Public Domain*
    - 1.3: Hybridization of orbitals and 3D structures of simple organic compounds - *Public Domain*
    - 1.4: Representing organic compounds - *Public Domain*
    - 1.5: Formal Charge - *Public Domain*
    - 1.6: Resonance - *Public Domain*
  - 2: Nomenclature and physical properties of organic compounds - *Public Domain*
    - 2.1: Hydrocarbons - *Public Domain*
    - 2.2: Functional group - *Public Domain*
    - 2.3: Functional groups containing sp3-hybridized heteroatom - *Public Domain*
    - 2.4: Functional groups containing sp2-hybridized heteroatom - *Public Domain*
    - 2.5: Functional groups containing mix of sp3- and sp2-, or sp-hybridized heteroatom - *Public Domain*
  - 3: Stereochemistry - *Public Domain*
    - 3.1: Introduction to stereochemistry - *Public Domain*
    - 3.2: Confirmations - *Public Domain*
    - 3.3: Configurations - *Public Domain*
  - 4: Organic reactions - *Public Domain*
    - 4.1: What is a reaction mechanism - *Public Domain*
    - 4.2: Free Radical Reactions - *Public Domain*
    - 4.3: Acid-base reactions - *Public Domain*
    - 4.4: Nucleophilic substitution and elimination reactions - *Public Domain*
    - 4.5: Nucleophilic acyl substitution reactions - *Public Domain*
    - 4.6: Nucleophilic Addition Reactions - *Public Domain*
    - 4.7: Electrophilic addition reactions - *Public Domain*
    - 4.8: Electrophilic aromatic substitution reactions - *Public Domain*
    - 4.9: Reduction and oxidation (redox) reactions - *Public Domain*
    - 4.10: Reactions with cyclic transition state - *Public Domain*
  - 5: Carbohydrates - *Public Domain*
    - 5.1: What are carbohydrates - *Public Domain*
    - 5.2: General class names and Common names of monosaccharides - *Public Domain*
    - 5.3: Cyclic structures of monosaccharides - *Public Domain*
    - 5.4: Reactions of monosaccharides - *Public Domain*
    - 5.5: Disaccharides - *Public Domain*
    - 5.6: Oligosaccharides - *Public Domain*
    - 5.7: Polysaccharides - *Public Domain*
  - 6: Lipids - *Public Domain*
    - 6.1: What are lipids? - *Public Domain*
    - 6.2: Fatty acyls - *Public Domain*
    - 6.3: Glycerolipids - *Public Domain*
    - 6.4: Glycerophospholipids - *Public Domain*
    - 6.5: Sphingolipids - *Public Domain*
    - 6.6: Saccharolipids, polyketides, and prenols - *Public Domain*
    - 6.7: Sterols - *Public Domain*
    - 6.8: Cell membrane - *Public Domain*
  - 7: Proteins - *Public Domain*
    - 7.1: What are proteins? - *Public Domain*
    - 7.2: Amino acids - *Public Domain*